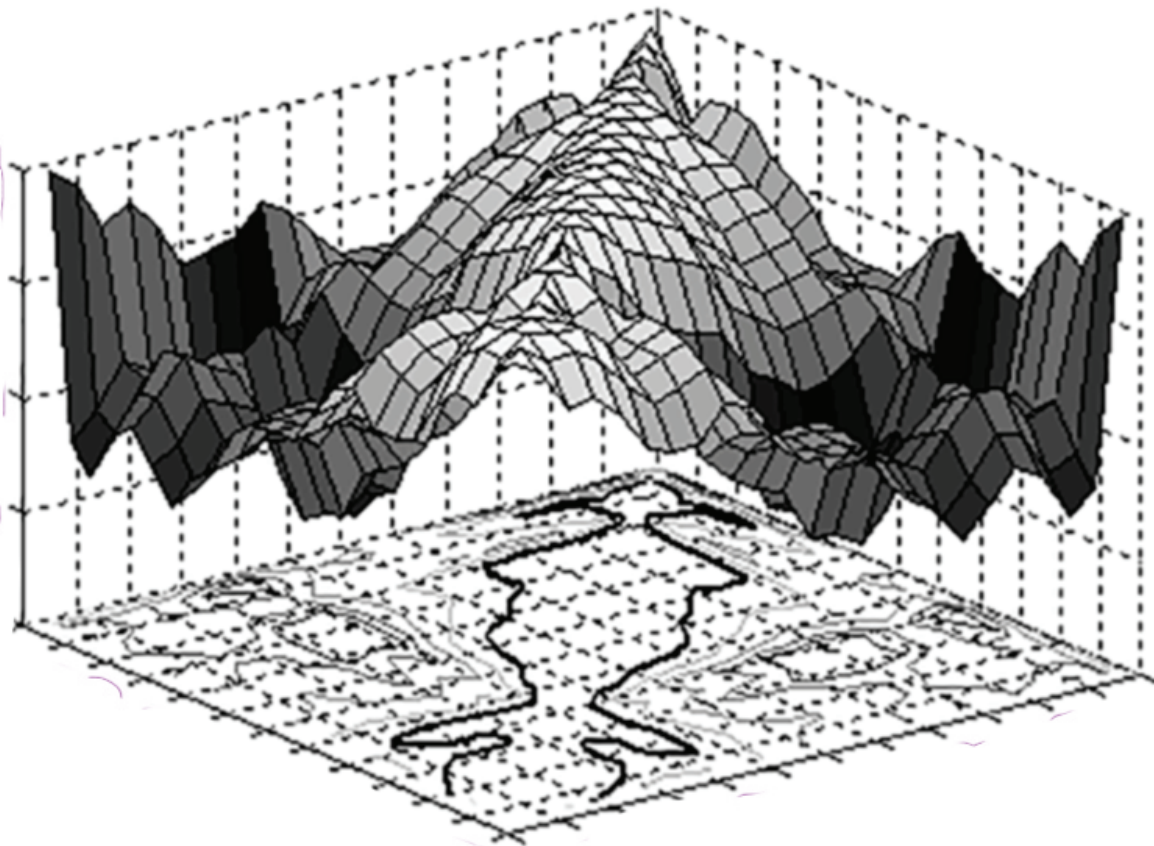


DYNA

Journal of the Facultad de Minas, Universidad Nacional de Colombia - Medellin Campus



DYNA 83 (197), June, 2016 - ISSN 0012-7353
Tarifa Postal Reducida No. 2014-287 4-72
La Red Postal de Colombia, Vence 31 de Dic. 2016.



UNIVERSIDAD
NACIONAL
DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS

DYNA is an international journal published by the Facultad de Minas, Universidad Nacional de Colombia, Medellín Campus since 1933. DYNA publishes peer-reviewed scientific articles covering all aspects of engineering. Our objective is the dissemination of original, useful and relevant research presenting new knowledge about theoretical or practical aspects of methodologies and methods used in engineering or leading to improvements in professional practices. All conclusions presented in the articles must be based on the current state-of-the-art and supported by a rigorous analysis and a balanced appraisal. The journal publishes scientific and technological research articles, review articles and case studies.

DYNA publishes articles in the following areas:

Organizational Engineering
Civil Engineering
Materials and Mines Engineering

Geosciences and the Environment
Systems and Informatics
Chemistry and Petroleum

Mechatronics
Bio-engineering
Other areas related to engineering

Publication Information

DYNA (ISSN 0012-73533, printed; 2346-2183, online) is published by the Facultad de Minas, Universidad Nacional de Colombia, with a bimonthly periodicity (February, April, June, August, October, and December). Circulation License Resolution 000584 de 1976 from the Ministry of the Government.

Contact information

Web page: <http://dyna.unalmed.edu.co>
E-mail: dyna@unal.edu.co
Mail address: Revista DYNA
Facultad de Minas Universidad Nacional de Colombia -
Medellín Campus
Carrera 80 No. 65-223 Bloque M9 - Of.:107
Telephone: (574) 4255068 Fax: (574) 4255343
Medellín - Colombia

© Copyright 2016. Universidad Nacional de Colombia

The complete or partial reproduction of texts with educational ends is permitted, granted that the source is duly cited. Unless indicated otherwise.

Notice

All statements, methods, instructions and ideas are only responsibility of the authors and not necessarily represent the view of the Universidad Nacional de Colombia. The publisher does not accept responsibility for any injury and/or damage for the use of the content of this journal.

The concepts and opinions expressed in the articles are the exclusive responsibility of the authors.

Institutional Exchange Request

DYNA may be requested as an institutional exchange through the e-mail canjebib_med@unal.edu.co or to the postal address:

Biblioteca Central "Efe Gómez"
Universidad Nacional de Colombia, Sede Medellín
Calle 59A No 63-20
Teléfono: (57+4) 430 97 86
Medellín - Colombia

Indexing and Databases

DYNA is admitted in:

The National System of Indexation and Homologation of Specialized Journals CT+I-PUBLINDEX, Category A1

Science Citation Index Expanded
Web of Science - WoS, Thomson Reuters
Journal Citation Reports - JCR
SCImago Journal & Country Rank - SJR
Science Direct
SCOPUS
Chemical Abstract - CAS
Scientific Electronic Library on Line - SciELO
GEOREF
PERIÓDICA Data Base
Latindex
Actualidad Iberoamericana
RedALyC - Scientific Information System
Directory of Open Acces Journals - DOAJ
PASCAL
CAPES
UN Digital Library - SINAB
CAPES
EBSCO Host Research Databases

Publisher's Office

Juan David Velásquez Henao, Director
Mónica del Pilar Rada T., Editorial Coordinator
Catalina Cardona A., Editorial Assistant
Amilkar Álvarez C., Diagrammer
Byron Llano V., Editorial Assistant
Landsoft S.A., IT

Reduced Postal Fee

Tarifa Postal Reducida # 2014-287 4-72. *La Red Postal de Colombia*, expires Dec. 31st, 2016



UNIVERSIDAD NACIONAL DE COLOMBIA

SEDE MEDELLÍN
FACULTAD DE MINAS

DYNA



UNIVERSIDAD NACIONAL DE COLOMBIA

SEDE MEDELLÍN
FACULTAD DE MINAS



**UNIVERSIDAD
NACIONAL
DE COLOMBIA**
SEDE MEDELLÍN
FACULTAD DE MINAS



DYNÁ
<http://dyna.medellin.unal.edu.co/>

COUNCIL OF THE FACULTAD DE MINAS

Dean

Pedro Nel Benjumea Hernández PhD

Vice-Dean

Ernesto Pérez González, PhD

Vice-Dean of Research and Extension

Santiago Arango Aramburo, PhD

Director of University Services

Carlos Alberto Graciano, PhD

Academic Secretary

Francisco Javier Díaz Serna, PhD

Representative of the Curricular Area Directors

Carlos Jaime Franco Cardona, PhD

Representative of the Curricular Area Directors

Eliana Isabel Arango Zuluaga, PhD

Representative of the Basic Units of Academic-Administrative Management

Rosa Elvira Correa Gutiérrez, PhD

Representative of the Basic Units of Academic-Administrative Management

Rafael Esteban Ribadeneira Paz, PhD

Professor Representative

Luis Hernán Sánchez Arredondo, MSc

Delegate of the University Council

León Restrepo Mejía, PhD

Student representative (alternate) undergraduate at the Faculty Council

Sergio Esteban Ospina Rendón

FACULTY EDITORIAL BOARD

Dean

Pedro Nel Benjumea Hernández, PhD

Vice-Dean of Research and Extension

Santiago Arango Aramburo, PhD

Members

Oscar Jaime Restrepo Baena, PhD

Juan David Velásquez Henao, PhD

Jaime Aguirre Cardona, PhD

Mónica del Pilar Rada Tobón, MSc

JOURNAL EDITORIAL BOARD

Editor-in-Chief

Juan David Velásquez Henao, PhD
Universidad Nacional de Colombia, Colombia

Editors

George Barbastathis, PhD
Massachusetts Institute of Technology, USA

Tim A. Osswald, PhD
University of Wisconsin, USA

Juan De Pablo, PhD
University of Wisconsin, USA

Hans Christian Öttinger, PhD
Swiss Federal Institute of Technology (ETH), Switzerland

Patrick D. Anderson, PhD
Eindhoven University of Technology, the Netherlands

Igor Emri, PhD
Associate Professor, University of Ljubljana, Slovenia

Dietmar Drummer, PhD
Institute of Polymer Technology University Erlangen-Nürnberg, Germany

Ting-Chung Poon, PhD
Virginia Polytechnic Institute and State University, USA

Pierre Boulanger, PhD
University of Alberta, Canadá

Jordi Payá Bernabeu, Ph.D.
Instituto de Ciencia y Tecnología del Hormigón (ICITECH)
Universitat Politècnica de Valencia, España

Javier Belzunce Varela, Ph.D.
Universidad de Oviedo, España

Luis Gonzaga Santos Sobral, PhD
Centro de Tecnología Mineral - CETEM, Brasil

Agustín Bueno, PhD
Universidad de Alicante, España

Henrique Lorenzo Cimadevila, PhD
Universidad de Vigo, España

Mauricio Trujillo, PhD
Universidad Nacional Autónoma de México, México

Carlos Palacio, PhD

Universidad de Antioquia, Colombia

Jorge García-Sucerquia, PhD

Universidad Nacional de Colombia, Colombia

Juan Pablo Hernández, PhD

Universidad Nacional de Colombia, Colombia

John William Branch Bedoya, PhD

Universidad Nacional de Colombia, Colombia

Enrique Posada, Msc

INDISA S.A, Colombia

Oscar Jaime Restrepo Baena, PhD

Universidad Nacional de Colombia, Colombia

Moisés Oswaldo Bustamante Rúa, PhD

Universidad Nacional de Colombia, Colombia

Hernán Darío Álvarez, PhD

Universidad Nacional de Colombia, Colombia

Jaime Aguirre Cardona, PhD

Universidad Nacional de Colombia, Colombia



CONTENTS

Risk management of occupational exposure to nanoparticles during a development project: A case study Francisco Silva, Pedro Arezes & Paul Swuste	9
Prevalence and symptomatology of musculoskeletal problems reported by home care service workers caring for the elderly Dina Chagas	17
Current state of the art and enduring issues in anthropometric data collection Sara Bragança, Pedro Arezes, Miguel Carvalho & Susan P. Ashdown	22
Understanding finger postures when touching targets on the touchscreen of mobile devices Mahrus K. Umami, Pedro M. Arezes & Álvaro M. Sampaio	31
Tackling the challenges of an aging workforce with the use of wearable technologies and the quantified-self Martin Lavallière, Arielle A. Burstein, Pedro Arezes & Joseph F. Coughlin	38
Analysis and characterization of starchy and cellulosic materials after enzymatic modification Juan A. Quintero, Javier A. Dávila, Jonathan Moncada, Oscar H. Giraldo & Carlos A. Cardona	44
Robust sample size for weibull demonstration test plan Manuel R. Piña-Monarez, Miriam L. Ramos-López, Alejandro Alvarado-Iniesta & Rey D. Molina-Arredondo	52
A new hardening rule to model the effect of density on soil behavior Márcio Muniz de Farias, Robinson Andrés Giraldo-Zuluaga & Teruo Nakai	58
Monitoring and evaluation of energy use in oil treatment facilities Laura Milena Corredor-Rojas, Ariel Uribe-Rodríguez, Juan Carlos Cobos-Gómez	68
Differences between the use of ferric sulphate and ferric chloride on biodesulfurization of a large coal particle Gerardo Andrés Calcedo-Pineda & Marco Antonio Márquez-Godoy	74
Human error and response to alarms in process safety Beata Mrugalska, Salman Nazir, Edwin Tytyk & Kjell Ivar Øvergård	81
Changes in EEG amplitude (Alpha and Beta waves) with Thermal environment Emanuel Tiago-Costa, Emilia Quelhas-Costa & João Santos-Baptista	87
Effects of joints and their waterstops on pressures spread over a slab subject to turbulent flow on a horizontal apron Mauricio González-Betancourt & Lilian Posada-García	94
Assessing the loss-of-insulation life of power transformers by estimating their historical loads and ambient temperature profiles using ANNs and Monte Carlo simulations Andrés Arturo Romero-Quete, Enrique Esteban Mombello & Giuseppe Rattá	104
Thermal assessment of ecological tiles in physical models of poultry houses Flavio Alves Damasceno a, Leonardo Schiassi, Tadayuki Yanagi Junior, Jairo Alexander Osorio-Saraz & Jofran Luiz-de Oliveira	114
Statistical performance of control charts with variable parameters for autocorrelated processes Oscar Oviedo-Trespalcacios & Rita Peñabaena-Niebles	120
Symbolic modeling of the Pareto-Optimal sets of two unity gain cells Said Polanco-Martagón, José Ruiz-Ascencio & Miguel Aurelio Duarte-Villaseñor	128
Thermodynamic properties of moisture sorption in cassava flour Alfredo A. Ayala-Aponte	138
Cr/CrN bilayer coating behavior under erosion-corrosion degradation by jet impingement testing Johanna Alegría-Ortega, Luz Marina Ocampo-Carmona, Fernando Rodríguez & Elena Forlerer	145
Validation of an inertial sensor-based platform to acquire kinematic information for human joint angle estimation Mauro Callejas-Cuervo, Andrés Felipe Ruiz-Olaya & Rafael M. Gutiérrez	153
Study of the evolution of the residual stress state in thermal barrier coatings sprayed on AISI 304 stainless steel Pedro Yáñez-Contreras, José Dolores Oscar Barceinas-Sánchez, Carlos Agustín Poblano-Salas, José Martín Medina-Flores, Adrián Luis García-García & Iván Domínguez-López	159

Burst erasure correction using LDPC codes constructed on base matrices generated by matched groups, nested polygons and superposed circulant matrices	166
Cassio André Sousa-da Silva & Evaldo Pellaes	
Heat flow assessment in an underground mine: An approach to improve the environmental conditions	174
Marc Bascompta, Ana María Castañón, Lluís Sanmiquel & Josep Oliva	
Secure point-to-point communication using chaos	180
Maricela Jiménez-Rodríguez, María Guadalupe González-Novoa, Juan Carlos Estrada-Gutiérrez, Cuauhtemoc Acosta-Lúa & Octavio Flores-Siordia	
Las TIC navegando hacia la sostenibilidad económica y ambiental en el transporte marítimo	187
Alfonso I. López-Díaz, Miguel Ángel Gutiérrez-García, Diego González-Aguilera, Alejandro Morales-Sánchez & Jaime Ruiz-Martí	
Prioritization of TQM practices in Brazilian medical device SMEs using Analytical Hierarchy Process (AHP)	194
Glaucio Henrique de Sousa-Mendes, Eduardo Gomes-Salgado & Bruno Enrico Moro-Ferrari	
Strategic guidelines for supply chain coordination in healthcare and a mathematical model as a proposed mechanism for the measurement of coordination effects	203
Delio Alexander Balcázar-Camacho, Cesar Amílcar López-Bello & Wilson Adarme-Jaimes	
Variables relevantes para la medición de la calidad percibida del servicio bancario	212
Carmen Regina Berdugo-Correa, Rodrigo Alberto Barbosa-Correa & Lina Margarita Prada-Angarita	
Coordinated recharge of electric vehicles in real time	222
Cristian C. Mendoza, Adriana M. Quintero, Francisco Santamaria & Alexander Alarcon	
Physicochemical characterization of four species of agaves with potential in obtaining pulp for paper making	232
Edith Jiménez-Muñoz, Francisco Prieto-García, Judith Prieto-Méndez, Otilio A. Acevedo-Sandoval & Rodrigo Rodríguez-Laguna	

Our cover

Image alluding to Article:

Assessing the loss-of-insulation life of power transformers by estimating their historical loads and ambient temperature profiles using ANNs and Monte Carlo simulations

Authors:

Andrés Arturo Romero-Quete, Enrique Esteban Mombello & Giuseppe Rattá



CONTENIDO

Gestión de riesgos de exposición ocupacional a las nanopartículas en un proyecto en desarrollo: Estudio de caso	9
Francisco Silva, Pedro Arezes & Paul Swuste	
Prevalencia y sintomatología músculo-esquelética auto-referida por los trabajadores del servicio de apoyo domiciliario en la prestación de atención a personas mayores	17
Dina Chagas	
Estado actual de la técnica y cuestiones perdurables en la recogida de datos antropométricos	22
Sara Bragança, Pedro Arezes, Miguel Carvalho & Susan P. Ashdown	
Comprensión de las posturas de dedos al tocar objetivos en la pantalla táctil de dispositivos móviles	31
Mahrus K. Umami, Pedro M. Arezes & Álvaro M. Sampaio	
Hacer frente a los desafíos de una fuerza laboral que envejece con el uso de tecnologías usables y la auto-cuantificación	38
Martin Lavallière, Arielle A. Burstein, Pedro Arezes & Joseph F. Coughlin	
Análisis y caracterización de materiales amiláceos y celulósicos después de modificación enzimática	44
Juan A. Quintero, Javier A. Dávila, Jonathan Moncada, Oscar H. Giraldo & Carlos A. Cardona	
Tamaño de muestra robusta para planes de demostración weibull	52
Manuel R. Piña-Monarez, Miriam L. Ramos-López, Alejandro Alvarado-Iniesta & Rey D. Molina-Arredondo	
Una nueva regla endurecimiento para modelar el efecto de la densidad en el comportamiento de los suelos	58
Márcio Muniz de Farias, Robinson Andrés Giraldo-Zuluaga & Teruo Nakai	
Monitoreo y evaluación del uso de la energía en estaciones de tratamiento de crudo	68
Laura Milena Corredor-Rojas, Ariel Uribe-Rodríguez, Juan Carlos Cobos-Gómez	
Diferencias entre el uso de sulfato férrico y cloruro férrico en la biodesulfurización de un carbón con tamaño de partícula grueso	74
Gerardo Andrés Caicedo-Pineda & Marco Antonio Márquez-Godoy	
Error humano y reacción a alarmas en seguridad de procesos	81
Beata Mrugalska, Salman Nazir, Edwin Tytyk & Kjell Ivar Øvergård	
Cambios en la amplitud del EEG (ondas Alfa y Beta) con el medio ambiente térmico	87
Emanuel Tiago-Costa, Emília Quelhas-Costa & João Santos-Baptista	
Efectos de las juntas y sus sellos en las presiones distribuidas en una losa sujeta a flujo turbulento en un canal horizontal	94
Mauricio González-Betancourt & Lilian Posada-García	
Evaluación de la pérdida de vida del aislamiento sólido en transformadores de potencia, estimando la historia de carga y los perfiles de temperatura ambiente por medio de redes neuronales artificiales y simulaciones de Monte Carlo	104
Andrés Arturo Romero-Quete, Enrique Esteban Mombello & Giuseppe Rattá	
Evaluación térmica de tejas ecológicas en modelos físicos de galpones avícolas	114
Flavio Alves Damasceno a, Leonardo Schiassi, Tadayuki Yanagi Junior, Jairo Alexander Osorio-Saraz & Jofran Luiz-de Oliveira	
Desempeño estadístico de cartas de control con parámetros variables para procesos autocorrelacionados	120
Oscar Oviedo-Trespalacios & Rita Peñabaena-Niebles	
Modelado simbólico del conjunto Óptimo de Pareto de dos celdas de ganancia unitaria	128
Said Polanco-Martagón, José Ruiz-Ascencio & Miguel Aurelio Duarte-Villaseñor	
Propiedades termodinámicas de humedad de sorción en harina de yuca	138
Alfredo A. Ayala-Aponte	
Comportamiento de un recubrimiento bicapa de Cr/CrN bajo degradación por erosión-corrosión por ensayo de impacto con chorro	145
Johanna Alegría-Ortega, Luz Marina Ocampo-Carmona, Fernando Rodríguez & Elena Forlerer	
Validación de una plataforma basada en sensores inerciales para adquirir información cinemática de estimación del ángulo articular humano	153
Mauro Callejas-Cuervo, Andrés Felipe Ruiz-Olaya & Rafael M. Gutiérrez	

Estudio de la evolución del perfil de esfuerzos residuales en recubrimientos barrera térmica depositados sobre acero inoxidable AISI 304	159
Pedro Yáñez-Contreras, José Dolores Oscar Barceinas-Sánchez, Carlos Agustín Poblano-Salas, José Martín Medina-Flores, Adrián Luis García-García & Iván Domínguez-López	
Corrección de borrado en ráfaga utilizando códigos LDPC construidos sobre matrices generadas por grupos combinados, polígonos anidados y matrices circulantes superpuestas	166
Cassio André Sousa-da Silva & Evaldo Pelaes	
Evaluación de los flujos de calor en una mina subterránea y enfoque para mejorar sus condiciones ambientales	174
Marc Bascompta, Ana María Castañón, Lluís Sanmiquel & Josep Oliva	
Comunicación punto a punto segura usando caos	180
Maricela Jiménez-Rodríguez, María Guadalupe González-Novoa, Juan Carlos Estrada-Gutiérrez, Cuauhtemoc Acosta-Lúa & Octavio Flores-Siordia	
ICT sailing to economic and environmental sustainability in the maritime transport	187
Alfonso I. López-Díaz, Miguel Ángel Gutiérrez-García, Diego González-Aguilera, Alejandro Morales-Sánchez & Jaime Ruiz-Martí	
Priorización de las prácticas de GCT en empresas de dispositivos médicos brasileños utilizando el Analytical Hierarchy Process (AHP)	194
Glaucio Henrique de Sousa-Mendes, Eduardo Gomes-Salgado & Bruno Enrico Moro-Ferrari	
Lineamientos estratégicos para coordinación en la cadena de suministro de medicamentos y propuesta de un modelo matemático para medir los efectos de la coordinación	203
Delio Alexander Balcázar-Camacho, Cesar Amílcar López-Bello & Wilson Adarme-Jaimes	
Relevant variables in the measuring of the perceived quality in bank service	212
Carmen Regina Berdugo-Correa, Rodrigo Alberto Barbosa-Correa & Lina Margarita Prada-Angarita	
Coordinación de la recarga de vehículos eléctricos en tiempo real	222
Cristian C. Mendoza, Adriana M. Quintero, Francisco Santamaría & Alexander Alarcon	
Caracterización fisicoquímica de cuatro especies de agaves con potencialidad en la obtención de pulpa de celulosa para elaboración de papel	232
Edith Jiménez-Muñoz, Francisco Prieto-García, Judith Prieto-Méndez, Otilio A. Acevedo-Sandoval & Rodrigo Rodríguez-Laguna	

Nuestra carátula

Imagen alusiva al artículo:

Evaluación de la pérdida de vida del aislamiento sólido en transformadores de potencia, estimando la historia de carga y los perfiles de temperatura ambiente por medio de redes neuronales artificiales y simulaciones de Monte Carlo

Autores:

Andrés Arturo Romero-Quete, Enrique Esteban Mombello & Giuseppe Rattá



Risk management of occupational exposure to nanoparticles during a development project: A case study

Francisco Silva,^{a,b} Pedro Arezes^b & Paul Swuste^c

^a Technological Centre for Ceramic and Glass, Coimbra, Portugal. fsilva@ctcv.pt

^b Human Engineering Group, Production and Systems Department, University of Minho, Guimarães, Portugal. perez@dpd.uminho.pt

^c Safety Science Group, Delft University of Technology, Delft, The Netherlands. P.H.J.J.Swuste@tudelft.nl

Received: December 09th, 2015. Received in revised form: March 18th, 2016. Accepted: April 12th, 2016.

Abstract

The production of nanotechnology based products is increasing, along with the conscience of the possible harmful effects of some nanomaterials. Along with technological advances, there is the need to improve knowledge of safety and health and apply that knowledge to the workplace. The “safety-by-design” approaches are attracting attention as helpful tools to develop safer products and production processes. The Systematic Design Analysis Approach could help to identify the solutions to control workplace risks by defining the emission and exposure scenarios and the possible barriers to interrupt them. When managing risks during a photocatalytic ceramic tiles development project, it was possible to identify relevant nanoparticles emission scenarios and related barriers. Possible ways to reduce them could then be defined, which would in turn, lead to an inherently safer production process.

Keywords: photocatalytic ceramic tiles; risk assessment; systematic design analysis; inherently safer process.

Gestión de riesgos de exposición ocupacional a las nanopartículas en un proyecto en desarrollo: Estudio de caso

Resumen

La producción de productos basados en la nanotecnología va en aumento, junto con la conciencia de los posibles efectos nocivos de algunos nanomateriales. Junto con los avances tecnológicos, existe la necesidad de mejorar el conocimiento de la seguridad y salud y aplicar ese conocimiento en los entornos laborales. Los enfoques “Safety-by-design” están atrayendo la atención como herramientas útiles para desarrollar productos y procesos de producción más seguros. El enfoque de Análisis Sistemática de Diseño podría ayudar a identificar las soluciones para el control de los riesgos laborales mediante la definición de los escenarios de emisiones y de exposición y los posibles obstáculos a interrumpirlos. Cuando la gestión de riesgos durante un proyecto de desarrollo de las azulejos cerámicos fotocatalíticos, fue posible identificar escenarios de emisiones de las nanopartículas relevantes y las barreras relacionadas. Así, las posibles formas de reducirlos podrían ser definidas, lo que, a su vez, pueden dar lugar a un proceso de producción inherentemente más seguro.

Palabras clave: azulejos cerámicos fotocatalíticos; evaluación de riesgos; análisis sistemática de diseño; proceso inherentemente más seguro.

1. Introduction

Photocatalytic ceramic tiles containing nano-sized titanium dioxide (TiO₂) have self-cleaning characteristics and are also able to transform some air pollutants like nitrogen oxides, contribute to a cleaner ambient air, and

reveal anti-bacterial properties [1].

In general, the in-vitro and in-vivo tests done with both fine (particles with nominal diameter > 100 nm) and ultrafine TiO₂ particles (with nominal diameter <100 nm, also called nanoparticles or nano-sized particles), have potentially harmful health effects in humans. TiO₂ nanoparticles induce

How to cite: Silva, F., Arezes, P. and Swuste, P., Risk management of occupational exposure to nanoparticles during a development project: A case study. DYNA 83(197), pp. 9-16, 2016.

inflammatory responses in the lung tissue, particularly in high doses [2]. The International Agency for Research on Cancer (IARC) classified TiO_2 as “possibly carcinogenic to humans”: a carcinogenic Group 2B substance [3]. In a review on the animal and human data relevant to assessing the carcinogenicity of TiO_2 , published in 2011, the National Institute for Safety and Health (NIOSH), concluded that exposure to ultrafine (or nano) TiO_2 should be considered a potential occupational carcinogenic agent, and recommended an airborne exposure limit of 2.4 mg/m^3 for fine TiO_2 and 0.3 mg/m^3 for ultrafine (including engineered nanoscale) TiO_2 [4].

Some authors have been defending the need for methodologies that address the risks related to nanotechnologies, based on the processes or product design [5-7]. One approach cited in the literature is the “Design for Safer Nanotechnology” proposed by Morose [8] in which the author suggests an intervention during the design stage for nano-objects and products that incorporate them. Schulte et al. [5] also mention the Prevention through Design (PtD) initiative as a valuable methodology to manage occupational risks. Swuste and Zalk [9] also propose the use of design analysis to achieve safer production processes in the nanotechnology field.

The aim of this paper is to present the work that is carried out to establish a safer production process resulting from a development project. The underlying research questions are:

- Does a design approach to the production line of photocatalytic ceramic tiles generate relevant emission scenarios and related barriers?
- What are the possibilities of the Systematic Design Analysis Approach (SYDAPP) reducing emission scenarios during the production of photocatalytic ceramic tiles?
- Could managing risks during the development phase of a new production process help to define safer processes?

2. Methodology

2.1. Framework

The work presented in this paper was performed during a photocatalytic ceramic tiles development project, using TiO_2 (anatase) and employing a common ceramics production processes. It was part of a SELFCLEAN funded research project.

The project lasted for approximately two years, from the first exploratory tests to the final product prototype. The OSH intervention, including the work described in this paper, lasted six-months, plus another two months to produce the OSH issues report.

The project team included several materials science researchers and engineers from a university, a technological institute and a ceramic tile company, and one occupational safety and hygiene (OSH) practitioner.

Discussions on the health and safety aspects for the project were held on an interdisciplinary knowledge basis. These discussions were complemented by observation and information collection during laboratory sessions and semi-industrial tests that were performed during the project.

OSH issues were included in the agenda of three of the project’s plenary meetings. For approximately 45 minutes in each of the meetings, the SYDAPP was presented and the team members had the opportunity to contribute their inputs to the process design analysis and related emission and exposure scenarios. The group discussions gathered contributions, particularly from the design analysis, the identification of emission and exposure scenarios and the possible barriers. The experts proposed alternative production principles and forms, including their feasibility evaluation, which helped to identify their impact on the possible scenarios. In parallel, several face-to-face informal meetings were held by the OSH practitioner with the other members of the group, including the ceramic company engineer and the university researchers, in order to refine knowledge on different options and confirm information collected during the meetings and project tests. Finally, the OSH practitioner, based on the collected information, produced a report for the project manager.

2.2. Systematic design analysis approach

Although occupational safety and hygiene research pays more attention to risk analysis [10], several authors in this domain have undertaken research in the safety by design field, especially the Safety Science Group from Delft University of Technology [11–13]. Swuste [10], for example, proposed a systematic approach towards solutions based on three complementary elements:

- A hazard process model;
- Design analysis;
- A problem-solving cycle.

The two first elements are the basis for the SYDAPP. Combining the process design analysis with the emission and exposure scenarios, it is possible to acquire a clear vision of how the different process operations will affect a worker’s exposure.

2.3. Hazard process model - Bow-tie

The bow-tie model is used in the safety science field as a tool to prevent the occurrence of accidents [14]. Its adaptation to the occupational hygiene field (see Fig. 1) helps to establish the necessary barriers to control risks arising from different workplace exposure scenarios [15]. The use of the bow-tie model as a support tool to risk management is also referred to by Fleury et al. [7]. An example of the use of this model, defining exposure scenarios and evaluating the risks during the production of carbon nanotubes polymer composites is presented in another article [16].

The bow-tie model also stresses the importance of management as the entity responsible for implementing the barriers [17].

When the bow-tie model and the design analysis are performed together, it is possible to have a detailed vision of the production process and the occupational risks related with the production process. Emissions and, consequently, exposure are identified on a production form level. Thus, the options to reduce emissions and exposure are usually limited to LEV and personal protective equipment. As these controls

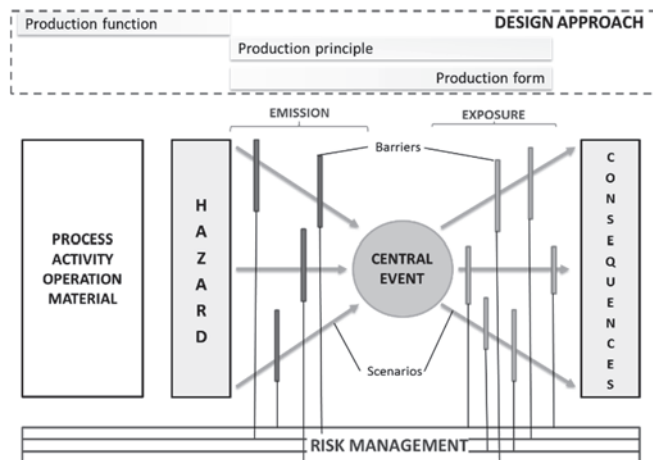


Figure 1. Bow-tie model with arrows representing different exposure scenarios.

Source: The authors

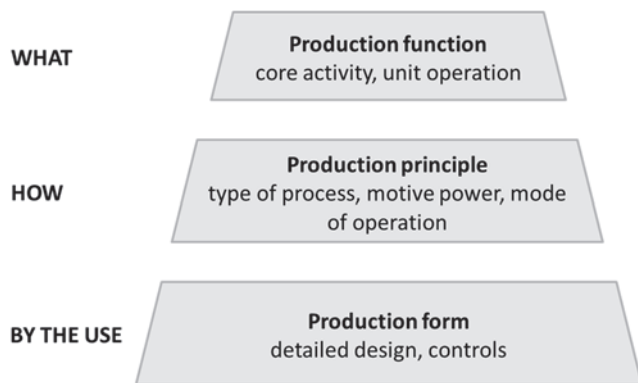


Figure 2. Design analysis hierarchy [10].

Source: The authors

could become ineffective due to the high level of exposure, or their own characteristics, it is useful to act with production principles or production functions, which provide more operative controls.

2.4. Design analysis

The design analysis methodology allows workplace conditions to be understood and studied. In design analysis, the production process is split into three decision levels (see Fig. 2), described below [10]:

- **Production function:** is the highest level and divides the production process into its core activities, similar to unit operations;
- **Production principle:** identifies the general process, motives, power, and operational control methods by which the production function can be achieved;
- **Production form:** is the lowest level and specifies the detailed design by which the production principle will be accomplished.

If there are a large number of production processes, the type of functions (or unit operations in rigor) in which each process can be broke down is relatively small. The main unit

operations categories are: material reception, material storage, transport and feed, processing, packaging, and waste disposal. The processing operations can be subdivided in subcategories that vary from one industry sector to other. When they are enumerated, they will allow the more effective and reasonable control measure or set of control measures to be applied in each particular situation to be studied. Some examples of processing production functions or unit operations in the ceramic tiles industry are milling, conformation, drying, glazing, firing and sorting, etc.

At the production principle level, it is possible to choose the type of process to achieve the function (eg., different shaping processes), the motive power (ex. electricity or fossil combustible), and the mode of operation (eg., manual operation, mechanical or automatic). There are hundreds of different production principles to undertake unit operations.

At the production form level the machine, the equipment, or set of equipment that will be used in the process is defined (e.g., the hydraulic press type, if shaping by press is the principle used to achieve the unit operation “conformation”). It is also at this level that the exposure controls are defined (e.g., a local exhaust ventilation (LEV) or a closed cabin).

From an occupational safety & health point of view, the focus on the production functions and principles will allow the less hazardous way to achieve the same production result to be found, the best available techniques to control the hazard can be chosen.

2.5. Risk and exposure assessment

In order to undertake risk assessment, a control banding based method was used: the CB Nanotool, which is a four by four matrix that relates severity parameters on one-axis and probability parameters on the other. The severity parameters consider physicochemical and toxicological properties of both nanomaterial and parent material, including, surface reactivity, particle shape and diameter, solubility, carcinogenicity and mutagenicity. The probability band scores are based on factors affecting the potential exposure to the nanomaterial, namely, the estimated amount of chemical used in one day, dustiness, number of employees with similar exposure, frequency of operation and operation duration. The obtained control bands by risk level can be classified in RL1 – general ventilation, to RL4 – seek specialist advice [18].

For exposure assessment, the usual occupational hygiene method was used, namely the NIOSH 0500 for total dust [19]. This consists of collecting the airborne particles in one filter through filtering workplace air. The samples were personal, thus the filter support was placed in the worker’s breathing area.

3. Results and discussion

3.1. Production process and the design analysis

After the preliminary tests, the planned photocatalytic ceramic tiles production process was defined and the use of already existing equipment in the ceramic production plant was proposed. Then, the first step was to detail the production

process, dividing its functions, principles and forms (see Table 1). This work was performed during the project meetings, by getting contributions from all the project's team members.

The production process is similar to the usual ceramic tiles production process. The most relevant unit operations in the process are those related to the processing of raw materials and surface coating.

During a project meeting when contributions from team members were obtained, it was possible to define alternative production principles and forms for the production process. The possible options were the automation of the sack emptying operation, ultrasound agitation for raw materials mixing, and a few non-spraying techniques to apply the TiO₂ aqueous suspension in the ceramic tiles (ex. roll printing, serigraphy or ink-jet). This information is presented in Table 2.

Beside the possible changes in the process itself, other possible action that has a positive impact in the emission and exposure scenarios, contemplated during the design analysis group discussions, was the acquisition of pre-prepared slurry. This would eliminate several unit operations, as the product would arrive at the facilities in liquid form. In particular, pouring raw materials (sack emptying) will be eliminated, which is a dusty operation in the production process.

Considering the bow-tie model together with the design analysis, it was possible to identify the emission scenarios and the barriers for each production function, and related principles and forms. The scenarios and barriers are defined for the normal functioning situations, process disturbances, facilities cleaning and equipment maintenance (Table 2). The identification of the possible emission scenarios and emission barriers was based on the knowledge of the processes and related engineering risk control measures.

It is possible to see that changing the production principle in the pouring raw materials function from the manual operation to the automatic operation will make it possible to introduce a barrier, a closed cabinet with LEV, in the emission scenario. Moreover, considering the acquisition of pre-prepared slurry, the emission scenario is eliminated.

Comparing the possible production principles for the surface coating, once again it is possible to eliminate the dust release emission scenario by choosing a non-spraying technique instead of the air-less spraying (or another spraying technique) to apply the TiO₂ on the ceramic tile surface.

3.2. Pilot-test

During the project, a pilot-test was performed, allowing part of the production process operations and tasks to be simulated. Previous to pouring raw materials, one additional operation was considered, weighing TiO₂. To undertake a risk assessment, 4 different tasks were considered: Task 1- Titanium dioxide weighing; Task 2- Pouring titanium dioxide; Task 3- Mixing slurry; Task 4- Surface coating.

Another relevant question was the use of fine TiO₂, instead of nano-sized form, which resulted from the fact that the photocatalytic properties were optimized with that material.

The risk assessment of the unit operations was performed with the CB Nanotool, which considered the possible use of nano-sized TiO₂. The severity factors are presented in Table 3.

Table 1.
Production functions, production principles and production forms for the photocatalytic ceramic tiles production process

Production Function	Production Principle	Production Form	Description
Reception of raw materials	Mechanical, discontinuous transport	Palletized bags, forklift	The nano-TiO ₂ is received in paper bags, which are packed in wood pallets. The pallets are handled with a forklift and/or an electric pallet-truck.
Storage of raw materials	Mechanical, discontinuous transport	Palletized bags, forklift	
Transport of raw materials	Mechanical, discontinuous transport	Palletized bags, forklift	
Pouring raw materials (sack emptying)	Manual operation	Emptying bags to a container	The TiO ₂ powder is poured into water and is dispersed in the liquid with a column stirrer to obtain homogenized slurry. When the suspension is homogeneous it is milled in a micro-balls mill in order to de-aggregate. Finally, the slurry is sieved.
Mixing raw materials	Mechanical stirring	Column stirrer and micro-ball mill, container	
Surface coating	Spraying, automatic	Air-less spraying, booth disk glazing, booth spray-gun, booth	The slurry is applied in the already fired ceramic pieces by spray technic.
Transport of materials	Mechanical, automatic	Parallel belt line, ceramic tiles loading and unloading machines, storage boxes	The coated tiles are transported over two parallel trapezoidal belts on the glazing line. At the end of the line the tiles are loaded on a "ceramic tiles box" for storage before firing. At the kiln, the tiles are unloaded from the box and transported to the kiln entrance over parallel trapezoidal belts and/or roller conveyor.
Processing - firing	Thermal, automatic	Roll kiln	After the coating, the pieces are fired (2nd fire) at a temperature of around 950 °C in a continuous roll kiln.
Sorting	Manual, mechanical automatic transport	Ceramic tiles sorting line	The fired pieces are sorted (defects on the surface and body of the pieces are checked for) and packed in cardboard boxes.
Packaging	Mechanical, automatic	Ceramic tiles packaging line	

Source: The authors

In Table 4 presents the exposure factors considered for the different tasks assessed.

Table 5 presents the CB Nanotool assessment results are.

Table 2.

Emission scenarios and related barriers related to possible options for the production principle

Production function	Production principle	Normally functioning		Process disturbances		Cleaning		Maintenance	
		Emission scenario	Emission barrier	Emission scenario	Emission barrier	Emission scenario	Emission barrier	Emission scenario	Emission barrier
Raw materials reception, storing and transport	Mechanical, discontinuous transport			Damaged bags, powder spills	Metal containers	Cleaning powder spills	Vacuum-cleaner		
Pouring raw materials	Manual operation	Dust release		Powder spills		Cleaning powder spills		Intervention for dirty equipment	
	Automatic process	Dust release	Closed cabinet	Powder spills	Closed cabinet	Cleaning powder spills	Vacuum-cleaner	Intervention for dirty equipment	
Pre-prepared slurry				Slurry spills	Closed containers	Cleaning dried slurry spills			
Mixing raw materials	Mechanical stirring			Slurry spills		Cleaning dried slurry spills		Intervention for dirty equipment	
	Ultrasound agitation			Slurry spills		Cleaning dried slurry spills		Intervention for dirty equipment	
Surface coating	Spraying, automatic	Spraying (aerosol release)	Closed cabin with LEV	Slurry spills, spray gun clog		Cleaning dried spills		Intervention for dirty equipment	
	All non-spraying techniques, automatic			Slurry spills		Cleaning dried spills		Intervention for dirty equipment	
Material transport	Mechanical, automatic			Tiles jam in line or loading /unloading machines		Removing jammed material			

Source: The authors

Table 3.

CB Nanotool Severity band factors

Hazard Factor	Answer
Parent material hazard	
OEL ($\mu\text{g}/\text{m}^3$)	2400
carcinogen?	yes
reproductive hazard?	no
mutagen?	no
dermal hazard?	no
asthmagen?	no
Nanoscale material hazard	
Surface reactivity	unknown
Particle shape	spherical
Particle diameter (nm)	>40
Solubility	insoluble
carcinogen?	yes
reproductive hazard?	unknown
mutagen?	unknown
dermal hazard?	unknown
asthmagen?	no

Source: The authors

During the pilot-test, the airborne particles concentration was measured using the NIOSH 0500 method in order to have a perception of the worker's exposure to TiO_2 particles during operations. Considering task durations and the workers present in the workplace, it was decided to sample

during the TiO_2 aqueous suspension, including weighing raw materials, pouring raw materials and mixing, and performing two personal samplings on both the workers operating the glazing line (surface coating and transport of materials). Table 6 presents the results of airborne sampling.

The sampling time corresponds to the whole working time. In the first attempt to produce the ceramic tiles, several disturbances occurred and the results should be considered to only represent the conditions of the test. They could not be considered as representing future exposure during industrial

Table 4.

CB Nanotool probability band factors

Probability factor	Answer			
	Task 1	Task 2	Task 3	Task 4
Estimated amount of chemical used in one day (mg)	10^6	10^6	10^6	10^6
Dustiness	Medium	High	Low	High
Number of Employees with Similar Exposure	1 - 5	1 - 5	1 - 5	1 - 5
Frequency of Operation (annual)	Monthly	Monthly	Monthly	Monthly
Operation Duration (hours per shift)	< 30 min	< 30 min	< 30 min	1 - 4 h

Source: The authors

Table 5.
Pilot-test risk assessment using CB Nanotool

Task	Severity band	Probability band	Overall risk band	Control required
1- Titanium dioxide weighing	Medium	Less Likely	RL1	General ventilation
2- Pouring titanium dioxide	Medium	Likely	RL2	Fume hood or local exhaust ventilation
3- Mixing slurry	Medium	Less Likely	RL1	General ventilation
4- Surface coating	Medium	Likely	RL2	Fume hood or local exhaust ventilation

Source: The authors

Table 6.
Airborne particles concentration during pilot-test tasks

Tasks	Sampling time (min)	Concentration (mg/m ³)
Worker during raw material weighing and slurry preparation	33	1.1 ¹
Surface coating in glazing line – Worker 1	160	0.9
Surface coating in glazing line – Worker 2	150	1.5

Source: The authors

production of this type of ceramic tiles, but they could give a rough estimation.

3.3. Discussion

The SYDAPP creates a cooperative environment between process engineers, safety practitioners and other people involved in the development of the process, and facilitating the communication and understanding inside the multidisciplinary team. With this approach it is possible to truly involve the designers and engineers in the occupational risk management.

The production functions and production principles are crucial to design solutions since emissions are directly related to the production functions applied. These functions will limit the number of possible principles, and consequently the number of forms. The actual emission that results in exposure always becomes visible in the production form. Conventional occupational hygiene control measures, such as LEV, enclosure, etc. will act on the production form level.

However, when the emission (and the related exposure) is too excessive, or the contaminants are too dangerous, (re)design approaches will be the only option left to reduce or eliminate emissions (apart from cancelling the whole production). (Re) design consists of changing production-principles under an unchanged production function, or changing or eliminating production functions. This last option is very effective, because the corresponding principles and forms will also be eliminated. Using pre-mixed slurries instead of mixing powdered raw materials is an example in which all functions related to raw materials processing are

eliminated. When a company introduces these changes, it is substantially reducing the sources of emission and exposure at the initial phase of the production process. Obviously, other companies will need to perform these production-functions, but when volumes are big enough, these firms can also modify their production methods, for example, by changing their mode of operation from manual to automatic.

Accordingly, the use of the supply chain with OSH purposes is one question raised by the SYDAPP. The design analysis performed along the supply-chain helps to identify opportunities to transfer higher risk operations to facilities that are prepared to address it. This allows others to focus on the core process operations, which will ultimately result in safer workplaces, by implementing cost-effective solutions. This approach is only acceptable if the risks are transferred to adequate facilities, not to less controlled subcontractors.

Both the CB Nanotool risk assessment and the airborne particles sampling pointed to potential risk to workers during the pilot-test, considering the possible use of nano-sized TiO₂. It is clear that the pilot-test conditions do not exactly replicate the future production conditions but could help to better understand the main emission and exposure scenarios. By replacing nano-TiO₂ by fine-TiO₂ it is possible to reduce the risk for workers. Based on the existing knowledge of the TiO₂ toxicological properties, it is clear that its nano form is more hazardous than the fine-TiO₂ [4]. Furthermore, the toxicological assays performed with nano-TiO₂ reveal potential effects to health resulting from the possible translocation of the nanoparticles in the human body and also from the capability of cell internalization. Considering the bow-tie model, acting on the hazard itself is an advantageous strategy to deal with the workplace risks as this takes place prior to the emission and, of course, the worker's exposure. The results obtained from the airborne particles sampling during the pilot-test show that the exposure to TiO₂ airborne particles is below the proposed limit value of 2.4 mg/m³, even when considering that all the airborne particles were TiO₂.

In the tests performed during the Selfclean Project, the medium size TiO₂ particles was in the 150-200 nm range, while the nano-sized TiO₂ particles have diameters below 100 nm. According, to the International Commission on Radiological Protection's (ICRP) respiratory tract deposition model for particles, quoted by the International Organization for Standardization, it is evident that the probability of the particles with sizes from 150 nm to 200 nm depositing in all respiratory tracts is lower than particles smaller than 100 nm [20].

Considering the lack of knowledge and the potential for harm of the different types of nano-objects, and the uncertainties related to risk and exposure assessment [21], the safety-by-design approaches become relevant. What has previously been learnt from the safety science field could help by defining ways to deal with potentially high-risk production processes. The inherently safer process concept developed in the late 1970's, which focuses on the avoidance or reduction of the hazard at source [22,23] is adaptable to the nanotechnologies field. The SYDAPP allows the project team to identify the unit operations with a lower emission potential.

¹ Result below quantification limit. The uncertainty is higher compared with the other results.

4. Conclusions

The use of the SYDAPP helps to find solutions to reduce the workers' exposure during their work with engineered nano-objects. As shown in the case presented in this model, it seems that there is an advantage to be had in applying it in a development project, or in other words, during the project phase and before the final process design is set.

With this approach, it was possible to generate emission scenarios resulting from the photocatalytic ceramic tiles production process operations. The bow-tie was a helpful concept model to achieve this.

Following identification of the emission scenarios, it was also possible to define emission reduction barriers. In the particular case of the production of photocatalytic ceramic tiles, it was possible to identify opportunities to reduce nanoparticle emission.

Risk management during the project phase allows safer production processes, changing materials, methods or equipment to be developed, the result being an inherently safer production process.

Acknowledgments

The research was part of the SELF CLEAN – Self-cleaning ceramic surfaces Project, funded by QREN – Technological R&D Incentives System – Co-operation projects, Project No. 21533. The authors would like to thank their project partners for their co-operation.

References

- [1] Chen, J. and Poon, C., Photocatalytic construction and building materials: From fundamentals to applications. *Building and Environment*, 44(9), pp. 1899-1906, 2009. DOI: 10.1016/j.buildenv.2009.01.002
- [2] Creutzenberg, O., Toxic effects of various modifications of a nanoparticle following inhalation, Germany, Federal Institute for Occupational Safety and Health, 2013, 405 P.
- [3] IARC, IARC Monographs on the evaluation of carcinogenic risks to humans, Vol. 93, Carbon black, titanium dioxide, and talc, Lyon: International Agency for Research on Cancer, 2006.
- [4] NIOSH, Current Intelligence Bulletin 63. Occupational Exposure to Titanium Dioxide, Cincinnati, OH, DHHS (NIOSH), 2011, 140 P.
- [5] Schulte, P., Geraci, C., Hodson, L., Zumwalde, R., Castranova, V., Kuempel, E., Methner, M.M., Hoover, M. and Murashov, V., Nanotechnologies and nanomaterials in the occupational setting, *Italian Journal of Occupational and Environmental Hygiene*, 1(2), pp. 63-68, 2010.
- [6] Amyotte, P.R., Are classical process safety concepts relevant to nanotechnology applications?. *Journal of Physics: Conference Series*, 304(012071), 2011. DOI: 10.1088/1742-6596/304/1/012071
- [7] Fleury, D., Bomfim, S. Metz, J.A.S., Bouillard, J.X. and Brignon, J.-M., Nanoparticle risk management and cost evaluation: A general framework, *Journal of Physics: Conference Series*, 304(012084), 2011.
- [8] Morose, G., The 5 principles of Design for Safer Nanotechnology, *Journal of Cleaner Production*, 18(3), pp. 285-289, 2010. DOI: 10.1016/j.jclepro.2009.10.001
- [9] Swuste, P. and Zalk, D.M., Risk management and nanomaterials, in Govil, J.N., Navani, N.K. and Sinha, S., Eds. *Nanotechnology Fundamentals and Applications*, Vol. 1, Houston, Studium Press LLC, 2013, pp. 155-173.
- [10] Swuste, P., *Occupational hazards, risks and solutions*, Delft: Delft Univ. Press. 1996.
- [11] Stoop, J., Scenarios in the design process. *Applied Ergonomics*, 21(4), pp. 304-310, 1990. DOI: 10.1016/0003-6870(90)90201-8
- [12] Schupp, B., Hale, A., Pasman, H., Lemkovitz, S. and Goossens, L., Design support for the systematic integration of risk reduction into early chemical process design. *Safety Science*, 44(1), pp. 37-54, 2006. DOI: 10.1016/j.ssci.2005.09.002
- [13] Hale, A., Kirwan, B. and Kjellen, U., Safe by design: Where are we now?. *Safety Science*, 45(1-2), pp. 305-327, 2007. DOI: 10.1016/j.ssci.2006.08.007
- [14] Visser, J., Developments in HSE management in oil and gas exploration and production, in Hale A. and Baram, M., Eds., *Safety Management: The challenge of change*, 1st Ed., Amsterdam: Pergamon, 1998, pp. 43-65.
- [15] Silva, F., Arezes, P. and Swuste, P., Risk assessment and control in engineered nanoparticles occupational exposure, in Arezes, P. et al, Eds. *Occupational Safety and Hygiene*, 1st Ed., Guimarães: CRC Press, 2013, pp. 197-202. DOI: 10.1201/b14391-41
- [16] Fleury, D., Bomfim, J.A.S., Vignes, A., Girard, C., Metz, S., Muñoz, F., R'Mili, B., Ustache, A., Guiot, A. and Bouillard, J.X., Identification of the main exposure scenarios in the production of CNT-polymer nanocomposites by melt-moulding process. *Journal of Cleaner Production*, 53, pp. 22-36, 2013. DOI: 10.1016/j.jclepro.2011.11.009
- [17] Guldenmund, F., Hale, A., Goossens, L., Betten, J. and Duijm, N.J., The development of an audit technique to assess the quality of safety barrier management. *Journal of hazardous materials*, 130(3), pp. 234-241, 2006. DOI: 10.1016/j.jhazmat.2005.07.011
- [18] Zalk, D.M., Paik, S.Y. and Swuste, P., Evaluating the control banding Nano tool: A qualitative risk assessment method for controlling nanoparticle exposures. *Journal of Nanoparticle Research*, 11(7), pp. 1685-1704, 2009. DOI: 10.1007/s11051-009-9678-y
- [19] National Institute for Occupational Safety and Health, NIOSH Manual of Analytical Methods, 4th Ed. Cincinnati, OH: NIOSH, 1994.
- [20] Technical Committee ISO/TC 146, TECHNICAL REPORT ISO/TR 27628 Workplace atmospheres — Ultrafine, nanoparticle and nano-structured aerosols — Inhalation exposure characterization and assessment, Geneva, ISO, 2007, 42 P.
- [21] IUTA, BAuA, BG RCI, VCI, IFA, & TUD, Tiered Approach to an Exposure Measurement and Assessment of Nanoscale Aerosols Released from Engineered Nanomaterials in Workplace Operations, Germany, 2011.
- [22] Kletz, T.A., Inherently safer plants. *Plant/Operations Progress*, 4(3), pp. 164-167, 1985. DOI: 10.1002/prsb.720040311
- [23] Mansfield, D. and Poulter, L., *Improving Inherent Safety*, HSE Books, 1996.

F.A.C. Silva, is an Occupational Hygiene Practitioner in the Portuguese Technological Centre for Ceramic and Glass. He undertakes occupational exposure assessments in chemical and physical agents in industry as well as other activities as a trainer and ISO 17025 auditor. He graduated in Ceramic and Glass Engineering in 1986 and has a MSc. in Human Engineering. He received a PhD in Industrial and Systems Engineering from University of Minho, Portugal. He is author and co-author of several safety and hygiene manuals and other technical publications. He is also a member of national and international standardization procedures, including European CEN/TC 352 Nanotechnologies. ORCID: 0000-0003-4234-5640

P.M. Arezes, received his PhD in Industrial and Systems Engineering from University of Minho, Portugal. He is a full professor in Ergonomics and Human Factors at the School of Engineering of the University of Minho. He is also a visiting fellow at MIT's AgeLab in the USA. He leads the Human Engineering research group and he is also the coordinator of the Engineering Design and Advanced Manufacturing (EDAM) area of the MIT Portugal Program at University of Minho, and the chair of the steering board of the "Leaders for Technical Industries (LTI)" PhD program at University of Minho. He is the (co)author of more than 60 papers published in peer-reviewed international journals, as well as a member of the editorial board of more than 15 international scientific journals. ORCID: 0000-0001-9421-9123

P.H.J.J. Swuste, is an associate professor of the Safety Science Group of the Delft University of Technology, The Netherlands. He finished his PhD. thesis 'Occupational Hazards and Solutions' in 1996. From 1980 onwards he has worked at the Safety Science Group and has been engaged in research concerning risk assessments of various occupational and process related hazards, and quality assessments of risk management systems of various industries. Examples are the international steel industry, the international rubber industry, the national transport sector, the international asbestos industry, and the international process industry. He publishes with co-authors very frequently on these topics, both in the scientific, international and national press, and in the professional press and media. He organized the postgraduate Master's course 'Management of Safety Health and Environment' with Andrew Hale from 1994-2008.

ORCID: 0000-0003-3702-1625



UNIVERSIDAD NACIONAL DE COLOMBIA

SEDE MEDELLÍN
FACULTAD DE MINAS

Área Curricular de Medio Ambiente

Oferta de Posgrados

Especialización en Aprovechamiento de
Recursos Hidráulicos

Especialización en Gestión Ambiental

Maestría en Ingeniería Recursos Hidráulicos

Maestría en Medio Ambiente y Desarrollo

Doctorado en Ingeniería - Recursos Hidráulicos

Doctorado Interinstitucional en Ciencias del Mar

Mayor información:

E-mail: acia_med@unal.edu.co
Teléfono: (57-4) 425 5105

Prevalence and symptomatology of musculoskeletal problems reported by home care service workers caring for the elderly

Dina Chagas ^a

^a Higher Institute of Education and Sciences, Lisbon, Portugal. dina.chagas2003@gmail.com

Received: November 30th, 2015. Received in revised form: April 07th, 2016. Accepted: April 12th, 2016.

Abstract

Work-related musculoskeletal disorders (WRMSDs) develop gradually and result from professional risk factors such as the repetition of tasks, emotional overburden, inadequate posture, excessive strain, and individual and organizational/psychosocial risks. The aim of this study was to identify prevalence and symptomatology of musculoskeletal problems reported by home care service workers caring for the elderly. The Nordic Musculoskeletal Questionnaire was used for the purpose. The sample contained 86 workers. The results show a high prevalence of musculoskeletal symptoms in different anatomical areas, affecting in particular the upper limbs and the lumbar, dorsal and cervical areas. There are also some important associations between the dorsal area and the lifting of patients weighing above 90 kg. It may be concluded that these workers run a considerable risk of acquiring musculoskeletal injuries as a result of their working conditions and from the ways in which their activities are currently performed.

Keywords: work-related musculoskeletal disorders; prevalence; symptomatology; home care services

Prevalencia y sintomatología músculo-esquelética auto-referida por los trabajadores del servicio de apoyo domiciliario en la prestación de atención a personas mayores

Resumen

Las lesiones músculo-esqueléticas relacionadas con el trabajo se desarrollan gradualmente y resultan de los factores de riesgo profesionales como la repetitividad de las tareas, sobrecarga emocional, postura inadecuada, aplicación de fuerza excesiva y de factores de riesgo individual y organizacionales/psicosociales. El objetivo de este estudio es identificar la prevalencia y sintomatología músculo-esquelética auto-referida por los trabajadores del servicio de apoyo domiciliario en la prestación de atención a personas mayores a través de una adaptación del Cuestionario Nórdico Músculo-esquelético. La muestra fue constituida por 86 trabajadores. Los resultados evidencian una prevalencia elevada de síntomas músculo-esqueléticos en diferentes zonas anatómicas, alcanzando en particular los miembros superiores, zona lumbar, zona dorsal y la zona cervical. También se encuentran asociaciones significativas entre la zona dorsal y levantar utentes con pesos superiores a 90 Kg. Podemos concluir que estos trabajadores tienen un considerable riesgo de adquirir lesiones músculo-esqueléticas resultantes de las condiciones de trabajo y de la organización en la que se lleva a cabo la actividad.

Palabras Clave: lesiones músculo-esqueléticas relacionadas con el trabajo; prevalencia; sintomatología; atención en domicilio

1. Introduction

The symptoms of work-related musculoskeletal disorders (WRMSDs) include chronic pain that can affect one or more body area during an activity [1]. WRMSDs are the

commonest occupational pathologies in the European Union (EU), affecting workers in all sectors.

In Portugal, the body responsible for health care, the *Direção-Geral da Saúde*, reports that these injuries are the result of high risk occupational actions, such as repetition,

How to cite: Chagas, D., Prevalence and symptomatology of musculoskeletal problems reported by home care service workers caring for the elderly. DYNA 83(197), pp. 17-21, 2016.

overload and posture. They are chronic pain syndromes that occur while certain activities are performed, and are therefore designated “work-related” [1].

WRMSDs are extremely significant in the EU, as they affect more than 40 million workers and are responsible for around 50% of sick leave for periods of three or more days and 60% of cases of permanent incapacity for work [2].

The body areas most affected by WRMSDs are the lumbar region, neck, shoulder, arm and hand, although recently, the lower limbs have also received some attention [3].

The rate of disorders for workers providing direct care services to the elderly is 61%. They are caused by excess effort [4] and by daily tasks, including helping in personal hygiene tasks, changing clothing and bed linen, moving elderly clients, etc. [5]. These tasks carry a high risk of causing musculoskeletal WRMSDs whose development is exacerbated physically and biomechanically (repetition of movements, inadequate posture and excess effort), organizational and psychosocial factors (demanding work, repetitive movements, lack of support) as well as individual and personal factors (previous medical history, physical abilities, age and obesity) [6].

The continuous lifting of weights, the movement of heavy items (as most clients are dependent), as well as the imposed adoption of incorrect postures are all hazards that are present in the daily lives of workers, which lead to the appearance of critical and sometimes chronic injuries [7]. In order to be able to respond to the demanding tasks asked of them, workers frequently adopt incorrect postures an aspect that is exacerbated by conditions that are not suitable for the use of available equipment, given the anthropometric status of the workers [8].

WRMSDs develop, and are further aggravated, by the way daily tasks are organized and by psychosocial risk factors [9].

Physical and psychosocial factors are generally identified by studies undertaken to determine the risk factors for WRMSDs, allowing the relationship between them and of symptoms and/or illnesses to be analyzed [10].

A study undertaken by the European Survey on Working Conditions (2005), cited by the European Agency for Safety and Health at Work, identifies standing postures or uncomfortable positions, repetitive work and the manual handling of loads as the most significant risk factors facing workers [11]. The expression “risk factor” generally refers to any attribute, experience or behavior that increases the likelihood of developing a disease or disorder.

Musculoskeletal disorders manifest themselves clinically in many different ways that frequently are not considered. They often comprise an ample set of signs and symptoms, from simple localized fatigue to intense pain, passing through situations of discomfort and mild pain [12].

We may assert that WRMSDs that have developed or that worsen in the workplace, as a result of continued exposure to risk factors - in particular, inadequate posture, repetitive movements, overload, excessive effort and so on, all of which affect different anatomical regions.

2. Methodology

The sample consisted of employees of a home care service that provides care for the elderly in various

institutions of the central region of Portugal. In total, 90 survey questionnaires were distributed to workers from various institutions and 86 were returned in all, which constituted the final sample.

A statistical analysis was conducted using IBM® SPSS® Statistics 21.0 software. This took into account the nature of the variables involved in the study, based on a previous exploratory and analytical study of the data, and using the most appropriate statistical techniques.

First, the descriptive statistical analysis (mean, standard deviation, simple frequency, percentage) will be described, before turning to look at the relationship between the principal variables, symptom recognition and other elements that contribute to the presence of grievances using the Chi-Square statistic. The significance level assumed in the statistical test was $p < 0.05$.

The questionnaire was adapted from the Nordic Musculoskeletal Questionnaire developed by Kuorinka et al. (1987), using a version adapted by Uva and Serranheira [13], to which several questions related to home care service and caring for the elderly were added. The delivery and collection of information using the questionnaire was conducted during July 2014.

The questionnaire was divided into two parts. The first was intended to identify the main socio-demographic and professional characteristics of each worker. Among other things it requested data on age, gender, anthropometrics and other elements related to the organization of tasks performed by the worker. It also included information on the type and characteristics of work-related complaints and musculoskeletal symptoms of respondents. Complaints referred to nine body regions (neck, shoulders, dorsal region, elbows, lower back, wrists/hands, thighs, knees and ankles/feet) which were chosen to assess respondents' self-reported pain/discomfort experienced over the previous 12 months and during the last seven days, and to ascertain whether, during this period of time, they had had to take time off from work. Each question had a diagram showing the body location to be evaluated.

Employee self-assessment was based on a scale that was used to register the intensity of pain, discomfort, swelling and fatigue. The scale had four levels: 1 = Light, 2 = Moderate, 3 = Intense, 4 = very intense. A four-level scale was also used to record the frequency of complaints and symptoms: 1 = Once a year; 2 = Twice or three times a year; 3 = Four to six times a year; and 4 = More than six times a year. To classify the workplace activities that were associated with symptoms and complaints a scale similar to a four point Likert scale was used, where: 1 = Not related to symptoms; 2 = Minor relation to symptoms; 3 = Strong relation to symptoms; 4 = Completely related to symptoms. When respondents did not know how to answer they could opt for choice 8 = “Don't know”, and if they did not wish to respond at all, then option 9 = “I do not wish to answer” was available.

The second part of the questionnaire was intended to identify aspects considered relevant to the health of the works (psychosocial factors and auxiliary equipment used to increase client mobility).

3. Results

The study surveyed 90 employees of the home care

service provider for the elderly. The response rate was 96%. In terms of demographic and anthropometric variables, all the respondents were female (100%). Ages ranged between 20 and 60, though the largest cohort (accounting for 37.2% of the total sample) was aged between 41 and 50 years old. The most affected upper limb was the right arm (58.1%). Most individuals were married (61.6%) and the commonest education level was the 9th grade (59.3%).

The average height of the respondents was 161cm, with a maximum height of 180 cm and a minimum of 149 cm (standard deviation 0.06075). The average weight of participants was 68.6 kg, with a maximum weight of 110 kg and a minimum of 45 kg (standard deviation 10.727 kg). The majority were right handed (58.1%), while 5.8% were left handed and 36.1% ambidextrous.

It was found that the respondents work an average of 38.8 hours a week, with a minimum of 35 hours and a maximum of 42 hours (standard deviation 1.45 hours). Average length of service was 8 years, with a minimum of one year and a maximum of 19 years (standard deviation 4.666 years). 70.9% worked fixed hours, while 29.1% worked shifts.

In terms of pain, discomfort, swelling and fatigue- it was found that the most affected areas during the previous 12 months had been the lumbar region (86.0%), the dorsal area (82.6%), the neck (66.3%), the right shoulder (50.0%), the left shoulder (14.0%), the right wrist/hand (45.3%) the left wrist/hand (5.8%), the right knee (30.2%), the left knee (5.8%), the right elbow (24.4%), the left elbow (1.2%), the right ankle/foot (19.8%), the left ankle/foot (7.0%), the right thigh (15.1%) and the left thigh (1.2%).

During the previous seven days symptoms were felt as follows: the dorsal area (43.0%), the lower back (37.2%), the wrist/right and left hand (both 26.7%), the right and left shoulder (both 17.4%), the neck (16.3%), the ankle/right and left foot (both 12.8%), the right and left knee (both 11.6%), the right and left elbow (both 8.1%) and the left and right thigh (7.0%).

Values were also recorded relating to absence from work with the following symptoms, shoulders (8.1%), neck, lower back, ankles/feet (7.0%), dorsal area, wrist/hands (4.7%) and knees (1.2%).

In terms of intensity of discomfort (pain, discomfort, swelling and fatigue) the highest level of discomfort recorded (30.2%: "very intense") was reported for the lumbar area, against 24.4% for the dorsal area. The lumbar region (44.2%) and the dorsal area (34.9%) also scored highest for "intense". A value of 20.9% was recorded for "moderate" discomfort in the neck and 4.7% of respondents recorded "light" levels for the dorsal area.

Regarding frequency (number of times per year) 47.7% reported experiencing lumbar area pain/discomfort "more than six times" in the year, while 34.9% reported pain/discomfort "4-6 times".

In terms of the relationship between work activity and self-reported symptoms mentioned above, it was found that workers generally reported that activities such as lifting and moving clients weighing between 70 kg to 90 kg (and above) was "strongly related to symptoms". They considered arm movements above shoulder height and repeated arm movements to be "completely related to symptoms".

87.2% of the sample said that they felt psychologically

tired, in particular following the afternoon shift (75.6%). Respondents who reported feeling physical fatigue (95.3%), said they felt tired when a shift was over, in particular in the case of the afternoon shift (82.6%).

Regarding the repetition of tasks, 95.3% considered the lifting of clients to be the task that is repeated most frequently, while 52.3% indicated that they had difficulty moving while performing certain tasks for the client. 73.3% considered that bed height was not ideal.

As for the height of the boards used in the bath tubs, 62.8% answered "no" when asked if they were at a comfortable height, while 60.5% mentioned that they did not use specialized equipment to perform tasks, including the transport and movement of clients.

3.1. Statistical analysis of associations

The significant statistical associations between the work activities of home care workers caring for the elderly and musculoskeletal complaints were evaluated using the Pearson correlation test.

The strongest association between symptoms felt in the neck, was for the variable "arms above shoulder height" ($r = 0.432$, $p = 0.003$). This was categorized as having a "Strong relation to the symptoms" and proved to be statistically significant. In the same anatomical area other variables were also considered significant, such as "body tilt" ($r = 0.369$; $p = 0.003$) and "body twist" ($r = 0.395$; $p = 0.003$).

For the shoulder region "repetitive arm movement" ($r = 0.265$; $p = 0.000$) and "arms above shoulder height" ($r = 0.354$; $p = 0.000$) were identified as contributing most to symptoms, being considered "highly related to the symptoms".

For the wrist/hand area "repetitive hand/finger movements" ($r = 0.400$; $p = 0.003$), were identified as the element that contributed most to complaints, being identified as "totally related to symptoms". "Applying effort with hands and fingers" ($r = 0.448$; $p = 0.000$), was also regarded as "totally related to symptoms".

Within the dorsal column there were statistically significant connections between the variables "lifting clients weighing more than 90 kg" ($r = 0.366$; $p = 0.000$), which contributed most to the association, being considered "totally related to the symptoms". In this anatomical area the variable "moving clients weighing between 70 and 90 kg" ($r = 0.002$; $p = 0.000$) was also considered significant.

A statistically significant association was also found with the variable "length of time working for the institution" ($r = 0.868$; $p = 0.011$) in which the element that contributed most to reported symptoms was seniority of service: on average around eight years.

For the lumbar region, the variable "tilt the body" ($r = 0.077$; $p = 0.000$) was found to be statistically significant. There is a possibility of musculoskeletal injuries occurring in certain anatomical areas that related to different kinds of position or movement adopted in the course of the working day [14].

4. Discussion

The data presented here refers to a group of adult females,

most of whom have received the minimum level of education required by the law.

The results for symptoms reported by workers during the previous 12 months were as follows: lower back (86.0%), dorsal areas (82.6%), neck (66.3%), shoulders (64.0%) and wrist/ hands (61.0%). Symptoms reported for the previous 12 months were for the same body areas as those reported for the previous 7 days.

The evidence suggests that the symptoms felt in the neck and shoulders were caused by the adoption of positions required by work activities carried out by the respondents. That is: cervical flexion and elevation of the upper limbs, often above shoulder height. It is important to note that these symptoms are related to the duration and the rhythm of work performed.

As for work-related absence caused by WRMSDs, the commonest cause was pain/discomfort in the shoulder region (8.1%) and “very intense” pain in the lumbar region (30.2%).

The prevention of occupational risk and the eradication of musculoskeletal disorders might lead to reduced levels of absence from work [15]. In this connection, the area that led to the greatest number of days off due to pain/discomfort was the lumbar area (47.7%) with “more than 6 times”.

Most employees associated activities such as lifting and moving clients weighing between 70 kg to 90 kg and above, with the classification “completely related to symptoms”, demonstrating that they have to handle very heavy weights. Added to this, the majority (60.5%), mentioned not using proper equipment to perform tasks including when transporting or moving clients (Fig. 1). Respondents stated that, in general, clients to whom they provide home care assistance are dependent, and as a result are not in a position to help their carers when they are providing services.

The use of excessive effort during frequent lifting, handling and moving of clients, incorrect posture and vertebral column crunches can cause health problems for workers, including fractures, back pain and varicose veins [8]. Such causal factors are related to ergonomic factors [16].

Respondents characterized arm activity above shoulder level and repetitive arm movements as having a “strong relation to symptoms”.

The majority of employees (95.3%) considered that the lifting of clients is the most repetitive task, performed daily (Fig. 2). Usually, repetitive movements are principally associated with quick finger or upper limb movements [17].

The results of the survey showed that 52.3% of respondents encountered problems related to movement during the performance of the different tasks required to help their clients. This was because of the arrangement of furniture and the need to maneuver wheelchairs between beds, which need to be moved every time a task is required to be performed.

It was also noted that 73.3% of respondents considered the height of beds not to be ideal, or noted that it was not possible to adjust bed height, a factor that contributes to the adoption of incorrect postures (Fig. 3). Other factors mentioned included that beds tend to be quite close to walls, with the result that workers do not have enough space to perform their tasks properly. In terms of the height of bath tub boards, 62.8% of respondents reported that, as they are not adjustable, a lot of additional effort is required when

bathing clients.

To tackle the problem of musculoskeletal disorders a scheme of risk prevention management should be integrated into organizational management systems [18].

87.2% of respondents said that they feel psychologically tired at the end of their shift - 75.6% referring specifically to the afternoon shift. The psychosocial risks faced by workers include the issue of becoming emotionally overburdened as a result of continual contact with the suffering of clients, with pain and death, shift work, the pace of work, the performance of multiple tasks and the scattered and repetitive nature of their tasks. These factors may lead to headaches, breathing problems, sleep disturbance, depression, anxiety, alcohol-use and isolation [19].

Regarding physical tiredness, 95.3% of respondents reported feeling tired at the end of their shift - 82.6% referring specifically to the afternoon shift.

Several authors have argued that there is a strong link between psychosocial and physical factors associated with the job itself and to the development of WRMSDs [17].



Figure 1. Moving clients.
Source: the author.



Figure 2. Lifting clients.
Source: the author.



Figure 3. Lifting clients up.
Source: the author.

5. Conclusions

This study aimed to identify the prevalence and symptoms of musculoskeletal disorders self-reported by workers and to determine whether or not there is a clear relationship between them and the work activity carried out.

After analyzing the data, it is clear that the assessment of pain/discomfort levels originating from the musculoskeletal system is related to work and the conditions under which it is performed. There is a high prevalence of upper body disorders and a significant connection between musculoskeletal symptoms and working postures involving the raising of the arms above shoulder height and of lifting clients who weigh over 90 kg. This is because the moving of individuals is one of the most commonly performed tasks by employees.

Auxiliary equipment should be used for handling and transporting clients and avoiding disorders, as many workers move heavy weights and have no concept of the maximum load that they should move or lift without help.

It is of paramount importance that more rest breaks should be introduced during shifts, as workers experience an enormous feeling of both physical and psychological tiredness when these end, after which they are required to repeat tasks during their next shift. The level of physical effort, (muscle effort, posture, repetition, etc.) in which workers engage is largely determined by the degree of dependence of their clients, which, as it is high, may cause problems in the upper limbs and cervical area.

The quality of life of workers is affected by their underlying performance, which is in turn related to their daily activities.

Specific training programs on caring for the elderly should be organized. These would contribute to reducing exposure to occupational risk factors and changing the ways that work is performed, in order to allow workers in the sector to improve their understanding of the problem, and prevent future disorders.

References

- [1] Uva, A., Camide, F., Serranheira, F., Miranda, L. e Lopes, M. Guia de orientação para a prevenção das lesões musculoesqueléticas e

- relacionadas com o trabalho: Programa nacional contra as doenças reumáticas. Lisboa: Direcção-Geral da Saúde, 2008.
- [2] Bevan, S., Reducing temporary work absence through early intervention: The case of MSDs in the EU. London: The Work Foundation – fit for work – Europe, 2013.
- [3] Punnett, L. and Wegman, D., Work-related musculoskeletal disorders: The epidemiologic evidence and debate. *Journal of Electromyography and Kinesiology*, 14, pp. 13-23, 2004. DOI: 10.1016/j.jelekin.2003.09.015
- [4] Faucett, J., Kang, T. and Newcomer, R., Personal service assistance: Musculoskeletal disorders and injuries in consumer-directed home care. *American Journal of Industrial Medicine*, 56, pp. 454-468, 2013. DOI: 10.1002/ajim.22133
- [5] Kim, S. and Lee, J., Development of an intervention to prevent work-related musculoskeletal disorders among hospital nurses based on the participatory approach. *Applied Ergonomics*, 41, pp. 454-460, 2010. DOI: 10.1016/j.apergo.2009.09.007
- [6] AESST. Introdução às lesões músculo-esqueléticas. Agência Europeia para a Segurança e saúde no Trabalho. Factsheet 71, 2007.
- [7] Johanning, E., Evaluation and management of occupational low back disorders. *American Journal of Industrial Medicine*, 37(1), pp. 94-111, 2000. DOI: 10.1002/(SICI)1097-0274(200001)37:1<94::AID-AJIM8>3.0.CO;2-X
- [8] Chagas, D., Direct action helpers at home care services: Consequences to the health of workers. In: Arezes et al., Eds. *Occupational Safety and Hygiene III*. London: Taylor & Francis Group, 2015, pp. 17-21. DOI: 10.1201/b18042-6
- [9] HSE – Health and Safety Executive. Upper limb disorders in the workplace. Sudbury: HSE Books, 2002.
- [10] Leino, P., Symptoms of stress predict musculoskeletal disorders. *Journal of Epidemiology and Community Health*, 43, pp. 293-300, 1989. DOI: 10.1136/jech.43.3.293
- [11] Schneider, E. and Irastorza, X., OSH in figures: Work-related musculoskeletal disorders in the EU-facts and figures. Luxembourg: Office for Official Publications of the European Communities, 2010.
- [12] Fonseca, R. e Serranheira, F., Sintomatologia musculoesquelética auto-referida por enfermeiros em meio hospitalar. *Revista Portuguesa de Saúde Pública*, 6, pp. 37-44, 2006.
- [13] Uva, A. e Serranheira, F., Lesões músculo-esqueléticas e trabalho - Alguns métodos de avaliação do risco. [Online], Lisboa, Sociedade Portuguesa de Medicina do Trabalho. Cadernos/Avulso 05. [date of reference January 25th of 2008], Available at: <http://www.spmtrabalho.com/downloads/ca05.pdf>.
- [14] García, M., Lite, A., Camacho, A. y Domingo, R., Análisis de métodos de valoración postural en las herramientas de simulación virtual para la ingeniería de fabricación. *DYNA*, 181, pp. 5-15, 2013.
- [15] Freitas, L., Manual de Segurança e Saúde do Trabalho. 1^a ed. Lisboa: Edições Silabo, 2008.
- [16] Carneiro, P. e Arezes, P., Lesões Músculo-esqueléticas relacionadas com o trabalho (LMERT) no contexto dos riscos psicossociais. In: Neto, Areosa, Arezes (Eds). *Manual sobre Riscos Psicossociais no Trabalho*. Porto: Civeri publishing, pp. 152-171, 2014.
- [17] Formoso, J., Couce, L., Rodríguez, G. and Guerreiro, M., Functions, responsibility, and authority of human resources in the implementation of a security and safety management system at work. *DYNA*, 172, pp. 180-186, 2012.
- [18] Bakker, A., Demerouti, E. and Sanz-Vergel, A., Burnout and work engagement: The JD-R Approach. *Annual Review of Organisational Psychology and Organizational Behaviour*, 1, pp. 389-411, 2014. DOI: 10.1146/annurev-orgpsych-031413-091235

D. Chagas, holds a PhD in Hygiene, Health and Safety at Work from the University of León, Spain. She is an invited professor at the Higher Institute of Education and Sciences, in Portugal. She has authored a book and (co)authored several articles published in magazines and book chapters revised by peers in the area of health and safety at work and work conditions. Her investigation interests are focused on health and safety at work in different contexts and work conditions.
ORCID: 0000-0003-3135-7689

Current state of the art and enduring issues in anthropometric data collection

Sara Bragança ^a, Pedro Arezes ^a, Miguel Carvalho ^b & Susan P. Ashdown ^c

^a Department of Production and Systems, University of Minho, Guimarães, Portugal. saraabraganca@gmail.com, parezes@dps.uminho.pt

^b Department of Textile Engineering, University of Minho, Guimarães, Portugal. migcar@det.uminho.pt

^c Fiber Science & Apparel Design, Cornell University, Ithaca, NY, USA. spa4@cornell.edu

Received: November 30th, 2015. Received in revised form: March 17th, 2016. Accepted: April 12th, 2016.

Abstract

The study of human body size and shape has been a topic of research for a very long time. In the past, anthropometry used traditional measuring techniques to record the dimensions of the human body and reported variance in body dimensions as a function of mean and standard deviation. Nowadays, the study of human body dimensions can be carried out more efficiently using three-dimensional body scanners, which can provide large amounts of anthropometric data more quickly than traditional techniques can. This paper presents a description of the broad range of issues related to the collection of anthropometric data using three-dimensional body scanners, including the different types of technologies available and their implications, the standard scanning process needed for effective data collection, and the possible sources of measurement errors that might affect the reliability and validity of the data collected.

Keywords: anthropometry; body scanners; errors; reliability.

Estado actual de la técnica y cuestiones perdurables en la recogida de datos antropométricos

Resumen

El estudio del tamaño y la forma del cuerpo humano ha sido un tema de investigación durante un tiempo muy largo. En el pasado, la antropometría utilizó técnicas de medición tradicionales para registrar las dimensiones del cuerpo humano y reportó la variación en las dimensiones del cuerpo en función de la media y la desviación estándar. Hoy en día, el estudio de las dimensiones del cuerpo humano se puede llevar a cabo utilizando maneras más eficientes, como los escáneres tridimensionales del cuerpo, que pueden proporcionar grandes cantidades de datos antropométricos más rápidamente que las técnicas tradicionales. En este trabajo se presenta una descripción de la amplia gama de temas relacionados con la recogida de datos antropométricos utilizando escáneres tridimensionales del cuerpo, incluyendo los diferentes tipos de tecnologías disponibles y sus implicaciones, el proceso de digitalización estándar necesario para la captura efectiva de datos, y las posibles fuentes de los errores de medición que podrán afectar la fiabilidad y validez de los datos recogidos.

Palabras clave: la antropometría; escáneres corporales; errores; confiabilidad.

1. Introduction

In traditional anthropometry the determination of human body dimensions can be achieved using a range of devices. Ever since Richer first used calipers in 1890, a standard set of anthropometric instruments has been used [1]. Simple, quick, relatively non-invasive tools include scales (to determine weight), measuring tapes (to measure

circumferences and linear body surface dimensions), anthropometers (to measure height and various transverse widths and depths of the body), spreading calipers (also to measure widths and depths of the body), sliding compasses (to measure short distances, e.g., on the nose, ears or hands), and head spanners (to measure the height of the head) [1]. All of these devices usually require calibration and the measurements taken are only as accurate as the techniques

How to cite: Bragança, S., Arezes, P., Carvalho, M. and Ashdown, S.P., Current state of the art and enduring issues in anthropometric data collection. DYNA 83(197), pp. 22-30, 2016.

used by the person who takes them. Therefore, it is generally necessary to take multiple measurements and to calculate average values. Additionally, there are typically differences between measurements taken by different people, although this can be reduced with consistent training.

With these traditional methods of collecting anthropometric data, the measuring process is time-consuming, expensive and error-prone. Moreover, traditional methods require the person being measured to adopt standardized postures that are prescribed when measurements are taken (and to maintain them during the measurement process). These standard measuring postures, defined in ISO 7250 [2], are based on the studies of several authors such as Kroemer and Kroemer [3], who explain the standard method of measuring a subject in detail. The primary measuring posture is referred to as the “anatomical position”, in which the participant’s body is placed in a defined, straight, upright posture, with the body segments at either 180, 0, or 90 degrees to each other. Participants are required to stand erect; heels together; buttocks, shoulder blades, and the back of head touching a vertical surface; arms vertical, fingers straight. The head is positioned in the Frankfurt plane; with the pupils on the same horizontal level; the right tragon and the lowest point of the right orbit are likewise aligned horizontally. When measurements are taken of a seated subject, the surfaces of the seat and the foot support must be arranged so that the thighs and feet are horizontal and the lower legs vertical. The measurements should be taken in the morning, because the human body tends to decrease in height during the day, and this practice can help remove one source of variation [4].

With the appearance of new ways of acquiring surface data, anthropometry has gained a new way of performing a deeper investigation of human body size and shape. These new digital shape analysis tools make it possible to acquire data on complex geometrical features, such as curvatures or partial volumes. Using these tools to acquire anthropometric data has the potential to be more practical, reliable, fast and -in comparison with traditional anthropometry- less expensive.

The study of the human body as a 3D object using digital capture tools began in 1973 with a light sectioning technique proposed by Lovesey [5]. This was labor intensive, as the interpretation of data was extremely time consuming. That technology evolved into what is now known as a three-dimensional body scanner.

A whole body scanner is an optical 3D measuring system that produces a digital copy of the surface geometry of the human body [6]. In most cases, three-dimensional body scanners capture the visible surface of the body by using optical techniques, in combination with light sensitive devices, and do not require physical contact with the body. The subject being scanned usually wears form-fitting clothing during the process. Despite the fact that there is no need for the measurer to touch the participants’ body, there are still some privacy issues. On the one hand there is more privacy because the body is not touched but, on the other, the recognizable image-capture of the semi-nude body results in sensitive personal images and data that may be stored in insecure circumstances and can potentially be made available across the Internet [7]. Nevertheless, with recent advances in

anthropometry and in digital human-shape reconstruction it is possible to provide a different perspective on the collection of anthropometric measurements.

This paper presents an overview of how anthropometric data is collected using three-dimensional body scanners. The available types of technology are discussed and the standard processes used for scanning test participants are described. In addition, the main causes of error are identified, including some discussion of the landmarking issue.

2. Use of 3D body scanners for collecting anthropometric data

One of the earliest 3D body scanning systems was the Loughborough Anthropometric Shadow Scanner (LASS), a shadow scanning method developed by Loughborough University in the UK [8]. This system was developed and used to digitize the human body, but the data had to be manipulated before body measurements could be derived from the scan. The original shadow data collection methods are different from other conventional structured lighting approaches since they require very little hardware other than a camera, a desk-lamp, a pencil and a checkerboard. LASS was an automated, computerized 3D measurement system based on triangulation, where the subjects stood on a rotating platform that was turned 360° in measured angular increments. Brooke-Wavell et al. [9] compared anthropometric measurements taken using the LASS with measurements acquired using traditional anthropometry and concluded that the results were similar. For women, statistical differences were found between various measurements (neck and chest circumferences, waist width, depth and height), whilst for men a significant difference was only found in the case of measurements of waist depth. These variations were explained as being due to landmarking and to difficulties in making horizontal measurements with a tape measure.

Body scanner hardware and software technologies have developed greatly since the early 1990s. Today, several alternative body scanning systems are available, using a variety of technologies.

2.1. Types of imaging techniques

Currently, several types of imaging techniques are used to create full body images. These imaging technologies, include: (i) 2D silhouette images converted into 3D models, (ii) white light phase-based image capture, (iii) laser-based image capture, and (iv) radio wave linear array image capture [10-12]. More recently, systems have been developed that use infrared light sources.

Body scanning systems normally consist of one or more light sources, one or more vision or capturing devices, software, computer systems and monitor screens to visualize the data capture process [6]. The major scan technologies in use are those employing laser and non-laser light. According to Daanen & Ter Haar [13], in 2013 the various types of technology used in three-dimensional body scanners were:

- **Laser line systems:** A laser line is projected onto the body from all sides and is viewed by cameras that are triangulated at a fixed angle. The advantage of a single line is that it is easily detected by the sensor which can

compute how the projected 2D line is deformed over the 3D surface very accurately. The sequentially captured 3D lines (generally taken at 1 or 2 mm increments) are then merged to form the complete 3D image.

- **Structured light systems:** A structured light system projects a structured light pattern onto the surface of the body from the front and from the back, and a full 3D image is calculated using the deformed pattern. The light pattern may consist of dots, bars, or any other pattern. The advantage of a structured light scanner is the speed with which it is capable of capturing data from the whole body. Structured light scanning is so fast that it can actually be used for 4D scanning: i.e. real time 3D scanning at 200 Hz. This offers opportunities to couple the registration of movements with 3D shape analysis.
- **Multi-view camera system:** A 3D image is acquired from two or more cameras. A stereo-camera records two images at the same time from a different viewpoint. From the content of the two images the depth to the body can be calculated and converted into a dense 3D image in real-time. The advantage of a stereo-camera system is that no laser line or light pattern is transmitted, meaning that environmental light cannot interfere with the pattern. However, using a laser line or patterns enables a higher resolution, more accurate, 3D image to be produced.
- **Millimeter waves:** Both active and passive millimeter wave scanners are available. Active scanners use the reflection patterns of millimeter waves projected onto the body. Passive scanners process the millimeter waves that are emitted by the human skin. Millimeter waves offer the advantage that they pass through most clothing ensembles but not the skin. Thus, the shape of the body can be captured without the client being required to undress. This offers an advantage in terms of time and effort, but may introduce an ethical problem because the private parts of the subjects can be seen. Millimeter wave scanners are currently employed at airports for the detection of metal parts under garments and offer an alternative to low radiation x-ray scanners.

Many studies are available that compare the various types of existing body scanners. Daanen et al., Daanen & Ter Haar and Olds & Honey [6,13-15] discuss the use of 3D whole body scanners in anthropometry as a whole, giving a good overview of the evolution of body scanning technology and the different scanners in use at the time their articles were written. Olds & Honey affirm that scanners using white light are generally faster and cheaper than laser scanners, but can produce lower quality scans, with areas of data missing. Despite concluding that body scanners are expensive, require technical expertise, and cannot measure skin-folds or compressed bone lengths, they agree that they offer the ability to collect greater amounts of data, can extract data when the subjects are no longer present and are able to use it directly in computer-aided design software applications. Mckinnon & Istook [16] compared two scanners available from the company TC2 at the time they were writing (2001), finding that the newer design was an improvement on the older version, as it produced data that replicated information obtained using traditional measurement methods more closely. They then anticipated that the extraction of fast and accurate anthropometric data would be

possible in the future, as is in fact the case, 15 years later.

2.2. Scanning process

3D scanning offers a technique for capturing body dimensions in a fast and reproducible way. However, the position of the subject in the scanning volume is important if reliable data that can be used in an anthropometric database are to be obtained. The postures used for measurements in traditional anthropometry are not suitable for body scanning because they occlude large areas at the axillae or the crotch. Instead, they require a scan posture with abducted arms and legs that affords the image capture devices a view of the inside surfaces of the limbs and torso. Kouchi et al. [17] state that this change to the basic posture may alter some measurements when compared to those acquired using standard anthropometric tools. Although occluded areas are smaller when arms and legs are abducted, these postures can also result in changes to the shape of the shoulders and to body dimensions around the shoulders and hips. In this study the authors concluded that the acromial height remains stable as long as the abduction angle is smaller than 20°, and the biacromial breadth smaller when the abduction angle is greater, approximately, than 5°.

As scanning systems are different from one another in the number and placement of image capture devices, the optimal scanning position may vary from system to system. When the optimal position is determined, it should be described precisely and used for all subjects. ISO Standard 20685 [18] suggests the four postures identified in Fig. 1. For all postures, quiet respiration (normal breathing) should be adopted. The shoulders should be straight without being stiff, and muscles should not be tense.

The positions adopted by subjects during the scanning process should be adapted to the study being conducted. As such, in the literature it is possible to find positions that differ from those indicated in ISO 20685. This is the case presented by Ashdown et al. [19] who argued (in a case in which a laser scanner with 8 paired cameras was stationed at four points equally spaced around the body) that the subjects' feet should be positioned about 30cm apart with the arms abducted from the body. This is a fundamental aspect of a good scan because other positions often result in holes or missing data for some portion of the body or obscure another area (such as under the arms) or for areas where the cameras cannot record data (such as surfaces parallel to the floor). Additionally, surfaces such as hair and dark-textured clothing decrease the quality of the scan by scattering the light and preventing the cameras from capturing a complete set of data points. A study by Tomkinson & Shaw [20] showed that most direct 3D scan measurements of standing posture had good repeatability, except the head and neck postures, whose repeatability was poor, as a result of significant postural errors. In this case they recommend that researchers aim at reducing postural and technical errors by strictly adhering to measurement protocols, undergoing extensive tester training, choosing appropriate test-retest intervals, minimizing diurnal variability, and taking multiple measurements.

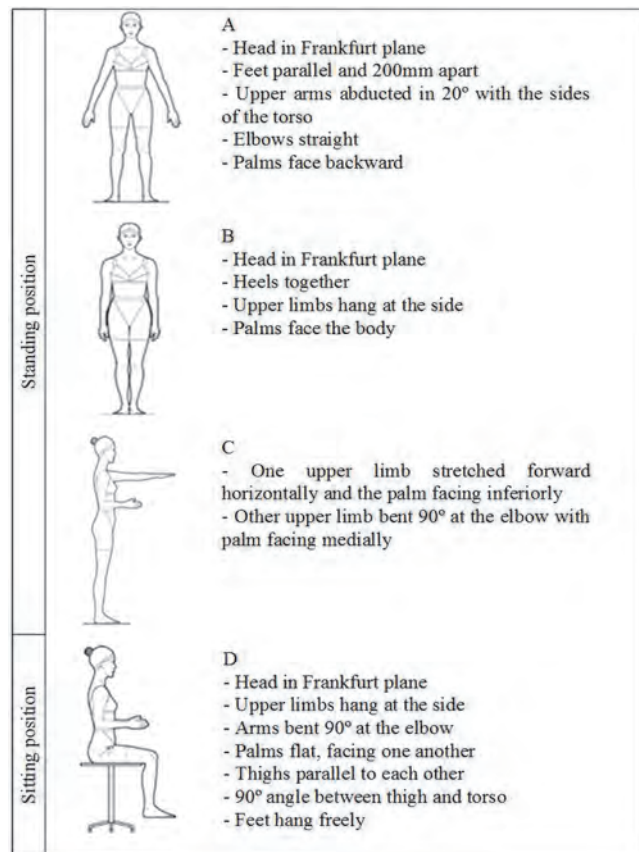


Figure 1. Subject positions during scanning.
Source: [18]

2.3. The different body scanners

There are two types of body scanners: high-end scanning systems that produce high-quality scans for sizing surveys and in-shop or in-house inexpensive scanning systems that produce lower quality scans for retail use [17]. Based on a publication from 2013 [13], several 3D whole body scanning systems are currently available on the market, including those presented in Table 1.

Table 1.
Currently available 3D whole body scanning systems.

Company	Product	City, Country	Technique
Cyberware	WBX	Monterey, CA, USA	Laser line
4ddynamics	Mephisto EX-pro or CX-pro	Antwerp, Belgium	Structured light projection
4ddynamics	Gotcha	Antwerp, Belgium	Structured light projection
Vitronics	Vitus Smart LC	Wiesbaden, Germany	Laser line
Vitronics	Vitus Smart XXL	Wiesbaden, Germany	Laser line
TC2	KX-16	Cary, NC, USA	Infrared
SizeStream	3D body scanner	Cary, NC, USA	Infrared
SpaceVision	Cartesia	Tokyo, Japan	Laser structured light
3dMbody	Flex8	Atlanta, GA, USA	Stereo photogrammetry

Source: Adapted from [13]

There are also new ways of creating three-dimensional images using systems that were not initially designed for the purpose. This is the case of Microsoft Kinect, which can be used for numerous other applications besides games. The Microsoft Kinect sensor is one of a class of devices known as depth cameras in the category of structured light systems [21]. Kinect may be considered a 3D markerless motion capture system because it provides a simplified skeleton in real time, without the need for special clothes or other equipment. Despite the fact that it cannot be used for extremely accurate studies [22-24], it may be deployed when there is no need for high levels of accuracy, for example in clothing or shoe sizing, indirect fat measurement, or clinical rehabilitation [25, 26].

2.4. Applications

Body scanners are used for a wide variety of applications. Jones & Rioux [14] divided their applications into:

- Medical: body deformity; glaucoma; orthodontics; orthopedics; surgery; lung function studies; custom prostheses; breast topography; pediatrics; medical management;
- Human systems engineering: work environment; population anthropology; helmets and face masks; gloves; clothing; human morphology; human motion analysis; forensic imaging; hearing studies;
- Virtual reality and communications: three-dimensional portraits; computer animation of human models.

More recently, according to Ashdown et al. [19], 3D scans have been used to create virtual models of customers for the apparel industry, which consumers can use to try on clothing virtually.

Another important use of body scanners is the creation of anthropometric databases. The first large-scale 3D anthropometry survey project carried out was the Civilian American and European Surface Anthropometry Resource (CAESAR). The CAESAR database contains anthropometric variability of men and women, with ages ranging from 18 to 65 years old. Representatives were asked to ensure the database contained samples for various weights, ethnic groups, gender, geographic regions, and socio-economic status. The study was conducted between April 1998 and early 2000 and included three scans per person for (i) standing posture, (ii) full-coverage posture and (iii) relaxed seating posture. The data collection methods were standardized and documented so that the database may be continually expanded and updated. High-resolution body scans were made using three-dimensional body scanners from Cyberware and Vitronics.

In addition to the CAESAR project, many other studies, such as Size UK, Size USA, Size Spain and Size China, have been conducted to classify entire populations using data collected with 3D body scanners.

3. Reliability and validity of anthropometric data

When using anthropometric data it is important to test its reliability and validity, as these factors may influence both the measurements and the interpretation of the results

obtained. Appannah et al. [27] argue that the validity of data is defined by the ability to achieve the “true value” of a measurement, while Johnson et al. [28] defined reliability as the ability to repeat, reproduce, or consistently obtain the same measurement under identical conditions. According to Mueller & Martorell [29], the reliability of a measurement relies on precision and dependability, the former being the most important determinant. Also important is intra-observer reliability, described as the ability of the same observer to obtain consistent measurement, and inter-observer reliability – ability of different measurers to obtain similar measurement. Kouchi et al. [17] stated that for anthropometric data users there are three essential quality parameters:

- Validity of the data, meaning that the target population is well defined by the subject population of an anthropometric survey;
- Comparability of measurement items, implying that the exact same method is used when taking the same one dimensional measurement;
- Accuracy and precision of measurements, which are affected by factors such as instruments, measurer skills or, even, the participants themselves.

3.1. Sources of errors

Despite the importance of all these issues (because of their impact on measurement error) they are sometimes neglected when conducting an investigation. In most studies the error limits are set prior to data collection, whilst the performance of the measurer is evaluated during the collection process (comparing it with previously defined standards).

Both in traditional anthropometry methods and in 3D anthropometry, there are some factors that may affect the incidence of errors. Kouchi et al. [30] presented a list of some of these factors (Table 2), where they state that the two principal sources of error are related to the devices used and to the persons involved in the data collection process.

The traditional instruments used when collecting anthropometric data are usually simple to calibrate and as such, they are only unreliable if they were poorly designed (e.g., a tape measure that is made from a material that will stretch). On the other hand the calibration of a 3D scanning system can be compromised either by the hardware or software. Most scanning systems have a calibration process that will verify and correct the calibration of the scanner by measuring a simple geometric shape of known dimensions.

The skill of measurement collection processes resides not only in the ability to produce several consistent measurements, but also in the ability to accurately identify the locations of the various landmarks. However, these two factors are very difficult to separate and assess individually as no “true values” are present in the human body [29]. Despite the fact that the repeatability of posture is marked as being a factor caused by the participants, a proper measuring posture and its repeatability are factors that are also related to the measurer, as they can be controlled if the measurer provides proper instructions [30]. As such, it may be said that observer error is the cause of most errors in traditional

Table 2.
Factors that affect errors in anthropometry.

Factors		Traditional anthropometry	3D anthropometry
Devices	Hardware	Accuracy of instrument	Accuracy of scanner system
	Software	-	Accuracy of landmarking software Accuracy of measurement calculation software
Humans	Measurer	Skill of measurement	Skill of landmarking
	Computer operator	-	Accuracy of landmarking software Accuracy of measurement calculation software
	Participant	Repeatability of posture	Repeatability of posture Body sway during scan

Source: Adapted from [30]

anthropometry since it includes imprecision in landmark location, subject positioning, and instrument usage [31]. The same authors also discuss the fact that when multiple observers are involved, this error can be accentuated, as happens in most large-scale anthropometric surveys, where the landmarking process is conducted by a single person but the body measurements are done by several [22].

Although when traditional methods are used the steps to achieving actual usable anthropometric data are not very complex, they are much more complex when a 3D body scanner is employed, as shown in Fig. 2. This is the reason why 3D anthropometry presents so many more factors that may influence the existence of errors in measurements.

Ever since 3D body scanners first appeared there have been authors who have tried to assess the accuracy and precision of the measurements derived from them. Attempts have been made to evaluate them in terms of: (i) comparability with measurements from traditional methods [32-34]; (ii) repeatability of scan-derived measurements [35,36]; (iii) and repeatability of scan-derived landmark locations obtained from the same image [35]. However, Kouchi et al. [30] state that the quality parameters of these studies are not usually consistent, because there is no explicit required accuracy standard and no widely accepted quality evaluation protocol.

Both Table 2 and Fig. 2 show that an important part of the three-dimensional anthropometry is based on landmarking. Most studies evaluate errors in body measurements rather than errors in landmark locations. However, Kouchi et al. [30] showed that the assessment of accuracy using measurement errors underestimates errors in landmark locations. This happens because when the same image is used, the scan-derived landmark locations are not always identical, but may vary according to the methodology – marker stickers indicating landmarks or landmark locations calculated from surface data – and to the way that markers are identified – by an operator, or calculated using software.

The following section goes into more detail about landmarking.

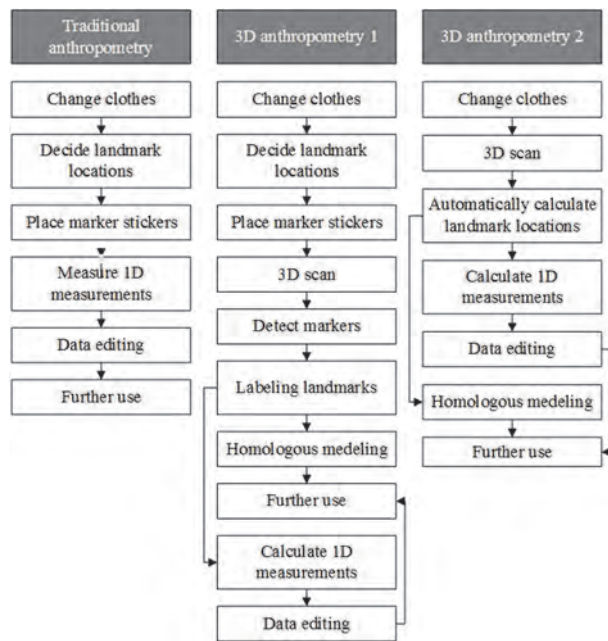


Figure 2. Flow diagram of anthropometric data collection.
Source: [17]

3.2. Landmarking

According to the manual of the International Society for the Advancement of Kinanthropometry [37], a landmark is an identifiable skeletal point which generally lies close to the body's surface and is the marker that identifies the exact location of a measurement site. They are found by palpation or measurement and can be used to define anatomical correspondence between individuals. Most commonly used landmarks are located on specific bones or are easily identifiable by soft tissue features such as nipples or the navel [17].

Wrongly identifying a body landmark is the main cause of observer error in the collection of anthropometric data [30]. As such, in any anthropometry-based study it is extremely important to agree on the body measurements to be recorded and the common points on the body to be identified.

In anthropometry, the same landmark is frequently used to measure several body dimensions. The first step in traditional landmarking is to mark the locations on the body sites that will be measured on the participants' skin using a non-smearing, skin pencil or skin-safe, washable, ink that can be easily removed using makeup remover [38, 39]. Anthropometrists usually use a small cross or dot as a marking symbol (Fig. 3).

Locating the required body landmarks can be a very difficult and time-consuming task. This may be especially problematic in people with more body fat over the bony landmarks [15], or people in wheelchairs with whom it can be hard to gain access to the required landmarks [32]. According to Paquette [41], in 1988, four hours were required to physically landmark, measure, and record the data of one individual in an anthropometric survey of US Army personnel.

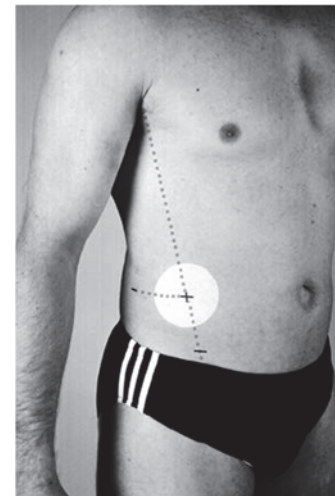


Figure 3. Landmarks identified manually.
Source: [40]

The traditional anthropometric measuring procedure has not changed much since this study was carried out in the late 1980s.

For the marking of landmark locations it is fundamental that participants adopt the correct posture for the measurement or the landmark location might be compromised and the results biased. Kouchi et al. [17] give the example of landmarking the tip of the spinous process of the seventh cervical vertebra, which can be easily palpated at the back of the base of the neck when the participant has the neck bent forward. They state that if the location were marked in this posture, the mark on the skin would slide away from it when participants lift their heads for orientation in the Frankfurt plane, compromising the reliability of the results.

As landmarking is the basis for obtaining valid results, it is important to quantify the measurement errors caused by landmarking. Despite the fact that the repeatability of landmarking has been considered an important factor that, if handled incorrectly, contributes to errors in anthropometry measurement, there are few studies that quantify the phenomenon. Kouchi et al. [30] argued that landmarks with large intra-observer errors also had large inter-observer errors. Additionally, they found that the errors in body dimensions were smaller than landmarking errors in 23 of the 35 measurements analyzed, suggesting that the magnitude of landmarking errors would likely be underestimated by examining errors in body dimensions.

When using 3D body scanners, the landmarking process is also crucial for the correct correspondence of anatomical locations between subjects and across scans. Moreover, measurements derived from reliable landmarks can be used for statistical analysis, for reconstructing variation in human body shape or even for creating homologous models [42, 43]. As such, the poor identification of landmark locations characteristic of 3D anthropometry has a significant effect on the derived data that is used to define participant body dimensions and to effect shape analysis. Landmarks can be placed manually (with traditional markers, for scanners that can sense color differentiation, or small hemispherical objects stuck to the skin for scanners that only capture surface geometry – Fig. 4). Or they may be identified by the



Figure 4. Landmark-based automated measurement process.
Source: [44]

scanner's software system, in which case they are automatically identified from body surface geometries. Fig. 5 shows a set of landmarks automatically derived from a scan based on surface geometries.

Before starting the scan measurement process it is necessary to identify some body parts, in order to achieve an appropriate reconstruction from the captured datapoints. The human model is usually divided at armpits and crotch (a process known as segmentation). Five body parts, including head and torso, both arms, and both legs can be identified. There are many ways to perform this segmentation, for example, Nurre et al. [45] proposed a cusp algorithm for segmenting the 3D scanning data of a human body while Wang et al. [46] applied fuzzy logic concepts to locate the armpits and crotch, and then to separate the arms and legs from the trunk.

The landmarking process is the most problematic aspect of 3D anthropometry since the landmarks placed on bony structures that are found below the skin surface, and palpated by the anthropometrist (in traditional methods) cannot be

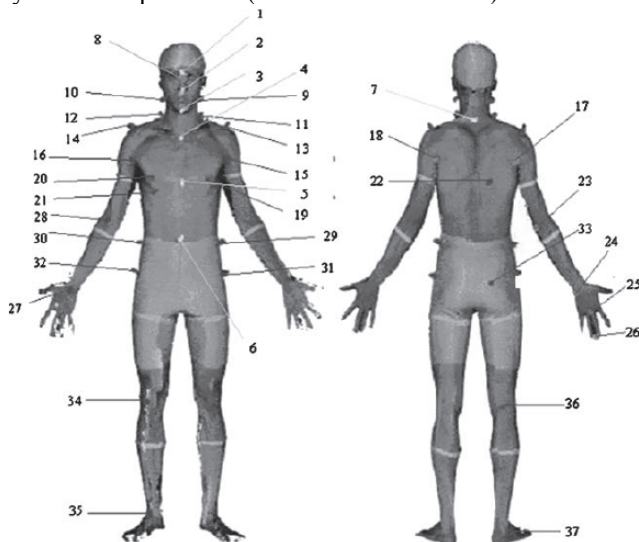


Figure 5. Landmarks identified with a 3D body scanner.
Source: [46]

accurately identified from the surface shape of the scan. As this is a difficult process, and integral to reliable data collection, some attention has been given to developing methods to enable correct landmark identification. Some studies present algorithms for automatically calculating landmark locations [47,48] while others propose alternatives for automatically detecting and calculating 3D coordinates of markers positioned by experienced anthropometrists [46]. However, as creating a good criterion for evaluating the performance of an algorithm is very difficult, the performance of very experienced anthropometrists is used as a criterion for evaluating the performance of algorithms used for calculating scan-derived body dimensions [30].

According to Wang et al. [46], landmark identification methods may in general terms be classified as: (i) premarking, (ii) human body mapping, (iii) geometry analysis and (iv) approximate determination of height location. Often the positions of the landmarks can be easily identified on the scanning image with the human eye, a process that is difficult to program into software. By using color information obtained from the cameras in the scanning heads, manually placed landmarks can also be identified by analyzing the RGB information in the scanned image [6,46,49].

However, the procedure of placing markers on the body surface is time consuming, relatively invasive, and may involve human error. Kouchi et al. [30] discuss the fact that the amount of error in identifying landmarks is not well known. They examined the landmarking of 40 individuals carried out by experienced and novice markers and compared the differences in measurements taken with reference to the landmarks identified. Differences in measurements obtained that were due to intra- and inter- observer error were sufficiently large that it was suggested that the explicit definition of landmarks in more detail might reduce landmarking errors.

Therefore, the possibility of developing marker-free techniques for landmarking becomes an important issue for analyzing 3D whole body scanning data. For the method of automated landmarking, analyzing the geometry of the human body using techniques such as silhouette analysis is a logical approach [50]. Douros et al. [51] used the method of reconstructing curves and surfaces to locate the landmarks. Allen et al. [42] proposed a method for identifying landmarks efficiently that employed a template mapping approach, which makes use of information from the existing database of 3D human models. Lu and Wang [52] used four algorithms (silhouette analysis, minimum circumference determination, gray-scale detection, and human-body contour plots) to locate 12 landmarks and 3 characteristic lines automatically, making it possible to obtain 104 body dimensions.

According to Kouchi et al. [30], the identification of the 3D coordinates of landmark locations can be done entirely manually, entirely automatically, or using a mix of both techniques:

- Deciding landmark locations on the body: a measurer decides landmark locations manually or a system calculates them automatically;
- Obtaining 3D coordinates of markers: an operator manually picks the centers of marker stickers or a system recognizes stickers automatically and calculates 3D coordinates;

- Naming landmarks (or labeling): either an operator or a system automatically names each marker.

Although selecting a manual method to decide landmark locations on the human body and placing a marker is time consuming, so too is doing it semi-automatically by validating marker centers and naming markers. If on the one hand, automatic calculations of landmark locations save time, on the other, they may not always match the landmark locations identified by experienced anthropometrists.

4. Conclusions

Creating anthropometric databases typically requires considerable resources (time, knowledge, funds, equipment and people). To overcome these limitations, technological developments in recent years, using three-dimensional digital forms, has made it possible to advance the study of human size and shape using fewer resources. 3D body scanners make anthropometric data acquisition more practical, faster, and less expensive. It also has the potential to produce valid and reliable measurements. With these advances it is now possible to explore the possibilities of anthropometric measurement processes and create new perspectives on its use.

3D scanning systems have evolved over the last few years. From LASS to Kinect, the technology is always advancing. The type of technology used differs from scanner to scanner. There are four main types of technology: - laser line systems, structured light systems, multi-view camera systems and millimeter wave systems. Currently, there is a significant range of three-dimensional body scanners available in the market (nine major products developed in different parts of the globe).

Depending on the technology and product used, the results and applications are different. Different products may have different reliability issues, potentially compromising the applicability of the data. Depending on the desired application -i.e., highly precise anthropometric data or less accurate data for apparel applications- a variety of body scanners may be used. For example, the Microsoft Kinect, which is less accurate, may be used for apparel applications, while the Vitrus Smart LC, because it is more precise, can be used for compiling an anthropometric database.

The same logic can be applied when selecting traditional methods versus 3D anthropometry, as which method should be used will depend on the type of data required (one-dimensional measurements or three-dimensional surface shapes) and on how it will be applied.

Acknowledgments

This work is financed by FEDER funds through the Competitive Factors Operational Program (COMPETE) POCI-01-0145-FEDER-007043 and POCI-01-0145-FEDER-007136 and by national funds through FCT – the Portuguese Foundation for Science and Technology, under the projects UID/CEC/00319/2013 and UID/CTM/000264 respectively.

References

- [1] Simmons, K.P. and Istook, C. L., Body measurement techniques: Comparing 3D body-scanning and anthropometric methods for apparel applications. *Journal of Fashion Marketing and Management: An International Journal*, 7(3), pp. 306-332, 2003. DOI: 10.1108/13612020310484852
- [2] ISO-7250., Basic human body measurements for technological design, 1996.
- [3] Kroemer, K. and Kroemer, H.J., *Engineering physiology: Bases of human factors/ergonomics*. New York: John Wiley & Sons, 1997.
- [4] Montagu, M.F. and Brožek, J.C., *A handbook of anthropometry*. Springfield: Charles C Thomas Publisher, 1960. DOI: 10.1037/12018-000
- [5] Lovesey, E.J., A method for determining facial contours by shadow projection. Royal Aircraft Establishment Technical Report TR66192, 1966.
- [6] Daanen, H.A.M. and van de Water, G.J., Whole body scanners. *Displays*, 19(3), pp. 111-120, 1998. DOI: 10.1016/S0141-9382(98)00034-1
- [7] Bindahman, S., Zakaria, N. and Zakaria, N., 3D body scanning technology: Privacy and ethical issues, *Proceeding of the International Conference on Cyber Security, Cyber Warfare and Digital Forensic (IEEE CyberSec)*, pp. 150-154, 2012.
- [8] Jones, P.R.M., West, G.M., Harris, D.H. and Read, J.B., The loughborough anthropometric shadow scanner (LASS). *Endeavour*, 13(4), pp. 162-168, 1989. DOI: 10.1016/S0160-9327(89)80014-6, 10.1016/S0160-9327(89)80004-3
- [9] Brooke-Wavell, K., Jones, P.R.M. and West, G.M., Reliability and repeatability of 3-D body scanner (LASS) measurements compared to anthropometry. *Annals of Human Biology*, 21(6), pp. 571-577, 1994. DOI: 10.1080/03014469400003572
- [10] Bragança, S., Arezes, P. and Carvalho, M., An overview of the current three-dimensional body scanners for anthropometric data collection, *Occupational Safety and Hygiene III*, pp. 149-153, 2015. DOI: 10.1201/b18042-32
- [11] Treleaven, P. and Wells, J., 3D body scanning and healthcare applications. *Computer*, 40(7), pp. 28-34, 2007. DOI: 10.1109/MC.2007.225
- [12] Istook, C.L. and Hwang, S., 3D body scanning systems with application to the apparel industry. *Journal of Fashion Marketing and Management*, 5(2), pp. 120-132, 2001. DOI: 10.1108/EUM0000000007283
- [13] Daanen, H.A.M. and Haar, F.B., 3D whole body scanners revisited. *Displays*, 34(4), pp. 270-275, 2013. DOI: 10.1016/j.displa.2013.08.011
- [14] Jones, P.R.M. and Rioux, M., Three-dimensional surface anthropometry: Applications to the human body. *Optics and Lasers in Engineering*, 28(1), pp. 89-117, 1997. DOI: 10.1016/S0143-8166(97)00006-7
- [15] Olds, T. and Honey, F., The use of 3D whole-body scanners in anthropometry. *Proceedings of the 9th International Conference of the International Society for the Advancement of Kinanthropometry*, pp. 1-12, 2006.
- [16] Mckinnon, L. and Istook, C., Comparative analysis of the image twin system and the 3T6 body scanner. *Journal of Textile and Apparel, Technology and Management*, 1(2), pp. 1-7, 2001.
- [17] Kouchi, M., Gupta, D. and Zakaria, N., Anthropometric methods for apparel design: Body measurement devices and techniques. *Anthropometry, Apparel Sizing and Design*, pp. 67-94, 2014. DOI: 10.1533/9780857096890.1.67
- [18] ISO-20685., 3-D scanning methodologies for internationally compatible anthropometric databases, 2010.
- [19] Ashdown, S.P., Loker, S., Schoenfelder, K. and Lyman-Clarke, L., Using 3D scans for fit analysis. *Journal of Textile and Apparel, Technology and Management*, 4(1), pp. 1-12, 2004.
- [20] Tomkinson, G.R. and Shaw, L.G., Quantification of the postural and technical errors in asymptomatic adults using direct 3D whole body scan measurements of standing posture. *Gait and Posture*, 37(2), pp. 172-177, 2013. DOI: 10.1016/j.gaitpost.2012.06.031
- [21] Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R. and Blake, A., Real-time human pose recognition in parts from single depth images. *Studies in Computational Intelligence*, 411(1), pp. 119-135, 2013. DOI: 10.1007/978-3-642-28661-2_5
- [22] Ye, M., Wang, X., Yang, R., Ren, L. and Pollefeys, M., Accurate 3D pose estimation from a single depth image. *2011 International Conference on Computer Vision*, pp. 731-738, 2011. DOI: 10.1109/ICCV.2011.6126310
- [23] Clark, R.A., Pua, Y.H., Fortin, K., Ritchie, C., Webster, K.E., Denehy, L. and Bryant, A.L., Validity of the Microsoft Kinect for assessment of

- postural control. *Gait and Posture*, 36(3), pp. 372-377, 2012. DOI: 10.1016/j.gaitpost.2012.03.033
- [24] Bragança, S., Carvalho, M., Xu, B., Arezes, P. and Ashdown, S., A validation study of a Kinect based body imaging (KBI) device system based on ISO 20685:2010. *Proceedings of the 5th International Conference on 3D Body Scanning Technologies*, pp. 372-377, 2014.
- [25] Fernández-Baena, A., Susin, A. and Lligadas, X., Biomechanical validation of upper-body and lower-body joint movements of Kinect motion capture data for rehabilitation treatments. *Proceedings of the 2012 4th International Conference on Intelligent Networking and Collaborative Systems*, pp. 656–661, 2012. DOI: 10.1109/INCoS.2012.66
- [26] Weiss, A., Hirshberg, D. and Black, M.J., Home 3D body scans from noisy image and range data. *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1951-1958, 2011. DOI: 10.1109/iccv.2011.6126465
- [27] Appannah, G., Haniff, J., Mohammad Nor, N.S., Wong, N.F., Kee, C. C., Zainuddin, A.A. and Yusoff, A.F., Reliability, technical error of measurements and validity of instruments for nutritional status assessment of adults in Malaysia. *Malaysian Journal of Nutrition*. Nutrition Society of Malaysia, 2008.
- [28] Johnson, T.S., Engstrom, J.L. and Gelhar, D.K., Intra- and interexaminer reliability of anthropometric measurements of term infants. *Journal of Pediatric Gastroenterology and Nutrition*, 24(5), pp. 497-505, 1997. DOI: 10.1097/00005176-199705000-00001
- [29] Mueller, W.H. and Martorell, R., Reliability and accuracy of measurement. *Anthropometric Standardization Reference Manual*, pp. 83-86, 1988.
- [30] Kouchi, M. and Mochimaru, M., Errors in landmarking and the evaluation of the accuracy of traditional and 3D anthropometry. *Applied Ergonomics*, 42(3), pp. 518-527, 2011. DOI: 10.1016/j.apergo.2010.09.011
- [31] Bennett, K.A. and Osborne, R.H., Interobserver measurement reliability in anthropometry. *Human Biology*, 58(5), pp. 751-759, 1986.
- [32] Sims, R.E., Marshall, R., Gyi, D.E., Summerskill, S.J. and Case, K., Collection of anthropometry from older and physically impaired persons: Traditional methods versus TC 2 3-D body scanner. *International Journal of Industrial Ergonomics*, 42(1), pp. 65-72, 2012. DOI: 10.1016/j.ergon.2011.10.002
- [33] Han, H., Nam, Y. and Choi, K., Comparative analysis of 3D body scan measurements and manual measurements of size Korea adult females. *International Journal of Industrial Ergonomics*, 40(5), pp. 530-540, 2010. DOI: 10.1016/j.ergon.2010.06.002
- [34] Lu, J.M. and Wang, M.J.J., The evaluation of scan-derived anthropometric measurements. *IEEE Transactions on Instrumentation and Measurement*, 59(8), pp. 2048-2054, 2010. DOI: 10.1109/TIM.2009.2031847
- [35] Kouchi, M. and Mochimaru, M., Evaluation of Accuracy in Traditional and 3D Anthropometry, 2008.
- [36] Robinette, K.M. and Daanen, H.A.M., Precision of the CAESAR scan-extracted measurements. *Applied Ergonomics*, 2006. DOI: 10.1016/j.apergo.2005.07.009
- [37] International Society for the Advancement of Kinanthropometry, available at www.isakonlinve.com
- [38] Roebuck, J.A., Kroemer, K.H.E. and Thomson, W.G., *Engineering Anthropometry Methods*, 1975.
- [39] O'Brien, R. and Shelton, W.C., *Women's Measurements for Garment and Pattern Construction*, 1941.
- [40] Marfell-Jones, T.O.A.S., L.C.M., Stewart, A. and Marfell-Jones, M., *International Standards for Anthropometric Assessment*, 2006.
- [41] Paquette, S., 3D scanning in apparel design and human engineering. *IEEE Computer Graphics and Applications*, 16(5), pp. 11-15, 1996. DOI: 10.1109/38.536269
- [42] Allen, B., Curless, B. and Popović, Z., The space of human body shapes. *ACM Transactions on Graphics*. ACM, 2003.
- [43] Mochimaru, M. and Kouchi, M., Statistics for 3D human body forms. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. SAGE Publications, 2000. DOI: 10.4271/2000-01-2149, 10.1177/154193120004403846
- [44] Human Solutions, AnthoScan - Body Dimensions Acquired Easily, [Online]. Available at: http://www.human-solutions.com/download/pdf/ANTHROSCAN_en.pdf
- [45] Nurre, J.H., Connor, J., Lewark, E.A. and Collier, J.S., On segmenting the three-dimensional scan data of a human body. *IEEE Transactions on Medical Imaging*, 19(8), pp. 787-797, 2000. DOI: 10.1109/42.876304
- [46] Wang, M.-J.J., Wu, W.-Y., Lin, K.-C., Yang, S.-N. and Lu, J.-M., Automated anthropometric data collection from three-dimensional digital human models. *The International Journal of Advanced Manufacturing Technology*, 2007. DOI: 10.1007/s00170-005-0307-3
- [47] Azouz, B.Z., Shu, C. and Mantel, A., Automatic locating of anthropometric landmarks on 3D human models. *Proceedings of the Third International Symposium on 3D Data Processing, Visualization and Transmission*, 2006.
- [48] Leong, I.F., Fang, J.J. and Tsai, M.J., Automatic body feature extraction from a marker-less scanned human body. *CAD Computer Aided Design*, 39(7), pp. 568-582, 2007. DOI: 10.1016/j.cad.2007.03.003
- [49] Burnsides, D., Boehmer, M. and Robinette, K., 3-D landmark detection and identification in the CAESAR project. *Proceedings Third International Conference on 3-D Digital Imaging and Modeling*, pp. 393-398, 2001.
- [50] Buxton, B., Dekker, L., Douros, I. and Vassilev, T., Reconstruction and interpretation of 3D whole body surface images. *Scanning*, 2000.
- [51] Douros, I., Dekker, L. and Buxton, B., Reconstruction of the surface of the human body from 3D scanner data using B-splines. *Proceedings of the Electronic Imaging '99*, pp. 234-245, 1999. DOI: 10.1117/12.341065
- [52] Lu J. and Wang, M., Automated anthropometric data collection using 3D whole body scanners. *Expert Systems with Applications*, 2008. DOI: 10.1016/j.eswa.2007.07.008

S. Bragança, is a PhD. student in the Department of Production and Systems, School of Engineering, University of Minho, Portugal. She completed her MSc. in Industrial Engineering and Management in 2012, also at the University of Minho, Portugal. Her MSc Thesis was in the Implementation of Standard Work and other Lean Production tools. Her current research focuses on ergonomics and anthropometry. ORCID: orcid.org/0000-0002-4765-3856

P. Arezes, has a PhD. in Industrial and Systems Engineering from the University of Minho, Portugal, where he is currently a full professor in Ergonomics and Human Factors. He is also a visiting fellow at MIT's AgeLab in the USA. He leads the Human Engineering research group and coordinates the Engineering Design and Advanced Manufacturing (EDAM) area of MIT's Portugal Program at the University of Minho, where he is also chair of the Steering Board of the PhD program "Leaders for Technical Industries (LTI)". ORCID: orcid.org/0000-0001-9421-9123

M. Carvalho, graduated in Textile Engineering; he has an MSc. in Design and Marketing and a PhD. in Textile Engineering from the University of Minho, Portugal. Since 1993 he has been a researcher at 2C2T - the University of Minho's Center of Science and Textile Technology. He is involved in the supervision of research projects in the areas of clothing and textile design; comfort; pattern design; anthropometrics; ergonomics and development of functional/multi-functional materials and textile products; and interactive textiles with applications in the health, automobile, sport and other sectors. ORCID: orcid.org/0000-0001-8010-6478

S. Ashdown, is Helen G. Canoyer Professor in the Department of Fiber Science & Apparel Design at Cornell University. Her research group investigates the impact of new technologies on apparel design, with a focus on sizing and fit, functional apparel design, and the use of the 3-D full-body scanners as a research tool for apparel designers. ORCID: orcid.org/0000-0002-0276-4122

Understanding finger postures when touching targets on the touchscreen of mobile devices

Mahrus K. Umami ^a, Pedro M. Arezes ^b & Álvaro M. Sampaio ^c

^a School of Engineering, University of Minho, Portugal, and School of Engineering, University of Trunojoyo Madura, Indonesia. mahrusku@gmail.com

^b Centre of ALGORITMI, School of Engineering, University of Minho, Portugal, parezes@dps.uminho.pt

^c Institute for Polymers and Composites IPC/I3N, School of Engineering, University of Minho, Portugal. amsampaio@dep.uminho.pt

Received: November 30th, 2015. Received in revised form: March 10th, 2016. Accepted: April 12th, 2016.

Abstract

This paper presents the results of a preliminary study on the interaction of the fingertip and a mobile device's touch screen. The objective of this study is to identify the finger postures that will be included in the main study on the contact area in the interaction between fingertip and a flat surface. Twenty participants (15 males and 5 females) took part in this study. They were asked to complete two tasks in a sitting posture. In the first task, they had to touch targets on the mobile device screens by tapping them sequentially, while in the second task they were asked to connect the targets with straight lines. The results showed that the participants used mainly their thumbs and index fingers to touch targets on the screen of the devices. Only a small number of participants used their middle finger, and only in a few touching activities.

Keywords: finger posture; touch screen device; mobile phone; tablet.

Comprensión de las posturas de dedos al tocar objetivos en la pantalla táctil de dispositivos móviles

Resumen

Este trabajo presenta los resultados de un estudio preliminar sobre la interacción de la yema del dedo y la pantalla táctil de un dispositivo móvil. El objetivo del estudio es identificar posturas de los dedos que se incluirán en el estudio principal en el área de contacto en la interacción entre las puntas de los dedos y una superficie plana. Veinte participantes (15 hombres), participaron en este estudio. Se les pidió que completasen dos tareas en una postura sentada. En la primera tarea, tuvieron que tocar objetivos en las pantallas de dispositivos móviles de una forma secuencial, mientras que en la segunda tarea se les pidió que conectasen posiciones objetivo con líneas rectas. Los resultados mostraron que los participantes utilizan principalmente sus dedos pulgares e indicadores para tocar posiciones objetivo en la pantalla de los dispositivos. Sólo un pequeño número de participantes usó su dedo medio, y en pocas actividades.

Palabras clave: postura; dedos; dispositivo con pantalla de toque; teléfono móvil; tableta.

1. Introduction

Nowadays, people use their mobile devices everywhere, adopting various postures with their fingers. They use their mobile phones when walking, standing, or sitting, as part of their daily activities. Karlson's study [1] identified that travellers operated their mobile devices using only one hand at an airport when walking, standing,

or sitting. The observation involved users of mobile phones and PDAs. There were two types of mobile phones recorded in the study, the candy bar and the flip types. A larger survey by Hooper [2] showed that people use their mobile devices in standing, walking and sitting positions, and while travelling on public transportation, such as buses or trains. The survey observed how people held their mobile devices at common places, such as universities, parks, offices, public transportation, and shopping malls.

How to cite: Umami, M.K., Arezes, P.M. and Sampaio, A.M., Understanding finger postures when touching targets on the touchscreen of mobile devices. DYNA 83(197), pp. 31-37, 2016

Hoober recorded a Samsung Galaxy Note 2 as the largest device in his data set. From both surveys, it was possible to conclude that people hold their mobile devices in different ways: one hand, two hands, and cradle. Additionally, people can also use their mobile devices by putting them on a table, especially when they interacting with tablets.

People change their hand postures very often when interacting with their mobile devices, changing postures during usage time. It is possible to see that they often change the way they hold their mobile device easily while using it. Hoober repeatedly saw individuals that adopted various hand postures while using the same device, in a short amount of time. He recorded individuals that used one hand at the beginning, then used the other for a different activity, then adopted cradle posture, and then returned to the initial posture [2]. In accordance with Hoober's findings, Azenkot and Zhai [4] found that all participants used at least two methods to complete the tasks required in their study on the use of two thumbs, one thumb, and index finger in human interactions with mobile devices. Regarding the most commonly used fingers, Hoober [2] and Azenkot and Zhai [4] noticed that most people used their thumbs and index fingers.

It is also possible to see that the target size—defined by considering the screen size of the device and designed according to the size of a human finger pad—of a tablet is usually larger than that of smaller devices [5]. Microsoft's Windows Phone Silverlight development guidelines recommend that the touch target size can be greater than or equal to a 9 mm square [6]. For the minimum target size, Microsoft allows their developers to use a 7 mm square when the smaller hit target is warranted [6]. The minimum target size recommended by Microsoft is the same as the size recommended by Google. Google's Android UI Guidelines suggest a 7 mm square (48 pixels square) [5], although for targets that are rarely used by users, the developers can apply a size smaller than 7 mm square. A smaller minimum target size is recommended by Apple. The iOS Human Interface Guidelines allow a 6.5 mm square (44 pixels square) for their developers [7]. All of the operating system developers define the minimum targets size by considering the touching accuracy [5-7].

Few studies showed that the target size has a significant relation to the user errors, performance and satisfaction. Parhi's study, for example, found that user errors decline as the size of the target increases [8]. Park's study on touch key design, showed that user performance and subjective satisfaction with the larger touch key size was higher than the satisfaction with smaller sizes [9]. Both studies showed that the target size should fit to the fingertip contact area on the target.

It is commonly known that users try to touch the targets on the touch screen devices accurately by manipulating the finger posture. Wang and Ren [10] observed two different forms of fingertip contact: the vertical touch and the oblique touch. They found that the size of the fingertip contact area is significantly different with each of the two forms of touching. Holz and Baudisch [11] studied the users' mental model of touch in their efforts to minimize error. The study

explored techniques used by the participants in targeting a crosshair accurately. They found that users tried to touch the target on their device by aligning the finger's feature and outline. Both studies observed the individuals touching the screen display in a flat posture on a table.

Since previous studies [2,4,10,11] have not completely described which fingers and respective postures are normally used by individuals in interacting with the touch screen devices, an in-depth study is required to acquire that information.

This paper presents the results of a pilot test on the interaction between the finger and the touch screen of mobile devices in various postures. The test is the preliminary observation of a proposed study on the contact area of fingertips and touch screen devices. The purpose of this study is to identify the user's typical finger posture in his interaction with the screen of a mobile device. The rationale for this observation is to determine the fingers and their pitches and rolls that should be included in the proposed main study on the fingertip contact area with the touch screen. A previous version of this paper with a smaller number of participants has been published elsewhere [3].

2. Materials and Methods

2.1. Participants

Twenty volunteer participants (15 males and 5 females) were involved in this study. Their age ranged from 19-58 years with a mean of 34.85 (SD. \pm 9.76). All of them were right handed.

2.2. Equipment and tools

Two mobile devices, a Samsung Galaxy S Duos mobile phone (4 inch screen) and a Samsung Galaxy Note 10.1 tablet (10 inch screen), were used to display the targets. There were 8 targets (5 x 5 mm² squares with blue outlines) set on the mobile phone screen, and 12 targets (7 x 7 mm² squares with blue outlines) set on the screen of the tablet. Fig. 1 and 2 show the position of the targets on the screens of the mobile phone and tablet respectively.

A Sony DCR DVD-403E PAL video camera and an Olympus VR-340 compact camera were used to record the hand postures of the participants while performing the tasks. The Sony camera was placed next to the participants' hands and was used to record their finger pitch angles when performing tasks, while the Olympus camera was set behind the observation desk and was used to record the finger roll angles.

A table with a desk height (DH) of 74 cm was used to place the equipment on. A chair with a seat height (SH) of 55 cm was provided for the participants to be able to perform the tasks comfortably in a sitting posture.

An AutoCAD 2010 software application was used in the measurements of the finger pitch and roll angles. A Microsoft Excel for Windows was used to handle and to analyse the data.

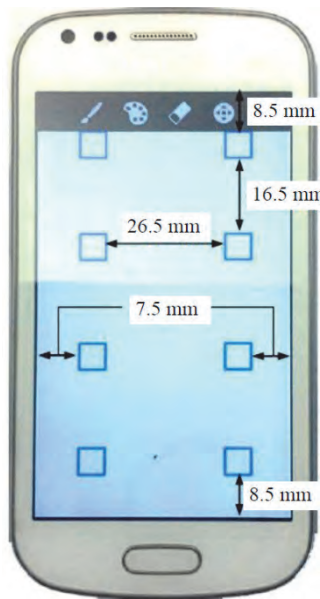


Figure 1. Position of targets on mobile phone screen.
Source: The authors

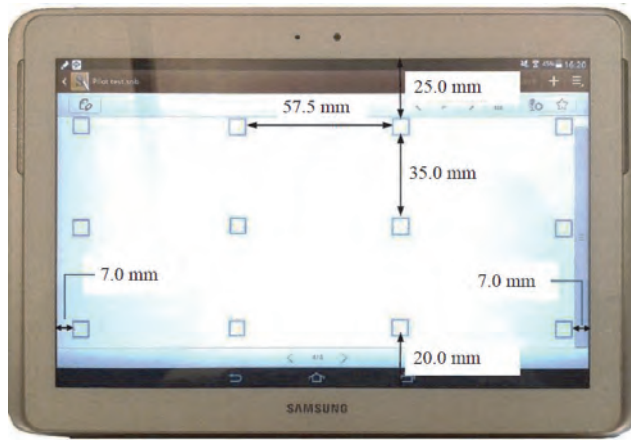


Figure 2. Position of targets on tablet screen.
Source: The authors

2.3. Experiment design and procedures

This study was designed to consider the position of targets, the type of tasks, and the hand postures adopted by the participants to touch the targets as independent variables, while the chosen fingers, finger pitches, and finger rolls were considered as response variables.

In the planning phase, a number of postures that might be performed by the users when interacting with mobile devices were identified: four postures in the interaction with the tablet and four postures for the interaction with the mobile phone. From the eight postures, the two hands with two thumb-screen interaction for the tablet was ignored given that in our previous study [3], a few participants found it difficult to complete the tasks in this posture. The two-hand cradle posture with one thumb-screen interaction for mobile phone use was also ignored. It was assumed that the use of a single

Table 1.
Hand postures observed in the study.

Devices	Hand postures when interacting with mobile devices
Mobile phone	Two hands: cradle
	Single hand: one thumb
	Two hands: two thumbs
Tablet	Two hands: cradle
	Single hand: tilt (45°) on table
	Single hand: flat on table

Source: The authors



Figure 3. Illustration of the finger pitch and the finger roll.
Source: The authors

hand with one-thumb interaction had adequately represented the thumb postures for the observation. Only the use of an index finger or a middle finger in the cradle posture was observed. Table 1 listed the users' hand postures observed in this study.

In the experimental phase, participants were asked to perform two tasks in a sitting posture. In the first task, the participants were asked to touch targets appearing on the devices by tapping them in various hand postures sequentially. Then, in the second task, they were asked to touch each target by connecting them as they usually drag an icon or an image on the touch screen. The participants were allowed to choose which finger they wanted to use for the required tasks.

In sum, the experiment was designed to include 20 participants \times 3 postures \times 2 tasks \times 12 targets = 1440 touches on the tablet and 20 participants \times 3 postures \times 2 tasks \times 8 targets = 960 touches on the mobile phone.

The observation focused on finger pitches and rolls when touching the targets on the screen. The definition of the finger pitch and roll used by Holz and Baudisch [11] was adopted. The finger pitch is the angle between the finger and the device screen, while the finger roll is the rotation of the finger in a clockwise or anti-clockwise motion. Fig. 3 illustrates the definition of the finger pitch and the finger roll.

2.4. Data processing and analysis

The relevant images of the hand postures were captured from the videos recorded by two cameras. Then, the images were imported to the AutoCAD 2010 software, so that the finger pitch and the finger roll angle could be measured. To simplify data recording and computation, the finger pitch and roll angles measured from the images were classified by following a guideline shown in Table 2.

Table 2.

Guideline for recording finger pitch angle used in this study

Range of finger pitch/roll angles measured from images	Finger pitch/roll angles used in data recording
0° - 37.5°	30°
>37.5° - 52.5°	45°
>52.5° - 67.5°	60°
>67.5° - 82.5°	75°
>82.5°	90°

Source: The authors

The data was summarised and analysed by using MS Excel for Windows on a computer. The frequency of the appearance of the finger pitch and the finger roll angles were counted manually and shown as the percentage of the sum of all the data in the related posture. All the data were analysed and presented in descriptive statistics.

3. Results and Discussion

It was found that most of the participants used their index finger when interacting with the tablet in all postures observed in this study, while only a small number of them used their middle fingers, and one participant was identified as not performing a cradle posture. It was also observed that two female participants used their middle fingers to touch the screen. They used their index and middle fingers, mainly because of their long fingernails.

In their interaction with the mobile phone, two participants did not use the single-handed posture. Besides, three participants did not use the cradle posture and four participants did not use the two thumbs posture. It was also observed that the participants performed the tasks by using their thumbs in the single- and two-handed interactions and their index fingers in the cradle interaction.

For the whole set of observations, we recorded 528 touching activities in 16 thumb postures and 288 touching activities in 9 index finger postures adopted by the participants when interacting with their mobile phones. Further, we recorded 1296 touching activities in 10 index finger postures and 120 touching activities in 3 middle finger postures adopted by the participants in their interaction with the tablet. Table 3 listed the number of the touching activities and the finger postures (pitch and roll angles combination) adopted by the participants.

Table 3.

Number of the touching activities and the finger postures adopted by the participants.

Device	Hand posture	Involved finger	Touching activities	Finger postures
Mobile phone	Single hand and two hands	Thumb	528	16
	Two hands: cradle	Index finger	288	9
Tablet	Two hands: cradle	Index finger	408	9
		Middle finger	48	2
	Single hand: flat on table	Index finger	456	7
		Middle finger	24	1
	Single hand: tilt (45°) on table	Index finger	432	10
		Middle finger	48	2

Source: The authors

3.1. Index finger postures in the interaction with tablet

In the interaction with the tablet in a tilt (45°) position, most of the participants adopted the 75° pitch and 0° roll (28.70%). For the flat position on the table, the participants equally performed 45° pitch and 0° roll, 60° pitch and 0° roll, and 75° pitch and 0° roll combinations. Fig. 4 and 5 respectively show the percentage of the finger pitch and roll combinations which occurred in the interaction with the tablet in tilt and flat positions on the table.

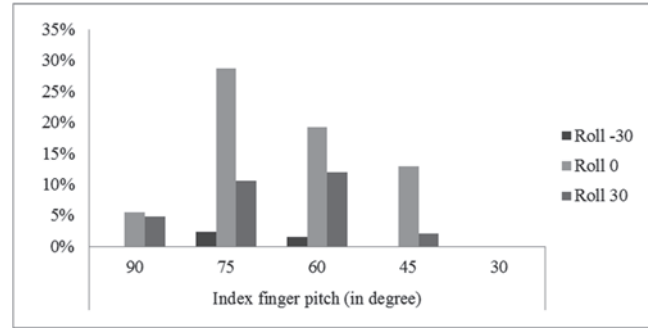


Figure 4. The index finger pitches and rolls occurred in the interaction with tablet in the tilt position on a table.

Source: The authors

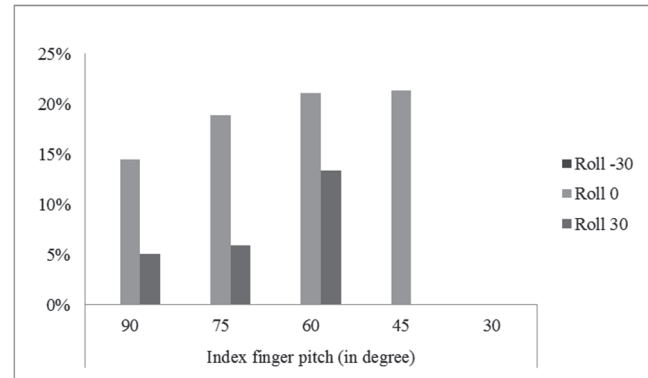


Figure 5. Index finger pitches and rolls in the participants' interaction with their tablet in the flat position on a table.

Source: The authors

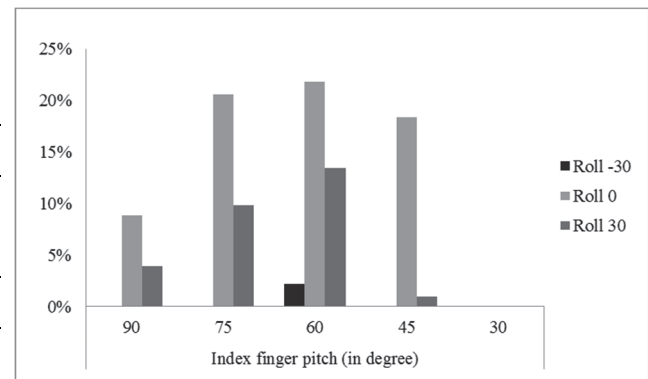


Figure 6. Index finger pitches and rolls in the participants' interaction with their tablet in the cradle posture.

Source: The authors

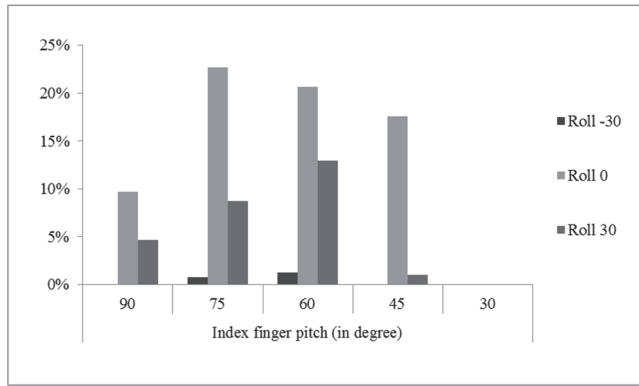


Figure 7. Index finger pitches and rolls in the observed interaction with tablets.

Source: The authors

Fig. 6 shows the index finger pitch and roll combinations occurred in the interaction with the tablet screen in the cradle posture. We can observe the five combinations of finger pitch and roll angles that were most commonly used by the participants, namely: 60° pitch and 0° roll (21.81%), followed by 75° pitch and 0° roll (20.59%), 45° pitch and 0° roll (18.38%), 60° pitch and 30° roll (13.48%), and 75° pitch and 30° roll (9.80%).

Fig. 7 summarizes the occurrence of various index finger pitches and rolls angles performed by participants in their interaction with the tablet in all three postures observed in this study. We can see six combinations of finger pitch and roll angles that were most commonly used by participants, namely: 75° pitch and 0° roll (22.69%), followed by 60° pitch and 0° roll (20.68%), 45° pitch and 0° roll (17.59%), 60° pitch and 30° roll (12.96%), 90° pitch and 0° roll (9.72%) and 75° pitch and 30° roll (8.72%).

3.2. Index finger postures in the interaction with mobile phone

In the interactions with the mobile phone, the index finger can be seen to be used in the cradle posture. Fig. 8 shows the percentage of the index finger pitch and roll angles combinations used by the participants when touching targets on the mobile phone screen in the cradle posture.

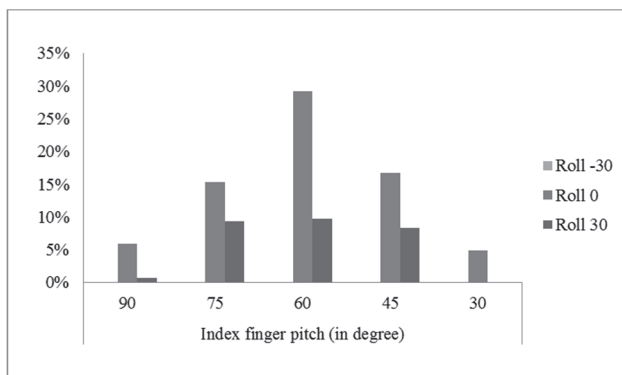


Figure 8. Index finger pitches and rolls in the interaction with mobile phones in the cradle posture

Source: The authors

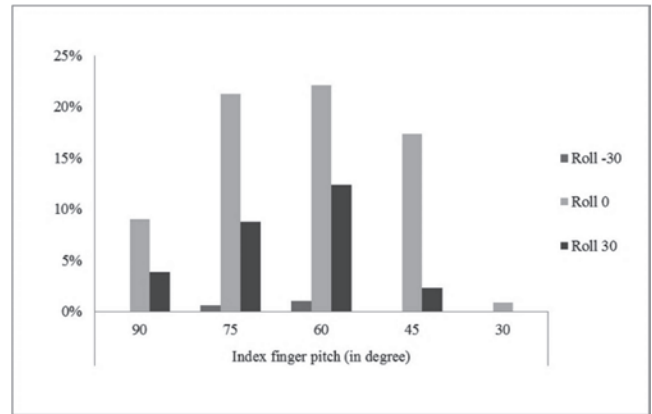


Figure 9. Index finger pitches and rolls in the observed interaction with both devices.

Source: The authors

We can see in Fig. 8 that the five most commonly used combinations of index finger pitch and roll angles for the interaction are: 60° pitch and 0° roll (29.17%), 45° pitch and 0° roll (16.67%), 75° pitch and 0° roll (15.28%), 60° pitch and 30° roll (9.72%) and 75° pitch and 30° roll (9.38%).

3.3. Index finger postures in the interaction with both devices

In Fig. 6 and 8, we can see that in the cradle interaction with the tablet or the mobile phone, the typical posture of the index finger adopted by the participants is the combination of 60° pitch and 0° roll angles. The other common postures are the combination of 75° and 45° pitches and 0° roll angles. Finally, it was also observed that there is no significant difference in the posture of the index finger adopted by the participants in the cradle posture when using both devices.

Fig. 9 shows the index finger postures most commonly used by the participants when touching the target on the screen of both devices. Fig. 9 allows us to conclude that for the whole observation, the participants performed the tasks by using their index finger in the 0° roll angles combined with 60°, 70°, and 45° pitches angles respectively.

3.4. Thumb postures in the interaction with mobile phone

Fig. 10 shows the thumb pitch and roll angle combinations used by the participants when interacting with the mobile phone screen.

Fig. 10 shows five combinations of pitch and roll angles of thumb most commonly used by participants in their interaction with the mobile phone, namely: 60° pitch and 0° roll (22.35%), 45° pitch and 0° roll (15.72%), 30° pitch and 0° roll (14.58%), 60° pitch and 30° roll (13.26%) and 45° pitch and 30° roll (7.77%).

It should be noted that two participants in this study used their middle fingers in the interaction with the tablet, performing combinations of 45° and 30° pitches and rolls respectively. They used these postures due to their long fingernails.

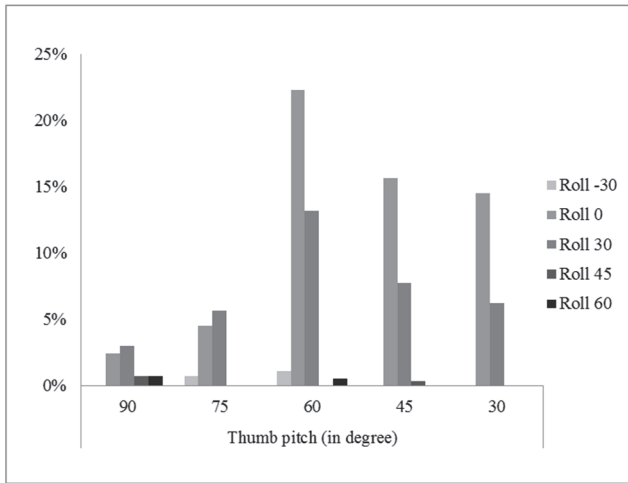


Figure 10. Thumb pitches and rolls in the interaction with mobile phones in single-handed, one-thumb and two-hand, and two-thumb interactions
Source: The authors

The current study found that the index finger postures commonly used by the participants are quite different from those observed in Holz's study [11] on the users' mental models of touch. Holz's study [11] found four finger pitch angles (65°, 45°, 25° and 15°) and five finger roll angles (-15°, 0°, 15°, 45° and 90°) from the exploration of the techniques used by the participants, which were included in his main study. This difference may occur given that in the current study, the participants were only asked to touch the targets freely regardless of the accuracy, while in Holz's study [11], the participants had to touch the targets accurately. Holz's findings suggest that users touch the targets precisely by aligning the finger's feature and outline [11].

The current study also completed our conclusions from the previous study [3] with a smaller number of participants. In both studies, the results showed that most of the participants performed the tasks by using their thumbs and index fingers in a combination of 60° pitch and 0° roll angles. The current study emphasized the conclusions of our previous study [3] in exploring the fingers and their postures, which will be included in our main study on fingertip contact area with touch screen devices.

4. Conclusion

This study is a preliminary observational study on the fingertip contact area with touch screen devices. The main purpose of this study was to understand finger postures when interacting with touch screen devices. It has been possible to conclude that the most common thumb and index finger posture used by participants in their interaction with the screen of a mobile device is a combination of 60° pitch and 0° roll angles. The study's results show that the female participants with long fingernails, touched targets on the mobile device screens by using their index or middle fingers, with a very low pitch. In sum, the findings of this study can help researchers determine the finger postures that should be

included in the study of the fingertip contact area with touch screen devices.

Acknowledgments

This project has been funded with support from the European Commission under the scope of the AREAS Project of the Erasmus Mundus program. This publication reflects the views only of the authors, and the Commission cannot be held responsible for any use that may be made of the information contained herein.

This study also had the financial support of FCT-Fundação para a Ciência e Tecnologia of Portugal, under the project ID/CEC/00319/2013.

References

- [1] Karlson, A.K., Bederson, B.B. and Vidal, J.L.C., Understanding single-handed mobile device interaction, Human-Computer Interaction Lab, University of Maryland, 2006.
- [2] Hooper, S., How do users really hold mobile devices? [Online]. [Date of reference October 23th of 2014], Available at: <http://www.uxmatters.com/mt/archives/2013/02/how-do-users-really-hold-mobile-devices.php>
- [3] Umami, M.K., Arezes, P.M. and Sampaio, A.M., Identifying finger postures when interacting with the touch screen of mobile devices. Proceeding of the International Symposium on Occupational Safety and Hygiene (SHO), 2015. pp. 415-417.
- [4] Azenkot, S. and Zhai, S., Touch behavior with different postures on soft smartphone keyboards, Proceeding of the 14th international conference on Human-computer interaction with mobile devices and services (MobileHCI), 2012. pp. 251-260. DOI: 10.1145/2371574.2371612
- [5] Google Developers. Page Speed Insights: Size tap targets appropriately [Online], [Date of reference March 26th of 2015], Available at: <https://developers.google.com/speed/docs/insights/SizeTapTargetsAppropriately>
- [6] Windows Dev. Center. Motion and interaction for windows phone: Interaction and usability with windows phone [Online], [Date of reference March 26th of 2015], Available at: <https://msdn.microsoft.com/en-us/library/windows/apps/hh202889%28v=vs.105%29.aspx>
- [7] iOS Human Interface Guidelines, UI Design Basics: Adaptivity and Layout [Online], [Date of reference March 26th of 2015], Available at: <https://developer.apple.com/library/ios/documentation/UserExperience/Conceptual/MobileHIG/LayoutandAppearance.html>
- [8] Parhi, P., Karlson, A.K. and Bederson, B.B., Target size study for one-handed thumb use on small touchscreen devices, Proceeding of the 8th conference on Human-computer interaction with mobile devices and services (MobileHCI), 2006. pp. 203-210. DOI: 10.1145/1152215.1152260
- [9] Park, Y.S., Han, S.H., Park, J. and Cho, Y., Touch key design for target selection on a mobile device. Proceeding of the 10th international conference on Human computer interaction with mobile devices and services (MobileHCI), 2008. pp. 423-426.
- [10] Wang, F. and Ren, X., Empirical evaluation for finger input properties in multi-touch interaction. Proceeding of the SIG CHI conference on Human factors in computing system (CHI), 2009. pp 1063-1072. DOI: 10.1145/1518701.1518864
- [11] Holz, C. and Baudisch, P., Understanding touch. Proceeding of the SIG CHI conference on Human factors in computing system (CHI), 2011. pp. 2501-2510. DOI: 10.1145/1978942.1979308

M.K. Umami, has a Sarjana Teknik (equal to BSc. Eng) degree in Mechanical Engineering and a MSc. degree in Industrial Engineering. He has been working at the Department of Industrial Engineering, School of Engineering, University of Trunojoyo Madura, Indonesia since 2004. He is

currently a PhD student at the Department of Production and Systems, School of Engineering, University of Minho, Portugal.
ORCID: orcid.org/0000-0002-1431-6060

P. Arezes, has a PhD. in Industrial and Systems Engineering from the University of Minho, Portugal, where he is currently a full professor in Ergonomics and Human Factors. He is also a visiting fellow at MIT's AgeLab in the USA. He leads the Human Engineering research group and coordinates the Engineering Design and Advanced Manufacturing (EDAM) area of MIT's Portugal Program at the University of Minho, where he is also chair of the Steering Board of the PhD program "Leaders for Technical Industries (LTI)".
ORCID: orcid.org/0000-0001-9421-9123

Á.M. Sampaio, is an Industrial Designer, with a MSc in Design, Materials and Product Management, and a PhD in Polymer Engineering, he works on product design and development, more specifically, in the field of user-centered design, additive manufacturing and interaction design. He is a researcher in the HMIEXCEL project: Human Machine Interaction Excellence Project – Critical R&D in the framework of the development and production cycles of advanced multimedia solutions for automobiles, a partnership between the Bosch Car Multimedia Portugal, S.A. and the University of Minho.
ORCID: orcid.org/0000-0002-8635-1682



UNIVERSIDAD NACIONAL DE COLOMBIA

SEDE MEDELLÍN
FACULTAD DE MINAS

Área Curricular de Ingeniería
de Sistemas e Informática

Oferta de Posgrados

Especialización en Sistemas
Especialización en Mercados de Energía
Maestría en Ingeniería - Ingeniería de Sistemas
Doctorado en Ingeniería- Sistema e Informática

Mayor información:

E-mail: acsei_med@unal.edu.co
Teléfono: (57-4) 425 5365

Tackling the challenges of an aging workforce with the use of wearable technologies and the quantified-self

Martin Lavallière ^a, Arielle A. Burstein ^a, Pedro Arezes ^b & Joseph F. Coughlin ^a

^a AgeLab, Massachusetts Institute of Technology, Cambridge, USA. mlav@mit.edu, burstein@mit.edu, coughlin@mit.edu

^b Industrial and Systems Engineering (ISE), University of Minho, Guimarães, Portugal. d140@dps.uminho.pt

Received: December 04th, 2015. Received in revised form: March 10th, 2016. Accepted: April 12th, 2016.

Abstract

The world's population is aging at an unprecedented rate, this demographic shift will change all aspects of life, including work. The aging of the workforce and a higher percentage of workers who will work past traditional retirement years presents significant challenges and opportunities for employers. Older workers are a valuable resource, but in order to ensure they stay in good health, prevention will be key. Wearable technologies are quickly becoming ubiquitous, individuals are turning to them to monitor health, activities and hundreds of other quantifiable occurrences. Wearable technologies could provide a new means for employers to tackle the challenges associated with an aging workforce by creating a wide spectrum of opportunities to intervene in terms of aging employees and extend their working lives by keeping them safe and healthy through prevention. Employers are already making standing desks available, and encouraging lunch time exercise, is it feasible for Wearables to make the jump from a tool for individuals to a method for employers to ensure better health, well-being and safety for their employees? The aim of this work is to lay out the implications for such interventions with Wearable technologies (monitoring health and well-being, oversight and safety, and mentoring and training) and challenges (privacy, acceptability, and scalability). While an ageing population presents significant challenges, including an aging workforce, this demographic change should be seen, instead, as an opportunity rethink and innovate workplace health and take advantage of the experience of older workers. The Quantified-Self and Wearables can leverage interventions to improve employees' health, safety and well-being.

Keywords: Aging, Health and Safety, Monitoring, Personalized intervention, Big Data.

Hacer frente a los desafíos de una fuerza laboral que envejece con el uso de tecnologías usables y la auto-cuantificación

Resumen

La población mundial está envejeciendo a un ritmo sin precedentes. El envejecimiento y un mayor porcentaje de trabajadores que trabajan más allá de los años de jubilación presentan importantes desafíos y oportunidades. Los trabajadores mayores son un recurso valioso, pero a fin de garantizar que permanezcan en buen estado de salud, la prevención será la clave. Tecnologías portátiles, ó wearables, están proporcionando un medio para hacer frente a el envejecimiento mediante la creación de un amplio espectro de oportunidades para intervenir y para prolongar la vida laboral de los colaboradores, manteniendolos seguros y saludables. El objetivo de este trabajo es exponer las implicaciones de este tipo de intervenciones con wearables (Control de salud, vigilancia, seguridad, y formación) y los desafíos (privacidad, aceptabilidad y escalabilidad). Los wearables pueden aprovechar y fortalecer las intervenciones para mejorar la salud, seguridad y el bienestar de los empleados.

Palabras clave: Envejecimiento, salud y seguridad, supervisión, intervención personalizada, big data

1. Introduction

The graying of the Baby Boomer generation, combined with a decrease in birth rates, has resulted in a dramatic shift

in worldwide demographics. In 1950, a large percentage of the population fell in the working age range of 18-64 years. This segment, of work aged adults supported the small population of individuals in their retirement years and the

How to cite: Lavallière, M., Burstein, A.A., Arezes, P. and Coughlin, J.F., Tackling the challenges of an aging workforce with the use of wearable technologies and the quantified-self. DYNA 83(197), pp. 38-43, 2016.

large demographic of children. In comparison, the U.S. Census estimates that the year 2050 presents a very different picture to the past, the demographic of people age 65 and over is projected to reach as high as 20.2% [1]. This shift, often referred to as the squaring of the demographic curve, shows a shrinking percentage of the population of children and those of working age, along with an expanding percentage of the population surviving into what has traditionally been considered the retirement years. Additionally, workers in the U.S. are working longer [2].

The aging of the workforce is a considerable issue worldwide. In Europe, the demographic of 55–64 year olds is predicted to increase by 16.2 % from 2010 to 2030 [3]. In the U.S., 9.9 % of the workforce was over the age of 55 in 2000, and it is expected that this number will increase to 15.3 % of the workforce by 2020 [4]. Such a demographic change has begun to affect many industries, including healthcare, finance, insurance, food, durable goods, electronics, the automotive industry, etc. Despite markets having recognized this demographic shift as an opportunity for products and services for an aging demographic also known as the “longevity economy” [5], less is known about how this will impact the future workforce. Industries need to react to changes not only in their consumer base to maintain market share, but perhaps more importantly, in their own workforce. In the discipline of human factors and ergonomics aging is considered a risk factor rather than an opportunity to retain skilled employees [6].

A review of epidemiological literature on the causes, and types of injuries and their related costs for the construction industry by age showed that workers age 50 and over, in fact, had a lower frequency of workplace injuries than younger workers, although the older workers had higher injury costs [7]. Similarly, the rate of injuries for 2012 with paid disability leave from work due to injuries by age range [8] showed that workers age 65 and over have a much lower rate of injuries with paid leave from work than the rest of the workforce. However, the length of disability leave for older workers tended to be longer, and they experienced a much higher fatality rate than younger workers [8].

On top of a leave of absence due to injuries or an accident in the workplace, it has been shown that health limitations appear to increase the likelihood of all types of retirement, suggesting that the beginnings of poor health are present before the decision to retire is made [9,10]. Therefore, it is of utmost importance to maintain or augment employees’ health to keep them healthy and safe and, as such, able to keep working. In a literature review, physical activity has been shown to help alleviate the negative impact of age on the body and the mind [11]. On the way to old age, there are landmark birthdays, which mandate certain screenings and tests for preventive purposes (e.g., the aviation and the trucking industry). The workforce may need to adopt similar practices in terms of health monitoring, be it checkups in real time or at intervals like the medical model in order to keep their older workforces healthy and working.

The technological developments in the portability (i.e., miniaturization), precision and accuracy of different types of health related sensors (e.g., heart rate monitors) and other sensors have transformed the realm of everyday calculability. These changes have come with an increased capacity to gather data about one’s performance and the capacity to track one’s

personal responses (i.e., emotional, physiological, etc.) to the environment. This movement known as the Quantified-Self can be characterized as self-monitoring or self-sensing using diverse types of Wearable sensors combined with the use of computing [12]. The Pew Research Center [13] showed that 60% of individuals surveyed recorded some type of information about their health (e.g., weight, diet, exercise routines or sleep patterns). For those who keep records of this information, methods vary considerably. Some people use electronic forms, while others use a pen and paper. The Wearables and Quantified-Self movement have enabled individuals to understand their personal life better through their health data, and such an approach could allow different industries to obtain a clearer picture of their employees through their lifespan [14].

Even though many ethical questions remain regarding their place in work settings [15], the use of Wearables technologies and Quantified-Self data represents an interesting opportunity to tackle the challenges associated with the aging of the workforce, both in terms of protecting and improving the health of employees and capturing information on working procedures to prevent the loss of knowledge due to retirement.

The aim of this work is to lay out the implications of using Wearable technologies and Quantified-Self data in the workplace to tackle the challenges of an aging workforce.

2. Methods

This work aims to expand an initial review that was conducted on the use of Quantified-Self and Wearable technologies in the workplace (for more details see Lavallière et al. [16]). The implications (Monitoring Health and Well-Being, Oversight and Safety, and Mentoring and Training) and challenges (Privacy and Acceptability, Scalability, Creativity) were addressed for a broader audience of health and safety practitioners, workers and industries. To extend this previous work and its relevance to the aging of the workforce, a similar structure will be used that will emphasize specific topics of these particular areas that would benefit the older employee. A narrative literature review was conducted to identify the main aspects that companies should take into consideration on how Wearables technologies can help to improve employees’ health, safety and well-being.

3. Implications

3.1. Monitoring health, safety, and well-being

In line with one’s ability to complete work, numerous reports have been made on the importance of health in the workplace [17]. The use of Wearables and Quantified-Self in such a setting allows for two types of interventions: identification of tasks that may involve risks for older employees and interventions in the workplace aimed at improving or maintaining employee health.

3.1.1. Identification of at risk situations for older employees

Based on the recent advances in technology development to collect multiple types of signals, all this new data can now

be gathered, recorded and tracked during work to gather information such as the level of physical or cognitive fatigue. Such physiological indicators include heart rate, electromyography [18] and stress level.

An example of such an application would be the mining industry. An investigation of the most common type of accident recorded in subsurface and surface mining in Spain throughout the 2003–2008 period showed “Physical over-exertion on the muscular-skeletal system” as the primary reason for accidents at work [19]. This result showed that over-exertion was the main descriptor for 21.7% (out of 25,362 accidents) and 28.3% (out of 28,094 accidents) in sub-surface and surface mining, respectively.

Recent technological development has allowed for smaller and smarter devices to detect specific events or levels of fatigue. Some have used a Smart Safety Helmet (SSH), which combines inertial measurement units (IMU) and electroencephalography (EEG) sensors to detect an operator’s level of fatigue [20]. The SSH sends a warning signal using a haptic device to alert the operator when a computed risk level (i.e., fatigue, high stress, or error) has reached a certain threshold.

Instantaneous measurement of physiological variables and environmental characteristics (e.g., light, noise, temperature, vibration) would enable decisions based on someone’s physical state and, enable employers to identify risk factors such as particular working conditions or fatigue onset among employees prior to reaching a state of over-exertion. Moreover, gathering information over time with such devices would enable industries to better evaluate the procedures involved in job requirements.

3.1.2. Interventions in the workplace aimed at improving employees’ health.

3.1.2.1. Modifying the work environment

As one ages, loss of physical strength and endurance can indicate potential declines in physical capacity [1]. There is also evidence that cognitive performance does not generally show any marked decrease until after 70 years of age [21]. Therefore, it is important to verify that the possible interventions implemented in the workplace are efficient at reducing the physical and cognitive burden of one’s job. By allowing individualized data collection using Wearables, information regarding workplaces adaptations and/or the use of mechanically efficient techniques to compensate for age decline can be gathered and analyzed to evaluate their efficacy and efficiency.

3.1.2.2. Health interventions in the workplace

Workplace health and wellness interventions have received a considerable amount of attention (for more details consult The NIOSH Total Worker Health™ Program [17]). These approaches integrate occupational safety, workplace health protection and promotion initiatives through ergonomics to prevent worker injury and illness and to promote worker health and well-being [22,23]. It has been recognized that to be successful and beneficial, these

approaches require the full participation of the employee [24]. Factors, such as higher attendance (i.e., greater participation in terms of number of employees) and better assessment of the level of involvement (i.e., intensity of one’s activity measured by a physiological value such as heart rate) with improved recording over time is needed to achieve significant results in augmenting worker productivity. Self-monitoring has been found to improve chronic conditions such as high blood pressure, excessive weight, and diabetes [25]. Moreover, two behaviors—healthful food consumption [26] and restful sleep patterns—which can be improved with Wearables and the Quantified-Self-concept, [27,28] have also been shown to greatly influence health. Since chronic health problems have been associated with decreased work ability and, to lower productivity at work, it is beneficial for employers to prevent chronic health conditions among their staff [29]. It is important to note that these programs benefit employee health throughout workers’ lifespans and not only once they are senior workers [30]. In sum, the use of Wearables, when embedded in a health promotion program, can help to improve employees’ health by reducing the prevalence of chronic conditions among employees and helping them to adopt and/or maintain a healthier lifestyle.

3.1.2.2. Return to work

In the case where all these interventions would have been insufficient in preventing employee’s injury, Wearables, and their associated data can be used to facilitate the employee’s return to work after sick leave. Results have shown that the longer an employee remains away from work due to disability or work related injuries, the less chance they have of returning [31,32]. Therefore, efforts should be made to guarantee the best treatment possible during sick leave. It has been shown that a formal follow-up with employees while on disability leave [33,34] and a proper rehabilitation program, on-site or using tele-rehabilitation [35] can greatly influence return to work programs and, as such, prevent prolonged disability. The integration of Wearable and ambient sensors has been suggested for home monitoring while undergoing treatment [14,36] and would increase accessibility to rehabilitation programs.

Overall, there are definite benefits to the inclusion of Wearables in health interventions aimed at either preventing risky situations in the workplace, improving employee health care through health promotion programs, or by improving the rehabilitation process if one is injured on the job.

3.2. Oversight and Safety

Different approaches have been used to monitor workers in the workplace in the past. Some have even made the assumption that we are transitioning from a world of surveillance to sousveillance [37]. Sousveillance has been described as the recording of an activity by an individual using a small Wearable or portable personal technology [38]. Such approaches allow for a direct monitoring of the working environment from an employee’s perspective (i.e., recording from first person perspectives - sousveillance), instead of the entire working environment (surveillance).

Systems can also be introduced to allow for remote monitoring of employees using an approach similar to *sousveillance* in combination with a GPS and personal emergency response system (PERS) to allow them to call for assistance at the push of a button or automatic emergency response.

The above mentioned the application of Wearables, and Quantified-Self data focused on tracking information regarding the health status of employees at work or while in recovery. With the aging of the population and the baby boomer generation quickly approaching retirement, multiple industries and job sectors are not only facing a shortage of workers, they are also at high risk of knowledge loss when these highly experienced and knowledgeable employees leave (e.g., oil, gas, and nuclear power facilities)[39].

3.3. *Mentoring and training*

While employers should assist their older employees in having long and healthy working lives, they do eventually leave. Employee knowledge can be captured and passed on by tapping into the underlying theme of human dynamics research. Pentland et al. [40-42] have suggested that people are social learners, so arranging work to increase productive face-to-face communication yields measurable benefits.

Wearable devices can be utilized to show a less experienced employee how a task should be performed. For example, recording devices such as smart glasses or a miniaturized video that gives a first person perspective (head, chest or shoulder mounted) would allow more experienced employees to gather information on how they navigate and interact in a work environment. Combining this video collection with a voice-over procedure of their execution of a task will not only enable the collection of precious information from experienced employees but also, help to identify whether there are disruptions in procedures that should be fixed. These procedures—used to gather employee knowledge—should serve all individuals as communicating vessels so older employees can also learn from younger colleagues. This type of approach aligns with Naegele & Walker's [43] work suggesting that when it comes to training employees, an age-neutral approach is particularly appropriate.

4. Challenges

4.1. *Privacy and Acceptability*

The use of more moment-to-moment information to improve the workplace would ideally lead to better management, and improved workplace conditions. All the information gathered by Wearables should be available for all parties involved—the employee as well as the employer—in order to be meaningful and prevent the stigma of technology as something that restricts autonomy.

4.1.1. Technology adoption among older workers

There is clear evidence that sensory abilities, such as visual [44] and auditory acuity [45,46], generally deteriorate with age. These declines can generally be compensated for by corrective technology, for example the use of personal

aids such as corrective lenses or a hearing aid, or by adaptations to the environment such as improved lighting, reduced glare or reduced background noise. Therefore, Wearable technologies have to be designed to be inclusive without raising a flag as a device for an older or impaired individual. As reported by Coughlin [47], the baby boomers represent a new generation of consumers. They have experienced mobility, new technology, and a growing trend in improvements in every field in their lives [48]. They therefore expect to continue this type of lifestyle and will demand more of industries [49]. They will not be drawn to what could be considered an “older worker Wearable.” Although the hope is that monitoring technologies will allow older adults to live independently and age in place, some may feel embarrassed and see technology as an admission of dependence on others. For example, patients, particularly older adults, are concerned about privacy, unfamiliarity with technology, and a potential decrease in social contact [50]. Therefore, great efforts and research should be undertaken in the domain of privacy concerns and willingness to use these devices among older individuals.

Sensors are becoming amazingly accurate, but Wearable products continue to be clunky and provide a poor user experience. In fact, studies have found that 40 percent of consumers who purchase a Wearable fitness tracker leave it sitting on their bedside after a month or two of wearing it. It is of utmost importance to overcome the barriers to the adoption of such technologies and recognize the potential benefits that they can deliver to older and younger adults alike.

4.2. *Scalability*

Such devices and quantified data have to provide insightful and accessible information to the user for an agreeable and compelling experience. Charts, graphs, and statistics can automatically be compiled, transforming what is essentially a large database of meaningless numbers into something that users can quickly parse and understand. This type of data translation is necessary to pass from a continuous data collection paradigm to information that is understandable and actionable [51,52].

4.2.1. Oligarchy

Implementations will have to rely on widespread distribution and usage among their industries to be successful. Moreover, great care should be taken that these tools are not only accessible to the wealthy. Open source code and the Internet of Things should alleviate such hurdles since cloud computing and crowdsourcing platforms will benefit us all by allowing the use of developments and tools created by others for free.

5. Conclusions

Instead of being viewed as a challenge, this upcoming aging workforce should be seen as an opportunity to review, rethink and reinvent how we cope with the aging workforce. The use of Quantified-Self information and Wearables

represents a new way to tackle the challenges associated with an aging workforce by creating a wide spectrum of opportunities to intervene with aging employees to extend their working lives by keeping them safe and healthy. However, to be effective, products and/or programs have to engage workers and have a meaningful impact on one's life in order to be successful.

The creation of an organization with a reputation for being an ethical, trustworthy, and a good place to work is likely to deliver additional benefits across all generations of employees [53].

Acknowledgement

Martin Lavallière was supported by a postdoctoral research grant - Recherche en sécurité routière : Fonds de recherche du Québec - Société et culture (FRQSC), Société de l'assurance automobile du Québec (SAAQ), Fonds de recherche du Québec - Santé (FRQS).

This work was partially developed with the financial support of the Luso-American Development Foundation - FLAD, through the research grant ref. rv14022, and of the MIT Portugal Program.

References

- [1] Administration on Aging. [Online]. 2010, Projected future growth of the population. Available at: http://www.aoa.gov/AoARoot/Aging_Statistics/future_growth/future_growth.aspx
- [2] Bureau of Labor Statistics, Labor Force Projections to 2020: A More Slowly Growing Workforce, Monthly Labor Review, pp. 43-64, 2012.
- [3] Ilmarinen, J., Towards a longer worklife. Ageing and the quality of worklife in the European Union, Finnish Institute of Occupational Health, Ministry of Social Affairs and Health, Helsinki, Finland 2006.
- [4] Toossi, M., A century of change: The U.S. labor force, 1950–2050, Monthly Labor Review, 135 pp. 15-28, 2002.
- [5] Irving, P. and Chatterjee, A. The Longevity Economy: From the elderly, a new source of economic growth. [Online] 2013. [Date of reference: June 10th of 2015]. Available: <http://www.forbes.com/sites/realspin/2013/04/02/the-longevity-economy-from-the-elderly-a-new-source-of-economic-growth/>
- [6] Crawford, J.O., Graveling, R.A., Cowie, H.A. and Dixon, K., The health safety and health promotion needs of older workers, Occup Med, 60(3), pp. 184-92, 2010. DOI: 10.1093/occmed/kqq028
- [7] Schwatka, N.V., Butler, L.M. and Rosecrance, J.R., An aging workforce and injury in the construction industry, Epidemiol Rev, 34(1), pp. 156-67, 2012. DOI: 10.1093/epirev/mxr020
- [8] Bureau of Labor Statistics (BLS). [Online]. 2014. Available: <http://data.bls.gov/gqt/InitialPage>
- [9] Henretta, J.C., Chan, C.G. and O'Rand, A.M., Retirement reason vs. retirement process: Examining the reasons for retirement typology, The Journal of Gerontology: Social Sciences, 47, pp. S1-S7, 1992.
- [10] Midanik, L., Soghikian, K., Ransom, L. and Tekawa, I., The effect of retirement on mental health and health behaviours: The Kaiser Permanente retirement study, Journal of Gerontology, vol. 50B pp. S59-S61, 1995. DOI: 10.1093/geronb/50B.1.S59
- [11] Bherer, L., Erickson, K.I., and Liu-Ambrose, T., A review of the effects of physical activity and exercise on cognitive and brain functions in older adults, J Aging Res, vol. 2013, p. 1-8, 2013. DOI: 10.1155/2013/657508
- [12] Mann, S., Smart clothing: The wearable computer and wearcam, Personal Technologies, 1(1), pp. 21-27, 1997. DOI: 10.1007/BF01317885
- [13] Pew Research Center, Tracking for Health, Washington, D.C. 2013.
- [14] Patel, S., Park, H., Bonato, P., Chan, L. and Rodgers, M., A review of wearable sensors and systems with application in rehabilitation, J Neuroeng Rehabil, 9(21), pp. 1-17, 2012. DOI: 10.1186/1743-0003-9-21
- [15] West, J.P. and Bowman, J.S., Electronic surveillance at work: An ethical analysis, Administration & Society, pp. 1-24, 2014. DOI: 10.1177/0095399714556502
- [16] Lavallière, M., Arezes, P., Burstein, A. and Coughlin, J.F., The Quantified-Self and wearable technologies in the workplace: Implications and challenges, in: International Symposium on Occupational Safety and Hygiene SHO2015, Guimarães, Portugal, 2015, pp. 161-163.
- [17] NIOSH, Research Compendium; The NIOSH Total Worker Health™ Program: Seminal Research Papers 2012.
- [18] Van-Eerd, D., Hogg-Johnson, S., Cole, D.C., Wells, R. and Mazumder, A., Comparison of occupational exposure methods relevant to musculoskeletal disorders: Worker-workstation interaction in an office environment, J Electromyogr Kinesiol, 22(2), pp. 176-85, 2012. DOI: 10.1016/j.jelekin.2011.12.001
- [19] Pera, L.S., Vintro, C. and Freijo, M., Characteristics of the 3 most common types of occupational accident in spanish sub-surface and surface mining, from 2003–2008, DYNA, 79(172), pp. 118-125, 2012.
- [20] Li, P., Meziane, R., Otis, M.J.-D., Ezzaidi, H. and Cardou, P., A smart safety helmet using IMU and EEG sensors for worker fatigue detection, in: The IEEE International symposium on Robotic and Sensors Environments (ROSE) Timisoara, Romania 2014. DOI: 10.1109/rose.2014.6952983
- [21] Fisk, A.D., Rogers, W.A., Charness, N., Czaja, S.J. and Sharit, J., Designing for older adults: Principles and creative human factors approaches, 2nd ed. Boca Raton, FL: CRC Press, USA, 2009. DOI: 10.1201/9781420080681
- [22] Hymel, P.A., Loeppke, R.R., Baase, C.M., Burton, W.N., Hartenbaum, N.P., Hudson, T.W., et al., Workplace health protection and promotion: A new pathway for a healthier--and safer--workforce, J Occup Environ Med, 53(6), pp. 695-702, 2011. DOI: 10.1097/JOM.0b013e31822005d0
- [23] Henning, R., Warren, N., Robertson, M., Faghri, P., Cherniack, M., and Team, C.-N.R., Workplace health protection and promotion through participatory ergonomics: An integrated approach, Public Health Rep, 124(1), pp. 26-35, 2009.
- [24] Pereira, M.J., Coombes, B.K., Comans, T.A. and Johnston, V., The impact of onsite workplace health-enhancing physical activity interventions on worker productivity: A systematic review, Occup Environ Med, 72(6) pp. 401-412, 2015. DOI: 10.1136/oemed-2014-102678.
- [25] Driscoll, K.A. and Young-Hyman, D., Use of technology when assessing adherence to diabetes self-management behaviors, Curr Diab Rep, 14(9), pp. 521, 2014. DOI: 10.1007/s11892-014-0521-1.
- [26] Spence, J.C., Cutumisu, N., Edwards, J., Raine, K.D. and Smoyer-Tomic, K., Relation between local food environments and obesity among adults, BMC Public Health, 9(192), pp. 1-6, 2009. DOI: 10.1186/1471-2458-9-192
- [27] Kim, Y., Wilkens, L.R., Schembre, S.M., Henderson, B.E., Kolonel, L.N. and Goodman, M.T., Insufficient and excessive amounts of sleep increase the risk of premature death from cardiovascular and other diseases: the Multiethnic Cohort Study, Prev Med, 57(4), pp. 377-385, 2013. DOI: 10.1016/j.ypmed.2013.06.017
- [28] Xiao, Q., Keadle, S.K., Hollenbeck, A.R. and Matthews, C.E., Sleep duration and total and cause-specific mortality in a large US cohort: Interrelationships with physical activity, sedentary behavior and body mass index, Am J Epidemiol, 180(10), pp. 997-1006, 2014. DOI: 10.1093/aje/kwu222
- [29] Leijten, F.R., van den Heuvel, S.G., Ybema, J.F., van der Beek, A.J., Robroek, S.J. and Burdorf, A., The influence of chronic health problems on work ability and productivity at work: A longitudinal study among older employees, Scand J Work Environ Health, 40(5), pp. 473-482, 2014. DOI: 10.5271/sjweh.3444
- [30] Vilela, B.L., Benedito-Silva, A.A., de Lira, C.A. and Andrade-Mdos, S., Workplace exercise and educational program for improving fitness outcomes related to health in workers: A randomized controlled trial,

- J Occup Environ Med, 57(3), pp. 235-240, 2015. DOI: 10.1097/JOM.0000000000000393
- [31] Lederer, V., Rivard, M. and Mechakra-Tahiri, S.D., Gender differences in personal and work-related determinants of return-to-work following long-term disability: A 5-year cohort study, J Occup Rehabil, 22(4), pp. 522-531, 2012. DOI: 10.1007/s10926-012-9366-0
- [32] Berecki-Gisolf, J., Clay, F.J., Collie, A. and McClure, R.J., Predictors of sustained return to work after work-related injury or disease: Insights from workers' compensation claims records, J Occup Rehabil, 22(3), pp. 283-291, 2012. DOI 10.1007/s10926-011-9344-y
- [33] McLellan, R.K., Pransky, G. and Shaw, W.S., Disability management training for supervisors: a pilot intervention program, J Occup Rehabil, 11(1), pp. 33-41, 2001. DOI: 10.1023/A:1016652124410
- [34] Shaw, W.S., Robertson, M.M., McLellan, R.K., Verma, S. and Pransky, G., A controlled case study of supervisor training to optimize response to injury in the food processing industry, Work, 26(2), pp. 107-114, 2006.
- [35] Tousignant, M., Moffet, H., Boissy, P., Corriveau, H., Cabana, F. and Marquis, F., A randomized controlled trial of home telerehabilitation for post-knee arthroplasty, Journal of Telemedicine and Telecare, 17(4), pp. 195-198, 2011. DOI: 10.1258/jtt.2010.100602
- [36] Callejas-Cuervo, M., Diaz, G.M. and Ruiz-Olaya, A.F., Integration of emerging motion capture technologies and videogames for human upper-limb telerehabilitation: A systematic review, DYNA, 82(189), pp. 68-75, 2014. DOI: 10.15446/dyna.v82n189.42066
- [37] Mann, S., Nolan, J. and Wellman, B., Sousveillance: Inventing and using wearable computing devices for data collection in surveillance environments, Surveillance & Society, 1(3), pp. 331-355, 2003.
- [38] Fernback, J., Sousveillance: Communities of resistance to the surveillance environment, Telematics and Informatics, 30(1), pp. 11-21, 2013. DOI: 10.1016/j.tele.2012.03.003
- [39] Deloitte, The talent crisis in upstream oil & gas strategies to attract and engage Generation Y, 2005.
- [40] de Montjoye, Y.A., Stopczynski, A., Shmueli, E., Pentland, A. and Lehmann, S., The strength of the strongest ties in collaborative problem solving, Sci Rep., 4, pp. 1-6, 2014. DOI: 10.1038/srep05277
- [41] Olguin, D.O., Waber, B.N., Kim, T., Mohan, A., Ara, K. and Pentland, A., Sensible organizations: Technology and methodology for automatically measuring organizational behavior, IEEE Transactions on Systems, Man and Cybernetics—Part B: Cybernetics, 39(1), pp. 43-55, 2009. DOI: 10.1109/TSMCB.2008.2006638
- [42] Pentland, A., Honest signals: How they shape our world. Cambridge, MA: MIT Press, USA, 2008.
- [43] Naegel, G. and Walker, A., A guide to good practice in age management, European Foundation for the Improvement of Living and Working Conditions, University of Dortmund, Germany and University of Sheffield, UK, 2006.
- [44] National Center for Chronic Disease Prevention and Health Promotion. The burden of vision loss. [Online]. 2009. Available at: http://www.cdc.gov/visionhealth/basic_information/vision_loss_burden.htm
- [45] Cruickshanks, K., Wiley, T., Tweed, T., Klein, B.E., Klein, R., Mares-Perlman, J., et al., Prevalence of hearing loss in older adults in Beaver Dam, Wisconsin, American Journal of Epidemiology, 148(9), pp. 879-886, 1998. DOI: 10.1093/oxfordjournals.aje.a009713
- [46] Strawbridge, W., Wallhagen, M., Shema, S. and Kaplan, G., Negative consequences of hearing impairment in old age: A longitudinal analysis, The Gerontologist, 40(3), pp. 320-326, 2000. DOI: 10.1093/geront/40.3.320
- [47] Coughlin, J.F., Longevity, lifestyle and anticipating the new demands of aging on the transportation system, Public Works Management & Policy, 13(4), pp. 301-311, 2009. DOI: 10.1177/1087724x09335609
- [48] Pak, C. and Kambil, A., Over 50 and ready to shop: Serving the aging consumer, Journal of Business Strategy, 27(6), pp. 18-28, 2006. DOI: 10.1108/02756660610710319
- [49] Coughlin, J.F., Speaking Silver: Lessons for product innovation & development in an aging marketplace, MIT AgeLab, Cambridge, MA, USA, 2007.
- [50] Lee, C., Adoption of smart technology among older adults: Challenges and Issues, Public Policy & Aging Report, 24(1), pp. 14-17, 2014. DOI: 10.1093/ppar/prt005
- [51] Swan, M., Emerging patient-driven health care models: An examination of health social networks, consumer personalized medicine and quantified self-tracking., International Journal of Environmental Research and Public Health, 6(2), pp. 492-525, 2009. DOI: 10.3390/ijerph6020492
- [52] Swan, M., The quantified self: Fundamental disruption in big data science and biological discovery, Big Data, 1(2), pp. 85-99, 2013. DOI: 10.1089/big.2012.0002
- [53] Hewitt Associates, Best employers to work for in Australia study 2000. Summary of findings, defence personnel executive, directorate of strategic personnel planning and research, Australian defence force, Canberra, Australia 2000.

M. Lavallière, completed his BSc., MSc. and PhD. in kinesiology at Université Laval, Québec, Canada. His PhD. work evaluated whether an in-simulator-based training program combined with specific feedbacks can improve on-road driving behaviors in older drivers. His postdoctoral work at the Massachusetts Institute of Technology AgeLab focused on the impact on driving performance of aging, navigation and communication technologies and active safety system in vehicles. His works also concern the development of a Multigenerational Workforce evaluation program to assess the “age readiness” of industries.
ORCID ID: 0000-0003-0247-0308

A.A. Burstein, is a researcher at the Massachusetts Institute of Technology AgeLab who translates demographic change into innovative social research. She holds a Bachelor of Arts degree in International Relations and Hispanic Studies from Wheaton College, Massachusetts. Her work looks at the social and behavioral aspects of aging to try to improve engagement, and develop new services and technology in the areas of health and wealth. Her research interests include financial decision-making and stress, technology adoption, and the role of empathy in professional advice and caregiving.
ORCID ID: 0000-0002-1683-4056

P.M. Arezes, completed his PhD. in Industrial and Systems Engineering at University of Minho, Portugal. He is a full professor in Ergonomics and Human Factors at the School of Engineering of the University of Minho, Portugal. He is also a visiting fellow at MIT's AgeLab in the USA. He leads the Human Engineering research group and he is also the coordinator at UMinho of the Engineering Design and Advanced Manufacturing (EDAM) area of the MIT Portugal Program, and the chair of the steering board of the PhD. program “Leaders for Technical Industries (LTI)” at University of Minho. He is the (co)author of more than 60 papers published in peer-reviewed international journals, as well as a member of the editorial board of more than 15 international scientific journals.
ORCID ID: 0000-0001-9421-9123

J.F. Coughlin, is the Director of the Massachusetts Institute of Technology AgeLab. He is a graduate of the State University of New York at Oswego, Brown University and Boston University. His research provides insights into how demographic change, technology, social trends and consumer behavior will converge to drive future innovations in business and government. He teaches policy and systems innovation in MIT's Department of Urban Studies & Planning as well as the Sloan School of Management Executive Education Program.
ORCID ID: 0000-0001-5288-2218

Analysis and characterization of starchy and cellulosic materials after enzymatic modification

Julián A. Quintero ^a, Javier A. Dávila ^b, Jonathan Moncada ^c, Oscar H. Giraldo ^d & Carlos A. Cardona ^e

^a Escuela de Ingeniería Bioquímica, Pontificia Universidad Católica del Valparaíso, Valparaíso, Chile. julian.quintero@ucv.cl

^b Chemical Engineering Program, Universidad Jorge Tadeo Lozano, Bogotá, Colombia. javier.davilar@utadeo.edu.co

^c Universidad Nacional de Colombia, sede Manizales, Colombia. jmoncadab@unal.edu.co

^d Department of Physics and Chemistry, Universidad Nacional de Colombia, sede Manizales, Colombia. ohgiraldo@unal.edu.co

^e Department of Chemical Engineering, Universidad Nacional de Colombia, sede Manizales, Colombia. ccardona@unal.edu.co

Received: March 21th, 2014. Received in revised form: November 5th, 2015. Accepted: December 20th, 2015.

Abstract

This study concerns the application of enzymatic hydrolysis to modifying starchy and cellulosic materials. Corn, cassava, upright elephant ear and sugarcane bagasse were used in order to evaluate the influence of the enzymatic hydrolysis over the structural, thermal stability and crystallinity of these materials. Differential Scanning Calorimetry (DSC) was used to evaluate the thermal properties while Scanning Electron Microscope (SEM), X-ray Diffraction (XRD) and Brunauer, Emmett and Teller (BET) methods were used for the structural and morphological analysis. Corn presented the highest starch yield (g of starch/g of raw material) with 40.4%. For all materials used, the crystallinity increased due to enzymatic hydrolysis suggesting that amorphous zones are attacked first. The gelatinization temperature of the starches increased as the crystallinity increased. The granules and fibers, except for upright elephant ear starch, did not change their size with enzymatic treatment and the superficial area did not increase significantly with the enzymatic treatment.

Keywords: Starch, Cellulose, Upright elephant ear, Cassava, Sugarcane bagasse, enzymatic hydrolysis, structural analysis, thermal analysis.

Análisis y caracterización de materiales amiláceos y celulósicos después de modificación enzimática

Resumen

Este estudio se enfoca en la modificación por hidrólisis enzimática de materiales amiláceos y lignocelulósicos. Maíz, yuca, bore y bagazo de caña de azúcar fueron usados para evaluar la influencia de la hidrólisis enzimática sobre las propiedades estructurales y térmicas de estos materiales. Calorimetría de barrido diferencial (DSC) fue usado para evaluar estabilidad térmica y cristalinidad, mientras métodos como Microscopía electrónica de barrido (SEM), difracción de rayos X (XRD) y Brunauer-Emmett-Teller (BET) fueron usados para análisis estructurales y morfológicos. El maíz presentó el rendimiento (g de almidón/g de materia prima) más alto con 40.4%. Para todos los materiales usados, la cristalinidad aumentó debido a la hidrólisis enzimática sugiriendo que las zonas amorfas son atacadas primero. La temperatura de gelatinización de los almidones incrementó conforme aumentó la cristalinidad. Los gránulos y fibras, excepto para el bore, no cambiaron su tamaño con el tratamiento enzimático y el área superficial no incrementó significativamente con la hidrólisis enzimática.

Palabras clave: Almidón, celulosa, bore, yuca, bagazo de caña de azúcar, hidrólisis enzimática, análisis estructural, análisis térmico.

1. Introduction

Starchy and cellulosic materials are important feedstocks in the agroindustrial sector worldwide because they turn into a

platform for sugar production and subsequently other important products. Crops such as corn, cassava, potato and tubers in general are good sources of starch, which can potentially be used in the commercial production of ethanol fuel [1, 2].

How to cite: Quintero, J.A., Dávila, J.A., Moncada, J., Giraldo, O.H. and Cardona, C.A., Analysis and characterization of starchy and cellulosic materials after enzymatic modification. DYNA 83(197), pp. 44-51, 2016.

Sugarcane, on the other hand, is one of the principal sources of cellulosic material. Both, starchy and cellulosic materials should be modified to increase their yield in the enzymatic hydrolysis. Starch has traditionally been hydrolyzed by acids, but the specificity of the enzymes, and the absence of secondary reactions among other advantages mean that enzymatic hydrolysis an excellent technique for such processes [3,4].

Both, starchy and cellulosic materials can be hydrolyzed with different types of enzymes. Materials such as corn stover, poplar, eucalyptus, corn cob, rice straw, wheat straw among others have been hydrolyzed using enzymes such as *endoglucanases*, *cellulases* and *xylanases* obtaining good yields [4]. The enzyme α -amylase has been used to carry out the hydrolysis of materials such as cornstarch and rice starch reaching yields above of 90% [5,6]. The sugar yield of the enzymatic hydrolysis of starchy and cellulosic materials can be defined by some important characteristics such as particle size, crystallinity, and accessible surface area [7]. These key parameters are defined for the feedstock characteristics such as the nature of the enzyme, its temperature, pH, and the operating conditions used in the enzymatic hydrolysis.

Starch is made up of two types of glucose polymers: amylose that is essentially a linear chain molecule, and amylopectin, which is branched. Amylose and amylopectin are packed into granules, representing the crystalline and amorphous parts of the structure. There is evidence that crystals consist of parallel double helices formed from the short chains of amylopectin. Two types of crystals or polymorph structures (A and B) have been identified in starch granules, which can be distinguished by the packing density of the double helices [8].

Sugarcane bagasse is a polydisperse particulate, which is present as fiber and rind particles. This material has high length-width ratios and corresponds mostly to stalk fibro-vascular bundles. Sugarcane bagasse matter is mostly made up of cells walls. Thus, bagasse particles have cavities that correspond to cell lumina. Besides, there are pith particles, which are finer and near unitary length-width ratio originated from stalk parenchyma [9].

Different techniques have been used to analyze some of the important characteristics of starch and cellulosic materials. Scanning Electron Microscope (SEM) has been used to relate granule morphology to starch genotype and sugarcane sizes [9, 10]. Shapes of starchy and cellulosic materials vary widely depending on the source. Generally, oval, elliptic, spherical and irregular shapes predominate among all the granule shapes. X-ray Diffraction (XRD) has been used to reveal the presence and characteristics of the crystalline structure of the starch granules and the patterns of sugarcane bagasse [11,12]. Differential Scanning Calorimetry (DSC) is used to measure changes of physicochemical properties in gelatinized starch and to measure the rate and extent of starch recrystallization [13]. The specific surface area of the starch granules is measured using Brunauer-Emmett-Teller (BET) technique [14].

Given that enzymatic hydrolysis is an interesting biorefining alternative to obtain a sugar-based platform from starchy and cellulosic materials, the analysis of the effect of enzyme treatment on specific material characteristics is of relevant importance to finding suitable raw materials and sustainable technologies. In this sense, this work shows a primary technical analysis of different starchy and cellulosic materials after enzyme treatment and their potential as future biorefinery feedstocks. Taking into account the

importance of factors such as particle size, crystallinity and accessible surface area in starchy and cellulosic materials, this paper also presents a study of the characterization of native and enzyme-treated starchy and cellulosic materials after enzymatic modifications. Corn, cassava and upright elephant ear are studied for starch recovery while, sugarcane bagasse is considered a source of fiber with a high cellulose content. Structural and thermal properties were compared for each material in its native and enzyme-treated states.

2. Materials and methods

2.1. Raw materials

Cassava (*Manihot esculenta*) and yellow dent corn (*Zea mays*) were obtained from a local Colombian market. Sugarcane bagasse (*Saccharum*) was obtained from a small mill, which produces commercial sugarcane juice. Upright elephant ear roots (*Alocasia macrorrhiza*), were collected from a farm belonging to the National Service of Learning (SENA) in Manizales-Colombia (center-west of Colombia). These raw materials were thoroughly washed with enough water to remove impurities. Cassava and upright elephant ear were manually peeled, and then processed for starch recovery. Corn grains were separated from the corncobs and subsequently processed for starch recovery. While, sugarcane bagasse was milled in a mill blade and the obtained fibers were dried (for 48 h at 75 °C) and sieved.

2.2. Starch recovery

Cassava and upright elephant ear starch were recovered by the conventional method used in tropical countries. This method consists of root washing, peeling, milling, and filtering [15]. The starch solution obtained from filtering was decanted for 24 h, after which, the residual liquid was extracted and the solid starch was dried at 60 °C to constant weight.

Starch from corn was recovered using two methods. In the first method (M1) exposed by [16], starch recovery was calculated as the ratio of the weight of starch recovered from wet milling to the total weight of the starch present in the corn. To give a better idea of the work undertaken, Fig. 1 shows the experimental procedure followed to recover starch. In the second method (M2), lactic acid and sodium bisulfite were used for the steeping step, and the starch content was measured according to the method reported by [17].

2.3. Enzymatic hydrolysis

The hydrolysis of starch from cassava, upright elephant ear and corn was conducted with α -amylase (HT-340L, Bacterial source: *Bacillus licheniformis*; supplied by Proenzimas, Cali, Colombia), which allows a partial hydrolysis under special conditions due to its action on starch polymers. On the other hand, Cellulase (Celulase CE 2, Fungal source: *Trichoderma longibrachiatum*; supplied by Proenzimas, Cali, Colombia) was used for sugarcane bagasse modification. Before hydrolysis, all materials were dried in an oven for 48 h at 60 °C [18]. Dried starches were milled in a ball mill (KM1, Germany) obtaining a dry powder. Enzymatic hydrolysis was carried out in a shaker (UNITRONIC OR, J.P. Selecta, S.A., Barcelona).

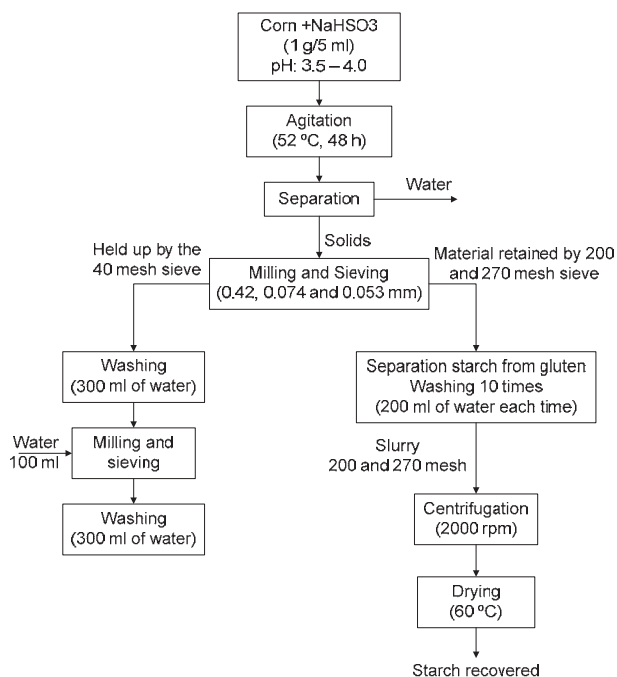


Figure 1. Experimental procedure to starch recovery (Method M1).
Source: The authors

Enzymatic modifications were based on procedures exposed by [19]. Starch modification using the α -amylase occurred by preparing a 1/50 dilution of enzyme in a buffer solution, which gave an activity level of 340.000 Modified Wohllhemuth Units (1 MWU = quantity of enzyme that will dextrinize one milligram of starch in 30 minutes at pH 6 and 50 °C). The buffer was prepared by combining deionized water with 2.4 g/L of sodium phosphate (monobasic; Sigma) and 0.39 g/L of sodium chloride (Sigma). The pH was adjusted to 6.9 using an aqueous solution of sodium hydroxide (Carlo ERBA Reagent). The starches were added at 2:3 v/v ratio of starch to enzyme solution. The modification took place at pH 6.9 at 25 °C for 7 days. The starch was then washed repeatedly and then dried in a convection oven for 24 h at 40 °C.

Cellulose hydrolysis was carried out with a 1/2 and 1/6 cellulase solution in citrate buffer at a ratio of 10 ml of the enzyme solution to 1 g of sugarcane bagasse. The mixture was kept at 50 °C in a shaker for 6 days. After enzymatic hydrolysis of the starchy materials and sugarcane bagasse, the reducing sugar concentrations were measured according to the method proposed by Nelson and Somogyi [20].

2.4. SEM, DSC, XRD and BET analysis

After hydrolysis, several analyses were carried out using Scanning Electron Microscope (SEM), Differential Scanning Calorimetric (DSC), X-ray Diffraction (XRD) and Brunauer-Emmett-Teller (BET). All samples were dried in a vacuum oven (3 mm Hg) at 60 °C for 3 days. SEM micrographs were obtained in a JEOL JSM-5910LV microscope. Calorimetric plots were measured in a DSC Q 100 (Module DSC Standard Cell FC, TA Instruments) unit in a temperature range of 25-200 °C for starchy materials and 25-355 °C for sugarcane bagasse, using a heating

rate of 10 °C/min, under nitrogen atmosphere with a flow of 50 mL/min. The X-ray diffraction plots were obtained in a Rigaku (MiniFlex II) unit with CuK α at 30 kV and 15 mA. The diffraction angle ranged from 35° to 2° with a scan rate of 5 °/min. Finally, the superficial area was determined in a Micromeritics ASAP 2020 unit (Micromeritics Instrument Corporation).

3. Results and discussion

3.1. Starch recovery

Table 1 shows the moisture, starch yield and the waste percentage determined for the starchy materials. From these results, we can see that the lowest yield corresponds to cassava due to its high moisture content therefore, for this raw material, removing water becomes in an important step that requires an efficient technique. [21]. Thus, the starch recovered from corn using the method M1 had the highest yield in contrast to method M2. The highest yield was obtained for corn also with method M2. On the other hand, cassava showed the highest waste production, while treated corn using method M1 showed the lowest. This fact suggests that the sugar yield and waste production in the starch recovery processes are strongly inversely related.

Additionally, the cornstarch recovery process generated other products, which can be considered commercial byproducts in a wet milling or dry milling concepts. In this sense, Table 2 shows additional fractions of fiber and gluten as byproducts from corn hydrolysis. Thus, starches obtained from the three selected raw materials were not completely pure given the remaining protein. The starch purities obtained were $74.03 \pm 0.520\%$, $73.92 \pm 0.410\%$ and 74.03 ± 0.658 for upright elephant ear, corn and cassava, respectively.

Table 1.
Moisture, starch yield and waste percentage from different starchy sources.

Raw material (RM)	Moisture (%)	Starch yield (starch mass / RM mass)*100 ^a	Waste percentage (waste mass / RM mass)*100
Upright elephant ear	63.19 \pm 0.320	22.25 \pm 0.021	53.24 \pm 0.019
Cassava	71.05 \pm 0.850	12.91 \pm 0.052	79.17 \pm 0.048
Corn "M1" ^b	46.49 \pm 0.540	40.40 \pm 0.030	16.03 \pm 0.017
Corn "M2" ^c	46.49 \pm 0.540	34.48 \pm 0.007	17.45 \pm 0.004

Source: The authors

Average of two replicates

^a Yield On dry matter percentage

^b M1: Method exposed by [16]

^c M2: Method exposed by [17]

Table 2.
Additional fractions obtained in the starch recovery from corn.

Corn fractions	Fresh corn	Method (M1)	Method (M2)
Fiber	15,32 \pm 0,051	8,50 \pm 0,017	13,21 \pm 0.021
Gluten	12,87 \pm 0,080	25,98 \pm 0,019	24,33 \pm 0.028
Starch	38,02 \pm 0,023	40,40 \pm 0,030	34,48 \pm 0.007

Source: The authors

Average of two replicates
All yields expressed in dry matter

3.2. Sugarcane bagasse treatment

Sugarcane moisture was 65.02 ± 0.080 %. The crust corresponded to 92.85 ± 0.002 % of the dry matter while marrow was only 7.15 ± 0.002 % of the dry matter. Consequently, the subsequent analyses were performed based on sugarcane crust due to its greater contribution.

3.3. Enzymatic hydrolysis

Conditions and conversion of enzymatic hydrolysis are shown in Table 3. The highest starch conversion was reached for upright elephant ear starch followed by cassava starch [22]. This fact suggests that starch granules from roots are more exposed to the enzymatic attack. Sugarcane bagasse with a dilution of 1:2 v/v presented greater conversion than that obtained from a dilution of 1:6 v/v. This behavior was expected due to the low enzyme concentration.

In the hydrolysis process, α -amylase breaks only the α -1,4 linkages presented in amorphous zones producing glucose units and partial hydrolysis is accomplished by correct hydrolysis conditions. This procedure was necessary for holding the starch granule structure while exposing the hydroxyl groups of the amorphous starch chain. A similar process occurs with the cellulose from sugarcane bagasse. In the partial acid hydrolysis, cellulose was broken down into cellobiose (glucose dimer), cellotriose (glucose trimer), and cellotetrose (glucose tetramer), whereas upon complete acid hydrolysis, it is broken down into glucose. The β -1,4 glycoside linkages cause cellulose to be in a low surface area crystalline form.

Table 3.
Conditions and conversions of enzymatic hydrolysis.

Raw material	Dilution ^a	Activity	Tem. (°C)	pH	% Conversion ^d
Upright elephant ear starch	1:50	340.000 MWU/ml ^b	25	6,9	15.17 ± 0.285
Corn starch	1:50	340.000 MWU /ml ^b	25	6,9	11.41 ± 0.621
Cassava starch	1:50	340.000 MWU /ml ^b	25	6,9	13.28 ± 0.542
Sugar cane bagasse	1:2	11.000 ECU/ g $\pm 5\%$ ^c	50	4,9	12.09 ± 0.747
Sugar cane bagasse	1:6	11.000 ECU/ g $\pm 5\%$ ^c	50	4,9	6.32 ± 0.097

^a Defined as enzyme volume : solution volume.

^b MWU = Modified Wohlhemuth = one unit is the quantity of enzyme that will dextrinize one milligram of starch in 30 minutes at pH 6 and 50 °C.

^c ECU/g $\pm 5\%$ = one unit is the quantity of enzyme that will hydrolyze one gram of de hydroxiethylcellulose at 50 °C and pH 4,8.

^d Defined as (gram glucose formed / gram material initially present)*100.

Source: The authors

3.4. SEM, DSC, XRD and BET analysis

3.4.1. Analysis of SEM micrographs

Scanning electron microscope (SEM) images of the starch granules and bagasse are shown in Fig. 2.

Native upright elephant ear starch granules (Fig. 2a) are particles with spherical and ovoidal shapes with diameters of between 5 and 15 μ m and a smooth surface. Enzyme-treated upright elephant ear starch tends to get agglomerated while native granules were scattered. Modified starch granules changed their diameters to 1-5 μ m, and their shape was altered obtaining irregular arrangements with a rough surface (see Fig. 2b). Native starch granules from corn are shown in Fig. 1c, the particles presented a spherical shape with some irregularities, diameters of between 5 and 10 μ m and a semi-porous surface. Enzyme-treated cornstarch granules kept their size but presented some fractures due enzymatic attack (see Fig. 2d).

Micrograph from Fig. 2e shows native cassava starch granules, which present irregular shapes and sizes (between

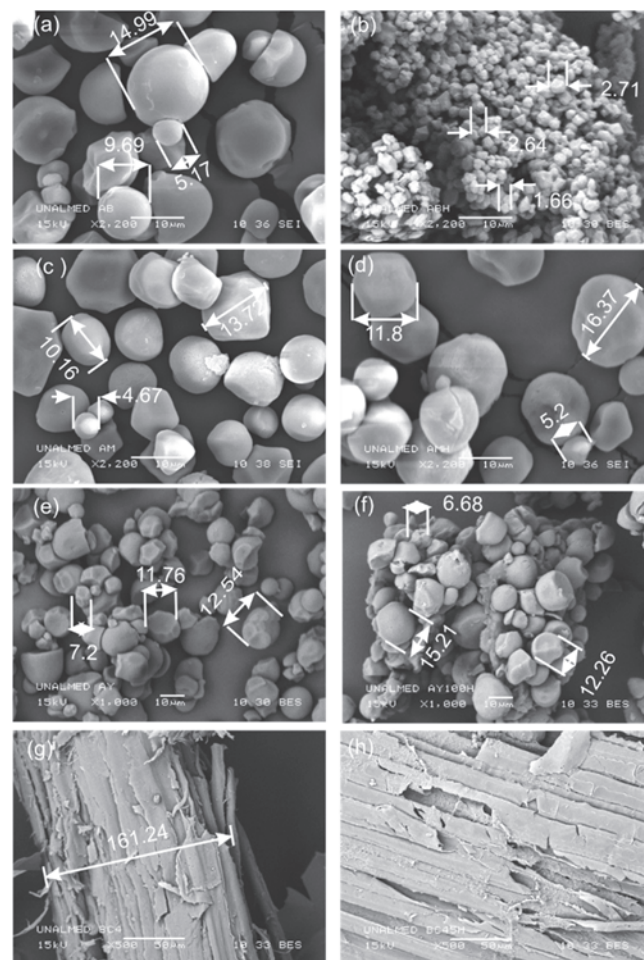


Figure 2. SEM micrographs of native and enzyme-treated materials: (a) native and (b) enzyme-treated upright elephant ear starch granules, (c) native and (d) enzyme-treated corn starch granules, (e) native and (f) enzyme-treated cassava starch granules, (g) native and (h) enzyme-treated sugar cane bagasse.

Source: The authors

5 and 10 μm) with a smooth surface. Enzyme-treated cassava starch granules, as well as upright elephant ear starch, tend to agglomerate but to a lesser extent. Fig. 2f shows the deterioration caused by enzymatic action, but with no change in the size. Enzymatic treatment generated deformations and breaking of the granules but no pores on the particle surface [22].

In general for all starches, enzyme treatment did not create new visible pores, but did cause some fractures and shape alteration, the most appreciable enzymatic effect occurred with upright elephant ear starch granules. Here, α -amylases did not have the ability to create pores on particle surfaces as observed by other authors using *glucoamylases* as a modifying agent [23,24].

Native bagasse fibers are shown in Fig. 1g. These are elongated with lengths of over 200 μm and are formed with several smooth surface layers. Hydrolyzed bagasse fibers with an enzyme to buffer ratio of 1:2 v/v are shown in Fig. 1h. There is evidence of layers breaking by enzymatic action exposing the inner channels of the fiber, but fiber length was kept after hydrolysis. Bagasse fibers hydrolyzed with an enzyme to buffer ratio 1:6 v/v, presented minor signs of enzymatic degradation.

3.4.2. Analysis of XRD plots

XRD patterns for starchy materials are shown in Fig. 3a. The patterns and characteristic of the starches are similar; however, the difference depends on the natural source of each starch. According to starch classification from X-ray diffraction, starch from Corn is classified as type A, because it presents the highest peaks at around 2θ (angle of diffraction) values of 14.92° , 17.18° , 17.8° and 22.7° , which are characteristic angles for this sort of starch [25]. Thus, upright elephant ear and cassava starch presented the maximum peak at 16.82° and 16.98° , respectively. Moreover, it has peaks of less intensity at 2θ values of 20.80° , 22.86° and 24.20° for upright elephant ear starch and 19.70° , 22.8° and 24.2° for cassava starch. As a result, upright elephant ear and cassava starches were classified as type B.

Crust and marrow bagasse have different structures and crystallinity because crust bagasse presented two diffraction peaks at 2θ values, (18.04° and 21.9°). However, marrow bagasse presented only a peak at 2θ values (21.86°), which is characteristic of the cellulose structures [26,27].

Fig. 3b shows the diffraction patterns for native and enzyme-treated upright elephant ear starch. Enzyme-treated upright elephant ear starch presents peaks with more intensity and definition than native starch suggesting that crystallinity was increased. Enzymatic attack promoted the formation of two peaks at 2θ values of 14.98° and 17.84° , the XRD pattern being more similar to a type A starch. Fig. 3c shows native and enzyme-treated cornstarch which did not present differences in 2θ values for their diffraction peaks. Nonetheless, these peaks were more defined and with greater intensity for enzyme-treated starch, indicating that hydrolysis generated a more crystalline structure. The latter also happened to upright elephant ear starch. Fig. 3d shows the enzyme-treated cassava starch which presents diffraction peaks at 2θ (14.92° and 17.76°), indicating that the XRD

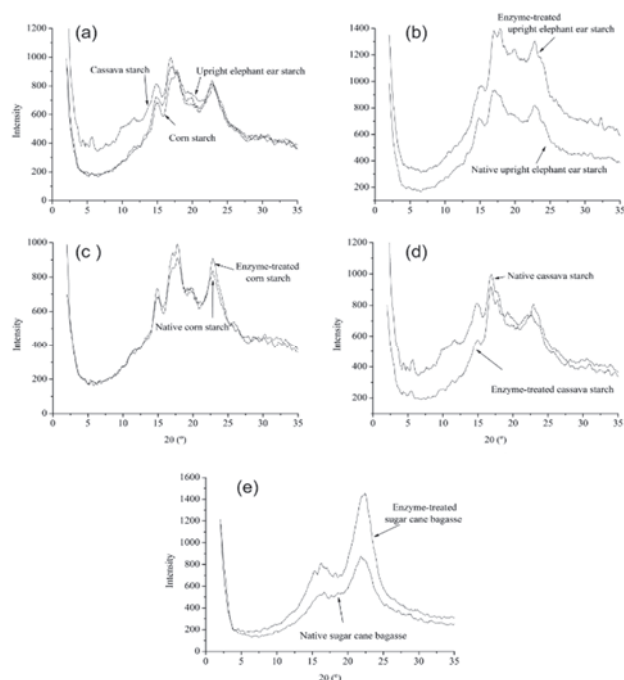


Figure 3. XRD plots of native and enzyme-treated materials: (a) native raw materials, (b) native and enzyme-treated upright elephant ear starch granules, (c) native and enzyme-treated corn starch granules, (d) native and enzyme-treated cassava starch granules, (e) native and enzyme-treated sugar cane bagasse (Enzyme/ buffer ratio 1:2).

Source: The authors

pattern was more similar to a type A starch. Diffraction peaks for this starch presented less intensity. Fig. 3e shows the hydrolyzed crust bagasse, which exhibited defined diffraction peaks and high intensity for both evaluated ratios (1:2 v/v and 1:6 v/v). This Figure also shows the diffraction pattern for enzyme to buffer ratio of 1:2 v/v. Bagasse became more crystalline with the enzyme treatment. As a consequence, enzymatic hydrolysis promoted the crystallinity by attacking the amorphous zones first.

3.4.3. Analysis of BET surface area

Nitrogen adsorption isotherms were obtained using the BET method. According to Brunauer's classification [28], all starchy materials exhibited type II isotherms (Isotherm with large deviation from Langmuir model absorption) with hysteresis from the middle zone of the curve indicating the presence of meso and macro pores. Sugarcane bagasse also showed a type II isotherm with a more pronounced hysteresis. This indicates that at high pressures, condensation in meso and macro pores is higher, which makes it more difficult to evaporate the nitrogen. Considering the latter, Table 4 shows the structural parameters obtained from the isotherms by the BET method.

Upright elephant ear starch was the material with the highest superficial area for both native and enzyme-treated cases. Enzymatic hydrolysis did not significantly change the superficial area for all tested materials. The specific surface areas obtained are low compared to other commercial materials. As a result, there is concern about the viability of

Table 4.
Structural and adsorption parameters for the evaluated materials.

Raw material	Superficial area BET (m ² /g)	D _p BJH ^a (Å)	V _p BJH ^b (cm ³ /g)
Upright elephant ear starch	2.1695 ± 0.0193	181.967	0.008031
Enzyme-treated upright elephant ear starch	2.8686 ± 0.0219	96.186	0.007675
Corn starch	0.9816 ± 0.0039	157.670	0.003761
Enzyme-treated corn starch	0.8267 ± 0.0064	168.129	0.002807
Cassava starch	0.4984 ± 0.0042	143.669	0.001266
Enzyme-treated cassava starch	0.6625 ± 0.0163	208.422	0.001125
Sugarcane bagasse	1.0058 ± 0.0913	136.012	0.003765
Enzyme-treated sugarcane bagasse	1.0945 ± 0.0436	172.946	0.003266

a BJH (Barrett-Joyner-Halenda). Desorption average pore diameter.

b BJH (Barrett-Joyner-Halenda). Desorption cumulative volume of pores.

Source: The authors

using these materials as adsorbents. However, it is important to evaluate the performance of these materials as adsorbents in dehydration processes. This is relevant since enzymatic treatment may have exposed more functional groups (hydroxyl groups) that can absorb water. However, such groups could not be found using the BET method. Thus, pore diameters obtained by the BJH method [29] confirmed the existence of macro pores in the studied materials.

3.4.4. Analysis of DSC plots

Fig. 4 shows the DSC patterns for native and enzyme-treated materials. DSC patterns for starches show the endothermic transition known as gelatinization, which occurs when the starch is heated in presence of water. Excess water was used to ensure the existence of a single peak. The onset, T_i , and the peak, T_p , temperatures were determined from the intercepts with the baseline. Fig. 4a shows the DSC pattern for native upright elephant ear, cassava, and corn, for which gelatinization peak temperatures were 62.02 °C, 62.79 °C and 73.07 °C respectively. This is in alignment with literature values for corn and cassava starch [30]. There is no open literature reporting on the thermal properties of upright elephant ear starch.

Peak temperatures for cassava and upright elephant ear starches were similar, while cornstarch had a peak temperature 10 °C higher than the roots. Other authors have reported that at higher temperatures, the granular structure has more ordered areas [31]. It has also been suggested that gelatinization of large starch particles occurs at a lower temperature than that of small particles [32].

Fig. 4b shows the change in onset and peak temperatures by enzymatic hydrolysis of upright elephant ear starch. Onset and peak temperatures were increased by 4 °C and 7 °C, respectively. The latter suggests that upright elephant ear starch became more crystalline. Degradation temperature was affected too, changing from 193.5 °C to 162 °C. This fact represents that thermal stability is decreased with partial hydrolysis. On the other hand, Fig. 4c shows that gelatinization

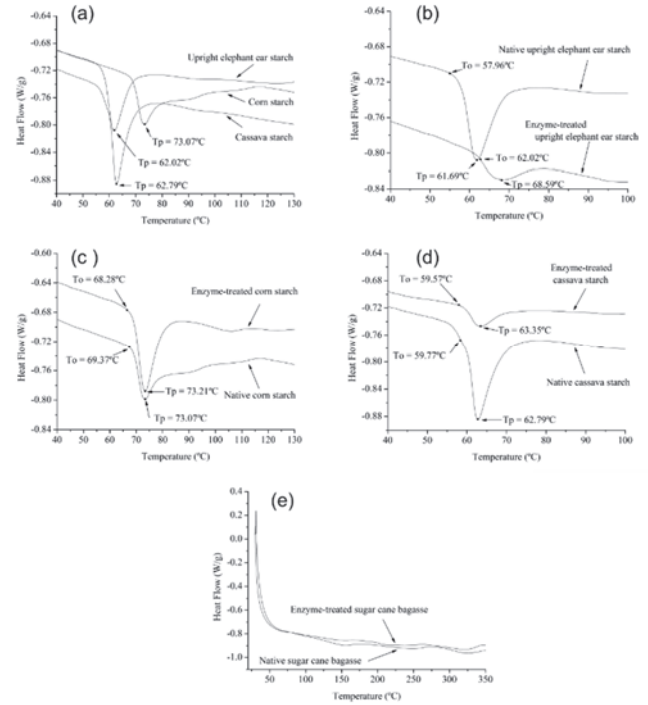


Figure 4. DSC plots native and enzyme-treated materials: (a) native raw materials, (b) native and enzyme-treated upright elephant ear starch, (c) native and enzyme-treated cornstarch, (d) native and enzyme-treated cassava starch, (e) native and enzyme-treated sugarcane bagasse.

Source: The authors

temperatures for cornstarch were not significantly altered. However, the degradation temperature decreased with de partial hydrolysis, implying the existence of some structural change. Degradation temperatures for native and enzyme-treated cornstarch were 174 °C and 152 °C, respectively.

Fig. 4d shows that gelatinization temperatures for cassava starch were not significantly altered. However, the degradation temperature increased with the partial hydrolysis, suggesting that some structural changes occurred. Degradation temperatures for native and enzyme-treated cassava starch were 168 °C and 179 °C, respectively. In general, gelatinization temperatures had the tendency increase with partial hydrolysis. This increment was related to the increase of crystallinity and decreasing particle size.

Finally, Fig. 4e shows that sugarcane bagasse did not present melting curves because of the high cristallinity of cellulose. Some small curves were obtained but these correspond to water elimination. In the evaluated range of temperatures, the degradation temperature was not reached.

4. Conclusions

Both, starchy and cellulosic materials presented changes after enzymatic modification. Size reduction, shape alterations, fractures, changes in the surface area, and pore formation the most important changes caused by the enzymatic modification.

From the raw materials used, corn and upright elephant ear presented the highest starch yield and reducing sugar

production respectively. Sugarcane bagasse did not require a further recovery process, and only a structural components separation (such as crust and marrow) was necessary. In general, enzymatic treatment increased crystallinity as shown in XRD patterns. Enzymatic hydrolysis generated alterations over starch granules as was observed in SEM images.

Upright elephant ear starch granules were more affected by hydrolysis than other materials, suffering a greater size reduction. The enzyme altered the surfaces of both, corn and cassava, but particle size was not affected. Enzymatic hydrolysis of sugarcane bagasse showed an appreciable fiber alteration exposing its inner channels. This work reveals that analyses of these types of materials are necessary to understand and evaluate the performance of the enzymatic hydrolysis as they offer valuable information about the influence of the structural and thermal properties on their adsorption performance.

Acknowledgment

The authors would like to thank the Nano-Structured and Functional Materials Laboratory (National University of Colombia at Manizales) for technical assistance in XRD and BET analyses, Dr. Andrés Rosales R., Magnetism and Advanced Materials Laboratory (National University of Colombia at Manizales) for technical assistance in DSC analyses, Dr. Medardo Pérez, Advanced Microscopy Laboratory (National University of Colombia at Medellín) for technical assistance in SEM analyses.

References

- [1] Cardona, C.A. and, Sánchez, O.J., Fuel ethanol production: Process design trends and integration opportunities. *Bioresource Technology* 98, pp. 2415-2457, 2007. DOI:10.1016/j.biortech.2007.01.002.
- [2] Quintero, J.A., Montoya, M.I., Sánchez, O.J., Giraldo, O.H. and Cardona, C.A., Fuel ethanol production from sugarcane and corn: Comparative analysis for a Colombian case. *Energy* 33 pp.385-399. 2008. DOI:10.1016/j.energy.2007.10.001
- [3] Sánchez, O.J. and Cardona, C.A., Trends in biotechnological production of fuel ethanol from different feedstocks. *Bioresource Technology* 99, pp. 5270-5295. 2008. DOI: 10.1016/j.biortech.2007.11.013
- [4] Van Dyk, J.S. and Pletschke, B.I., A review of lignocellulose bioconversion using enzymatic hydrolysis and synergistic cooperation between enzymes—Factors affecting enzymes, conversion and synergy. *Biotechnology Advances* 30, pp. 1458-1480. 2012. DOI: 10.1016/j.biotechadv.2012.03.002
- [5] Apar, D.K. and Özbek, B., α -Amylase inactivation during corn starch hydrolysis process. *Process Biochemistry* 39, pp. 1877-1892. 2004, DOI: 10.1016/j.procbio.2003.09.014
- [6] Apar, D.K. and Özbek, B., α -Amylase inactivation during rice starch hydrolysis. *Process Biochemistry* 40, pp. 1367-1379. 2005. DOI: 10.1016/j.procbio.2004.06.006
- [7] Zhang, Q., Zhang, P., Pei, Z.J. and Wang, D., Relationships between cellulose biomass particle size and enzymatic hydrolysis sugar yield: Analysis of inconsistent reports in the literature. *Renewable Energy* 60, pp. 127-136. 2013. DOI: 0.1016/j.renene.2013.04.012
- [8] Shujun, W., Jinglin, Y., Jiugao, Y., Jiping, P. and Hongyan, L., Structure characterization of C-type starch granule by acid hydrolysis. *Food Hydrocolloids* 22, pp. 1283-1290. 2008. DOI: 10.1016/j.foodhyd.2007.06.012
- [9] Driemeier, C., Oliveira, M.M., Mendes, F.M. and Gómez, E.O., Characterization of sugarcane bagasse powders. *Powder Technology* 214, pp. 111-116, 2011. DOI: 10.1016/j.powtec.2011.07.043
- [10] Fannon, J.E., Hauber, R.J. and Bemiller, J.N., Surface pores of starch granules. *Cereal Chemistry*, 69, pp. 284-288, 1992.
- [11] Hoover, R., Composition, molecular structure and physicochemical properties of tuber and root starches: A review. *Carbohydrate Polymers*, 45, pp. 253-267, 2001. DOI: 10.1016/S0144-8617(00)00260-5
- [12] Loh, Y.R., Sujun, D., Rahman, M.E. and Das, C.A. Sugarcane bagasse—The future composite material: A literature review. *Resources, Conservation and Recycling*. 75. pp. 14-22, 2013. DOI: 10.1016/j.resconrec.2013.03.002
- [13] Hyang-Aee, L., Nam-Hie, K. and Nishinari, K., DSC and rheological studies of the effects of sucrose on the gelatinization and retrogradation of acorn starch. *Thermochimica Acta*, 322, pp. 39-46, 1998. DOI: 10.1016/S0040-6031(98)00469-9
- [14] Chen, Y., Huang, S., Tang, Z., Chen, X. and Zhang, Z., Structural changes of cassava starch granules hydrolyzed by a mixture of α -amylase and glucoamylase. *Carbohydrate Polymers*, 85, pp. 272-275, 2011. DOI: 10.1016/j.carbpol.2011.01.047
- [15] Alarcon, M.F. y Dufour, D., Almidón agro de yuca en Colombia. CIAT Centro Internacional de Agricultura Tropical 1998. ISBN: 958-9439-67-5
- [16] Haros, M. and Suarez, C., Effect of drying, initial moisture and variety in corn wet milling. *Journal of Food Engineering*, 34, pp. 473-481, 1997. DOI: 10.1016/S0260-8774(97)00099-X
- [17] Serna, S.S.O., Refinación de almidón y producción de jarabes glucosados a partir de sorgo y maíz. *Transferencia de Programas de Graduados e Investigación*. [Online]. 1998. Available at: <http://www.mty.itesm.mx/die/dre/transferencia/Transferencia42/eli-01.htm>.
- [18] Giraldo, G., Orrego, C., Grajales, L., Telis, V., Gabas, A. and Telis J., Effect of drying methods on the thermodynamic properties of blackberry pulp powder. *DYNA*, 78(168), pp 139-148, 2011.
- [19] Beery, K.E., Gulati, M., Kvam, E. and Ladisch, M., Effect of enzyme modification of corn grits on their properties as an adsorbent in a skarstrom pressure swing cycle dryer. *Adsorption*, 4, pp. 321-335, 1998. DOI: 10.1023/A:1008846003116
- [20] Nelson, N., A Photometric adaptation of the Somogyi method for the determination of glucose. *Journal of Biological Chemistry*, 153, pp. 375-380, 1944.
- [21] Breuninger, W.F., Piyachomkwan, K. and Sriroth, K., Chapter 12 - Tapioca/Cassava starch: Production and use. In: James.B. and Roy, W., Eds. *Starch* (Third Edition). San Diego: Academic Press, 2009, pp. 541-568. DOI: 10.1016/B978-0-12-746275-2.00012-4
- [22] Quintero, J.A. and Cardona, C.A., Ethanol dehydration by adsorption with starchy and cellulosic materials. *Industrial Engineering Chemistry Research*, 48, pp. 6783-6788, 2009. DOI: 10.1021/ie8015736
- [23] Aggarwal, P. and Dollimore, D., A thermal analysis investigation of partially hydrolyzed starch. *Thermochimica Acta*, 319, pp. 17-25, 1998.
- [24] Aggarwal, P. and Dollimore, D., Degradation of starchy food material by thermal analysis. *Thermochimica Acta*, 357-358, pp. 57-63, 2000. DOI: 10.1016/S0040-6031(00)00368-3
- [25] Sajilata, M.G., Singhal, R.S. and Kulkarni, P.R., Resistant starch. A review. *Comprehensive Reviews in Food Science and Food Safety*, 5, pp. 1-17, 2006. DOI: 10.1111/j.1541-4337.2006.tb00076.x
- [26] Filho, G.R., de Assunção, R.M.N., Vieira, J.G., Meireles, C.d.S., Cerqueira, D.A., da Silva-Barud, H., Ribeiro, S.J.L. and Messaddeq, Y., Characterization of methylcellulose produced from sugar cane bagasse cellulose: Crystallinity and thermal properties. *Polymer Degradation and Stability*, 92, mpp. 205-210, 2007. DOI: 10.1016/j.polymdegradstab.2006.11.008
- [27] Zhao, H., Kwak, J.H., Conrad, Z.Z., Brown, H.M., Arey, B.W. and Holladay, J.E., Studying cellulose fiber structure by SEM, XRD, NMR and acid hydrolysis. *Carbohydrate Polymers*, 68, pp. 235-241, 2007. DOI: 10.1016/j.carbpol.2006.12.013
- [28] Andrade, P.R.D., Lemus, M.R. and Perez, C.C.E., Models of sorption isotherms for food: Uses and limitations. *VITAE Revista de la Facultad de Química Farmaceutica*, 18, pp. 325-334, 2011.
- [29] Webb, P.A. and Orr, C., Analytical methods in fine particle technology. *norcross*, First Edition ed, Micromeritics Instrument Corporation, USA, 1997. ISBN-13: 978-0965678308

- [30] Sandoval, M.L., Camerucci, M.A., Cavalieri, A.L. y Sian, A.N., Conformado por consolidación directa con almidón de precursores de materiales porosos de cordierita. [En línea]. 2007. Congresso Brasileiro de Cerâmica, No. 51, Salvador Brasil, [Consultado: Febrero de 2013] Disponible en: http://www.conicet.gov.ar/new_scp/detalle.php?keywords=&id=20192&congresos=yes&detalles=yes&congr_id=1186567
- [31] Qi, X., Tester, R.F., Snape, C.E., Yuryev, V., Wasserman, L.A. and Ansell, R., Molecular basis of the gelatinisation and swelling characteristics of waxy barley starches grown in the same location during the same season. Part II. Crystallinity and gelatinisation characteristics. *Journal of Cereal Science*, 39, pp. 57-66. 2004. DOI: 10.1016/S0733-5210(03)00065-1
- [32] Bar, W.H., Park, Y.K. e Papini, R.S., Relacao entre intumescimento, gelatinizacao suscetibilidade dos amilos de mandioca e de milho . alfa-amilase bacteriana. *Revista Brasileira de Tecnologia*, 2, pp. 95-100, 1971.

J. Quintero, completed his BSc in Chemical Engineering in 2005, his MSc degree in Chemical Engineering in 2008, and his PhD degree in Automatic Engineering in 2011, all of them at Universidad Nacional de Colombia, Manizales, Colombia. Since 2011, he has been a postdoctoral researcher at the Universidad Católica de Valparaíso, Valparaíso, Chile and since 2014 he has been an assistant professor at the School of Biochemical Engineering at the same university. His main research interests are: first and second generation biofuels production, process modelling and simulation, life cycle analysis, fermentation, enzymatic conversion and non-conventional separation operations.

ORCID: 0000-0003-1224-5649

J. Davila, completed his BSc. Eng in Chemical Engineering in 2005 at Universidad Nacional de Colombia, Manizales; his MSc degree in Chemical Engineering in 2008, at Universidad del Valle, Cali, Colombia and his PhD degree in Automation in 2016, at Universidad Nacional de Colombia, Manizales, Colombia. From 2009 to 2012, he worked for Chemical Engineering Department at the Universidad de los Andes, Bogotá, Colombia. Currently, he is a full professor in the Engineering Department, Universidad Jorge Tadeo Lozano, Bogotá Colombia. His research interests include: simulation, modeling and design of processes, alternative energy, supercritical fluids, and biorefineries.

ORCID: orcid.org/0000-0002-4583-7512

J. Moncada, works as a junior researcher at the Copernicus Institute of Sustainable Development, Utrecht University, in the Netherlands. He holds BSc. and Msc. degrees in Chemical Engineering. He has worked as researcher and consultant in many projects related to the techno-economic and environmental assessment of biofuels and biorefineries. His research interests include the design and sustainability assessment of biorefinery systems, bio-based economy, sustainability assessment of novel bio-based routes for chemicals production, and integrated multiproduct biorefineries.

ORCID: orcid.org/0000-0003-3620-8594

O.H. Giraldo, completed his BSc. in Chemical Engineering in 1992, at Universidad del Valle, Cali, Colombia, and his Dr. degree in 2000, at the University of Connecticut, USA. He is a full professor in the Physics and Chemistry Department, Universidad Nacional de Colombia, Manizales Colombia. His research interests include: Inorganic and nuclear chemistry, chromatographic, nanotechnology.

ORCID: orcid.org/0000-0002-9213-6348

C.A. Cardona, completed his BSc. En in Chemical Engineering in 1993, his MSc. degree in Chemical Engineering in 1995 and his Dr. degree in 2001, all of them at the Estatal Academy Lomonosov, Russia. He is a full professor in the Chemical Engineering Department, Universidad Nacional de Colombia, Manizales, Colombia. His research interests include: industrial biotechnology, biofuels, agroindustrial waste usage, biorefineries and reactive distillation.

ORCID: orcid.org/0000-0002-0237-2313



UNIVERSIDAD NACIONAL DE COLOMBIA

SEDE MEDELLÍN

FACULTAD DE MINAS

Área Curricular de Ingeniería Química e Ingeniería de Petróleos

Oferta de Posgrados

Maestría en Ingeniería - Ingeniería Química
Maestría en Ingeniería - Ingeniería de Petróleos
Doctorado en Ingeniería - Sistemas Energéticos

Mayor información:

E-mail: qcaypet_med@unal.edu.co

Teléfono: (57-4) 425 5317

Robust sample size for weibull demonstration test plan

Manuel R. Piña-Monarez ^a, Miriam L. Ramos-López ^b, Alejandro Alvarado-Iniesta ^c
& Rey D. Molina-Arredondo ^d

Industrial and Manufacturing Dep. of the IIT Institute, Universidad Autónoma de Ciudad Juárez, Chihuahua, México.
^amanuel.pina@uacj.mx, ^bal127816@alumnos.uacj.mx, ^calejandro.alvarado@uacj.mx, ^drey.molina@uacj.mx

Received: August 13th, 2014. Received in revised form: September 07th, 2015. Accepted: March 11th, 2016.

Abstract

The efficiency of a Weibull demonstration test plan is completely determined by the total experimental time (T_a), which depends on the unknown sample size (n) and on the Weibull shape parameter (β). Thus, once β was selected, T_a depends only on n . Unfortunately, because n was estimated by the parametrical binomial approach, then if the confidence level C was higher than 0.63, n , and as consequence T_a , was overestimated, (For $C < 0.63$, they were underestimated). On the other hand, in this paper, because the intersection between n and β , for which T_a was unique, was found with n depending only on $R(t)$, then the estimation of T_a was optimal. On the other hand, since once β was selected, η was completely determined, then β and η were used to incorporate the expected failure times of the operational level in an accelerated life test analysis (ALT). Numerical applications are given.

Keywords: Weibull demonstration test plan, Success run testing, Lipson equality, Accelerated life testing.

Tamaño de muestra robusta para planes de demostración weibull

Resumen

La eficiencia de un plan de demostración Weibull está completamente determinada por el tiempo total de experimentación (T_a) el cual depende del tamaño de muestra desconocido (n) y del parámetro de forma Weibull (β). De esa forma, una vez que β fue seleccionada, T_a depende sólo de n . Desafortunadamente, debido a que n es estimada a través del método binomial paramétrico, entonces si el nivel de confianza C es mayor de 0.63, n y como consecuencia T_a , son sobre-estimados (para $C < 0.63$, estos son subestimados). Por otro lado, en este artículo, debido a que la intersección entre n y β , para la cual T_a es única, se encontró con n dependiendo sólo de $R(t)$, entonces la estimación de T_a es óptima. Por otro lado, dado que una vez que β fue seleccionada, η está completamente determinada, entonces β y η fueron utilizadas para incorporar los tiempos de falla esperados del nivel operacional en un análisis de prueba de vida acelerada (ALT). Aplicaciones numéricas son dadas.

Palabras Clave: Planes de demostración Weibull, Rachas exitosas, Desigualdad de lipson, Pruebas de vida acelerada.

1. Introduction

In reliability engineering, because of its flexibility the Weibull distribution is one of the most commonly used probability density functions to model the behavior of a product or process through the time [6]. Moreover, since the lower reliability index could be seen as an index of stability (quality through the time), the Weibull demonstration test plans are performed without failures in order to determine whether the product fulfills its designed reliability $R(t)$. On the other hand, to perform the test plan, we must know; $R(t)$, the designed life time (t_d), the operational environmental, the

desired confidence level (C) and the Weibull shape parameter β . Regardless of this knowledge, given that the test plan is completely determined by the total experimental time T_a , and because for a known β , T_a depends only on the sample size n , then the efficiency of the test plan depends on the accuracy with which n is estimated. In practice, the parametrical binomial approach, considering a constant failure rate p ([2] chapter 9, [5] and section 2.2), is used to estimate n . Unfortunately, if C higher than 0.63, n is overestimated, and as a consequence T_a is overestimated too. To solve this problem, and based on the fact that for constant β , T_a is directly related to the Weibull scale parameter η , then based

How to cite: Piña-Monarez, M.R., Ramos-López, M.L., Alvarado-Iniesta, A. and Molina-Arredondo, R.D., Robust sample size for weibull demonstration test plan. DYNA 83(197), pp. 52-57, 2016.

on the addressed relations among η , n , $R(t)$ and β , and on the found intersection between n and β for which T_a is unique, in this paper a method to estimate n in closed form, but independent of β , is given. Moreover, because n was found to be depending only on the known $R(t)$ index, the estimated T_a is completely representative of the designed test plan.

On the other hand, in order to show how to proceed under time or lab restrictions, by using the above estimated n , the Lipson equality is applied to perform a tradeoff between n and the experimental time T_i for which T_a is constant. Finally, because n is directly related with $R(t)$, in section 5.1, we show how it could be used in the median rank approach to incorporate the expected lifetimes of the operational level if an accelerated life test analysis (ALT) has to be used. The paper structure is as follows. Section 2 addresses the problem statement. Section 3 presents the proposed method. Section 4 offers the application and comparison between the proposed method and the binomial approach. Section 5 outlines the steps to incorporate the failure times of the operational level into the ALT analysis. Section 6 presents the conclusions. Finally, the paper ends with the references in section 7.

2. Problem statement

Since a Weibull demonstration test plan is performed without failures, the Weibull parameters β and η could not be estimated. Thus, the efficiency of the test plan depends completely on the accuracy on which T_a is estimated. But, because T_a for constant β depends only on n , then the efficiency of the test plan now only depends on the accuracy with which n is estimated. Regardless of this, since n is estimated by the parametrical binomial approach considering a constant failure rate p , and a confidence interval C , then when C higher than 0.63 is selected, n is overestimated and for C lower than 0.63 n is underestimated. The overestimation (or underestimation) of n directly implies that T_a is overestimated (or underestimated) also. Observe that this means that the test plan fails to demonstrate whether the product fulfills its designed $R(t)$ index. On the other hand, although in practice, β is selected from a historical data set (or engineering knowledge), because its value depends on the material characteristics [18], and on the variability of the manufacturing process [13], here the analysis is presented in two parts. The first is conducted to address the effect that the uncertainty of β has on T_a , and the second is conducted to statistically identify the disadvantages that the use of the binomial approach has on the estimation of n . To do this, let us first present the β analysis.

2.1. Shape parameter analysis

Since for non-failures Weibull analysis T_a is cumulated as

$$T_a = n * T_i^\beta \quad (1)$$

Where T_i is the unite experimental time, which is selected as the designed time ($T_i = t_d$), the value of β has a big impact on T_a , [see [12] and [15] sec. 2.3] and because in the Weibull demonstration test plan there is no failures information to estimate β , and due to the fact that its estimation depends on the variability of the manufacturing process [13], in practice β is selected from tabulated data sets. Moreover, because of the

Weibull closure property (β has to be constant), once the β value is selected it has to be considered constant in the analysis [16]. On the other hand, regardless which value of β we had selected, once it was assigned, the efficiency of the test plan depends on the scale parameter η . Thus, we are now interested in how to estimate T_a in function of η . However, because there are no failures, the lower expected limit of $R(t)$ has to be used. That is, that the lower limit of η , here called (η_L), has to be estimated. According to [10] and [11], η_L is given by

$$\eta_L = \left[\frac{2 \sum_{i=1}^n T_i^\beta}{\chi_{\alpha, 2r+2}^2} \right]^{\frac{1}{\beta}} \quad (2)$$

Where α is the significant level and r is the number of observed failures. On the other hand, since equation (2) for zero failures, with C representing the desired confidence level and T_i equal to t_d , is given by

$$\eta_L = \left[\frac{\sum_{i=1}^n T_i^\beta}{-\ln(1-C)} \right]^{\frac{1}{\beta}} = \left[\frac{n T_i^\beta}{-\ln(1-C)} \right]^{\frac{1}{\beta}} \quad (3)$$

Then by selecting $C = (1 - e^{-1})$ in (3), the relation between η_L and T_a is given by

$$T_a = \eta_L^\beta \quad (4)$$

Clearly from (4), T_a is in function of β , which in practice is selected from a data set (or engineering knowledge). Now, let us focus on the uncertainty of n .

2.2. Binomial approach analysis

Given that the Weibull demonstration test plan, the parametric binomial approach used to determine n (see [2] and [5] chapter 9) is based on the binomial distribution given by

$$p(x) = \frac{n!}{(n-x)!x!} p^x q^{(n-x)} \quad x=0,1,2,\dots,n \quad (5)$$

which, instead of considering the time and the risk function to determine n , considers a constant failure rate p and a confidence level C to model the uncertainty on $R(t)$, then n is not optimal. Given this, *first* let us analyze how the estimation of n is formulated. In doing so, we can see that because no failures are allowed ($x=0$), the lower confidence of $R(t)$ has to be used. Thus, C based on the fact that if n items are tested and $k=1,\dots,n$ of them fails, C is given by

$$C = 1 - \sum_{i=1}^k \frac{n!}{(n-i)!i!} R^{n-i} (1-R)^i \quad (6)$$

In (6), R represents the lower confidence of $R(t)$. In (6), R is used instead of $R(t)$ because the binomial approach does not consider the time variable t . Therefore, equation (6) with zero failures ($x = 0$) is given by

$$R^n = (1 - C) \quad (7)$$

Finally, by rearranging terms, n is given by

$$n = \frac{\ln(1 - C)}{\ln R(t)} \quad (8)$$

Function (8) is known as success run testing [5]. *Second*, suppose that our customers are asking us to demonstrate using $C=0.90$, if their product fulfills $R(t)=0.96$ for $t_d=1500$ hrs. In addition suppose, that from historical data, we know that $\beta=2.5$. Thus, by using (8), we have to test without failures $n=57$ pcs for 1500hrs each. Thus, observe from (1) that $T_a=57*1500^{2.5}=4967101142$ and from (4) that $\eta=7558.59718$. With this information, since we now know β and η , then by using these parameters in the Weibull reliability function given by

$$R(t) = \exp\left(-\left(\frac{t}{\eta}\right)^\beta\right) \quad (9)$$

We note that the demonstrated $R(t)$, for $t_d=1500$ hrs, is $R(t)=0.9826$, instead of the planned $R(t)=0.96$. Thus, because $R(t)$ and β are known, we conclude that the C value used in (8) overestimated n , and that, as a consequence, T_a defined in (1) was overestimated too.

In contrast, observe that for $C=0.50$, $n=17$ pcs, $T_a=1481416130$ and $\eta=4658.7652$, with $R(t)=0.9428$. That is to say that for $C=0.50$, n was underestimated. Thus, a C value between 0.5 and 0.9 for which n is optimal exists. In the next section, this value is statistically addressed and generalized to any desired $R(t)$ and β value.

3. Proposed method

Given that the goal of the proposed method is to remain practical, first let us show how the binomial approach, the Lipson equality, and the Weibull reliability function are related. In doing so, we can first see that because T_a completely determines $R(t)$, the analysis is based on T_a . *Second*, we can see that for constant β , T_a depends only on n . Finally, note that regardless how β , and n were estimated, once their values were selected, a tradeoff between n and T_i for which T_a remains constant could be made by applying the Lipson equality (see [5]) as follows. The Lipson equality is formulated by replacing R given in (7) with the Weibull reliability function defined in (9). After the replacement, the equality is given by

$$R(t) = (1 - C)^{\frac{1}{n}} = \exp\left(-\left(\frac{t}{\eta}\right)^\beta\right) \quad (10)$$

From (10), by taking logarithms and by rearranging terms, the sample size n in the Lipson equality is given by

$$n = \frac{\ln(1 - C)}{L^\beta \ln R(t)} \quad (11)$$

In (11), L represents one life of the product $L = t_d$. Function (11) is known as Lipson equality (or extended life approach) and it relates the binomial approach and the Weibull reliability function. Finally, from (11), it is clear that it works regardless of how n and β were selected. Thus, in the same way as T_a , its efficiency depends on how n is estimated.

On the other hand, in order to estimate T_a accurately, by substituting (8) in (2), η_L is found to be depending only on $R(t)$ and β as in (12).

$$\eta_L = \left[\frac{T_i^\beta}{-\ln(R(t))} \right]^{\frac{1}{\beta}} \quad (12)$$

Then based on (12), and by selecting n as in (13).

$$n = -1 / \ln(R_{td}) \quad (13)$$

the relation between T_a and η_L as in (4), is given by

$$T_a = \eta_L^\beta = [n * T_i^\beta] \quad (14)$$

From (14), T_i is given by

$$T_i = \left[\frac{-t_d^\beta}{n * \ln(R(t_d))} \right]^{\frac{1}{\beta}} \quad (15)$$

And clearly, since there no failures are allowed, then we can select $T_i = t_d$, thus η is given by

$$\eta = n^{1/\beta} T_i = n^{1/\beta} t_d \quad (16)$$

From (16), since $T_i = t_d$ is known, and n is directly related to $R(t)$ as in (13), η now depends only on the selected β value. On the other hand, observe that n in (13) is estimated regardless of the value of β . Seeing this numerically, suppose we are determining η_L defined in (12) by a designed time $t_d=1500$ hrs, and suppose we desire to demonstrate a reliability of $R(t)=0.90$. Furthermore, suppose that from historical data (or engineering knowledge), we know that β ranges from $1.5 \leq \beta \leq 3$. Then by testing different n values for $\beta=1.5$ and $\beta=3$, as in Table 1, we found that T_i shifts its behavior from higher to lower, implying that an intersection for which T_i is equal to both β values exists. And because, this intersection corresponds exactly to the n value defined in (13), and it does not depend on β , we conclude that by estimating n using (13), the proposed method is robust under the uncertainty that β has over T_a . In particular, we can see from Table 1 and Fig. 1 that for $T_i = t_d$, n is as in (13) and $R(t)$ is as expected for both values of β , and that η_L is as in (12).

Table 1.
Experimental time for different sample sizes.

Sample N	Shape Parameter		Reliability		Experimental Time	
	$\beta=1.5$	$\beta=3$	$R(t)\beta=1.5$	$R(t)\beta=3$	Ti $\beta=1.5$	Ti $\beta=3$
5	4386.03	2564.96	0.82	0.82	2299.64	1857.27
6	4952.89	2725.68	0.85	0.85	2036.44	1747.76
7	5488.96	2869.40	0.87	0.87	1837.56	1660.22
8	6000.00	3000.00	0.88	0.88	1681.05	1587.94
9	6490.12	3120.13	0.89	0.89	1554.10	1526.81
9.4912	6724.18	3175.89	0.90	0.90	1500.00	1500.00
11	7419.13	3335.97	0.91	0.91	1359.49	1428.02
12	7862.22	3434.14	0.92	0.92	1282.88	1387.20
13	8293.16	3527.00	0.93	0.93	1216.22	1350.67
14	8713.18	3615.21	0.93	0.93	1157.59	1317.72
15	9123.30	3699.32	0.94	0.94	1105.55	1287.76

Source: The authors

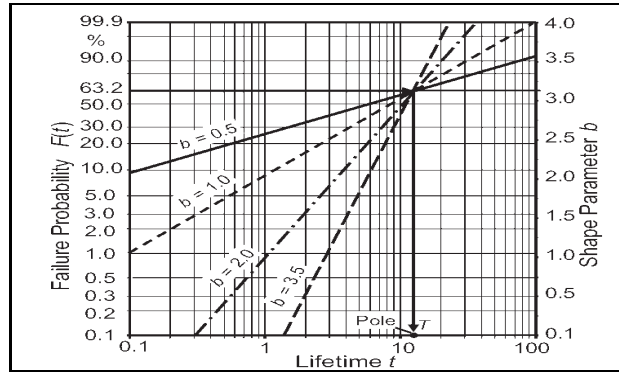


Figure 1. Intersection for the shape parameter
Source: [2], p.46

That is, data in this intersection holds with equations (12) to (16). The steps to apply the proposed method are as follows.

3.1. Steps of the proposed method

In order to demonstrate whether a product fulfills its designed reliability, the following steps were taken.

1. Determine the reliability level $R(t)$ index to be demonstrated, the operational environmental to be tested and the design time t_d .
2. Determine the value of β to be used. A base line product, engineering knowledge or historical data could be used to select the most suitable β value.
3. By using (13) with the $R(t)$ level of step 1, determine the sample size n to be tested without failures during the t_d lifetime each.
4. If there are experimental or time restrictions, perform the desired tradeoff between n and t_d , using (11) with $C=0.63212$.
5. Test each specimen by t_d lifetime and, if neither of them fails, go to step 6. If one of them fails, go to step 7.
6. By using (12) or (16), estimate the expected η_L value. And by using η_L , t_d and the selected β value in (9), determines the demonstrated $R(t)$ value, and draw your conclusions.
7. Correct and reinforce the design (or process) and go to step 1.
8. If you are performing an accelerated life testing, and the

normal operational conditions could not be applied to the experiment, follows the steps given in section 5.1.

4. An application

As an application, first consider the data in section 2.2, ($R(t)=0.96$, $C=0.90$ and $C=0.50$, $t_d=1500$ hrs and $\beta=2.5$). With this information obtained by using (13) in step3, we have to test $n=24.49 \approx 25$ pcs for 1500hrs each. And by using (1) or (14) with $\beta=2.5$, $T_a=2134685635$ hrs and by using (12) or (16), $\eta_L=5391.797$. Thus, by applying (9) the demonstrated reliability, is $R(t)=0.96$ as was planned.

On the other hand, suppose that because of the lab capacity, we can run each test for no more than 3000hrs; ($L=2$ lives), then by using (11) with $C=0.63212$, we have to run $n=4.33 \approx 5$ pcs without failures for 3000hrs each, and from (1) or from (14) $T_a = 5 * (3000)^{2.5} = 2464751.509$ hrs and from (12) or (16) $\eta_L=5710.96$ hrs, and by using η_L in (9), the demonstrated reliability, as planned, is $R(t)=0.96$.

Finally, since n in (13) depends only on $R(t)$, and because $R(t)$ is used in the response variable of the median rank approach, in the next section, n is used to incorporate the expected normal operational lifetimes into an accelerated life time analysis.

5. Weibull accelerated life test planning

In Weibull accelerated life test analysis (ALT) for constant and interval valued variables, the shape parameter β is considered constant in the analysis due to the Weibull closure property [16]. Thus, the reliability index $R(t)$ depends only on η ([4], [7] and [9]), which in ALT is estimated as a linear function of the covariates by using a life/stress models $r\{X_{Li}(t)\}$ [1], [3], [7] and [10], as follows

$$\eta_{Li} = r\{X_{Li}(t)\} \quad (17)$$

Here, it is important to note that in (17) for constant over time stress variables $r\{X_{Li}(t)\}$ is parametrized as

$$r\{X_{Li}(t)\} = e^{-\beta'Z} \quad (18)$$

Where β is a vector of regression coefficients to be estimated and Z is a vector of the effect of the related stress variable (e.g. in Arrhenius $Z=1/T$ where T is the temperature in Kelvin degrees). Thus, by using the Weibull density function given by

$$f(t) = \frac{\beta}{\eta} \left(\frac{t}{\eta} \right)^{\beta-1} \exp \left\{ - \left(\frac{t}{\eta} \right)^{\beta} \right\} \quad (19)$$

the Weibull/life/stress parameters are estimated by substituting η in (19) with the corresponding $r\{X_{Li}(t)\}$ model defined in (17) or (18). For example suppose that the stress variable is the temperature, then the Arrhenius model is used. The Arrhenius model is given by

Table 2.
Doe for ALT analysis

Stress Level	Replicates				Parameters	
	R1	R2	...	Rm	Shape	Scale
1	Y ₁₁	Y ₁₂	...	Y _{1m}	β	η_1
2	Y ₂₁	Y ₂₂	...	Y _{2m}	β	η_2
3	Y ₃₁	Y ₃₂	...	Y _{3m}	β	η_3

Source: The authors

$$\eta = r\{X_{Li}(t)\} = C e^{\left(\frac{B}{T}\right)} \quad (20)$$

Thus, the Weibull/Arrhenius model is given by

$$f(t/T) = \frac{\beta}{A} e^{-\left(\frac{B_1}{T}\right)} \left(\frac{t}{A} e^{-\left(\frac{B_1}{T}\right)} \right)^{\beta-1} \exp \left\{ - \left(\frac{t}{A} e^{-\left(\frac{B_1}{T}\right)} \right)^{\beta} \right\} \quad (21)$$

in (21) the Weibull/Arrhenius parameters are estimated in joint form by using the maximum likelihood method.

On the other hand, observe that the lifetime data is collected by using an experiment design (DOE) (see Table 2), and that for each replicated level, we estimate its corresponding Weibull parameters.

Thus, since, as in (16), n in the DOE determines the accuracy with which η is estimated, in order to show how n could be used to incorporate the expected failure times in an ALT analysis, first note that because under multiple linear regression, the Weibull parameters of each replicate level of the DOE are estimated by the median rank, approximation given by

$$\text{Median Rank (50\%)} = F(t) = \frac{i - 0.3}{n + 0.4} \quad (22)$$

where, i represents the ordered rank statistic and $F(t)$ is the empirical estimation of the cumulative probability function [$F(t) = 1 - R(t)$], then (22) highly depends on n too. Second, observe that because (22) was constructed based on an area given by $2a - (n-1)b = 1$ (for details see [8]), where b represents the amplitude (width) of the $(n-1)$ intervals given by $b = 1/(n+0.4)$ which for high percentiles ($p > 0.85$) tends to be $R(t) [1/(n+0.4) \approx R(t)]$ (e.g. for $R(t) = 0.96$,

$b = [1/(24.4966 + 0.4) \approx 0.96]$), then because n in (13) depends only on $R(t)$ and since $R(t)$ is generally higher than 0.85, then the use of (13) in (22) is useful. On the other hand, by taking the linear form of (9) as

$$\ln(-\ln(R(t))) = -\beta \ln(\eta) + \beta \ln(t) \quad (23)$$

The expected times of the operational level can be estimated and incorporated to the ALT analysis as in the next section.

5.1. Application to ALT analysis.

As an application, let us use data given in Table 3. Data was published by [17]. Suppose the normal operational temperature level is 323K and the design time is $t_d = 15000$ hrs. The analysis to incorporate the expected lifetimes of the 323K level into the analysis is as follows.

1. For each ALT level, by applying (13) with the desired $R(t)$ level, determine the number of replicates to be tested. Here $R(t) = 0.90$ was used, thus $n = 10$ pcs.
2. Perform the experiment, and for each replicated level, estimate the Weibull parameters β and η (here β should be constant).
3. Using the estimated n and β value in (16), estimate the corresponding ηL value. (here it is $\eta L = 101/4.2206$ (15000) = 25883.58312hrs).
4. Using (22) with n given by (13), estimate the expected $R(t_i)$ value for each of the order statistics. Then by using these $R(t_i)$ values, the β and the ηL values in (23), solve (23) to $\ln(t)$ and then determine the expected failure times of the operational level to each order statistic. (In our case, the expected failure times appear in the last row of Table 3).
5. By using the incorporated expected times, and the experimented accelerated lifetimes, perform the estimation of the Weibull/Life-stress parameters. Here, the analysis, by using the ALTA routine and (21), yield to $\beta = 4.3779$, $B = 2308.4884$ and $A = 19.8667$.
6. Finally, by using the parameters of step 5, determine the desired reliability indexes.

To conclude, observe that the data in Table 3, for $t_d = 15000$, $R(t)$ is of $R(t) = 0.90$ as was expected. In particular, note that without incorporating the operational times, the estimated $R(t)$ for the operational level is of 323K, instead of the designed $R(t) = 0.90$, would be $R(t) = 68.33\%$.

Table 3.
Data for Weibull Arrhenius analysis.

Stress	Time										Parameters by RXX	
	t ₁	t ₂	t ₃	t ₄	t ₅	t ₆	t ₇	t ₈	t ₉	t ₁₀	η	β
393K	3850	4340	4760	5320	5740	6160	6580	7140	7980	8960	6668.7828	4.2206
408K	3300	3720	4080	4560	4920	5280	5640	6120	6840	7680	5716.0996	4.2206
423K	2750	3100	3400	3800	4100	4400	4700	5100	5700	6400	4763.4163	4.2206
Expected data for operation stress level												
323K	13769.50	17206.91	19468.73	21304.73	22945.13	24508.69	26086.29	27783.98	29795.04	32746.95	25883.5831	4.2206

Source: The authors

6. Conclusions

In Weibull demonstration test plans, the binomial approach for confidence levels higher (lower) than 0.63212, overestimate (sub-estimate) $R(t)$; thus, C should not be selected lower than 0.63212, and if it is selected higher, it should be selected close to 0.63212, say 0.7. In the proposed method, given that n depends only on $R(t)$ which is always known, the designed and the demonstrated $R(t)$ value always holds. On the other hand, observe that because data is gathered by using an experiment design where each row represents a different form to run the process, then the β parameter does not remain constant and as a consequence, the multivariate approach using the Taguchi method as in [14], should be used. Given that in the estimation process, the value of β depends on the variance of the logarithm of the failure times, which depends on the control of environmental factors (see [13] and [14]), its value must be selected from a data set that covers the variability of the manufacturing process in its interval.

Although, the proposed method allows practitioners to incorporate the expected operational times to the ALT analysis, it is important to note that their efficiency corresponds to the efficiency of the median rank approach and on the assumption of a constant β . Finally, it is important to note that although it may seem that n could be used to incorporate the expected times in the ALT analyses where the Weibull closure property does not hold (β is not constant), as is the case of ALT analysis with several variables, more research is required.

References

- [1] Bagdonavičius, V. and Nikulin, M., Accelerated life models, modeling and statistical analysis, Ed. Chapman and Hall/CRC, Boca Ratón, FL, USA, 2002.
- [2] Bertsche, B., Reliability and automotive and mechanical engineering, 1ra. ed., Stuttgart, Springer-Verlag Berlin Heidelberg, 2008. DOI: 10.1007/978-3-540-34282-3.
- [3] Cox, D.R. and Oakes, D., Analysis of survival data, 1st Ed, Chapman and Hall, Boca Ratón, FL, USA, 1984.
- [4] Escobar, L.A. and Meeker, W.Q., A Review of accelerated test models. Statistical Science, 21(4), pp. 552-577, 2006. DOI: 10.1214/088342306000000321.
- [5] Kleyner, A.V., Reliability demonstration: Theory and application. Reliability and Maintainability Symposium (RAMS). Las Vegas, USA, Tutorials CD, 2008.
- [6] Manotas, E., Yañez, S., Lopera, C. y Jaramillo, M., Estudio del efecto de la dependencia en la estimación de la confiabilidad de un sistema con dos modos de falla concurrentes. DYNA. 75(154), pp. 29-38, 2007.
- [7] Meeker W.Q. and Escobar L.A., Statistical methods for reliability data, New York, John Wiley & Sons, USA, 2014.
- [8] Mischke, C.R., A distribution-independent plotting rule for ordered failures, Journal of Mechanical Design, 104(3), pp. 593-597, 1979. DOI: 10.1115/1.3256391.
- [9] Nelson, W.B., Accelerated testing statistical models, test plans and data analysis, John Wiley & Sons, New York, USA, 2004.
- [10] Nelson, W.B. Applied life data analysis, John Wiley & Sons, New York, USA, 1985.
- [11] Nelson, W.B., Weibull analysis of reliability data with few or no failures, Journal of Quality Technology, 17(3), pp. 140-146, 1985.
- [12] Nicholls, D. and Lein, P., Weibayes testing: What is the impact if assumed beta is incorrect?, Reliability and Maintainability Symposium (RAMS), Fort Worth, Texas, Tutorials CD, 2009.
- [13] Piña-Monárrez M.R., Avila-Chavez C. and Marquez-Luevano C.D., Weibull accelerated life testing analysis whit several variables using multiple linear regression, DYNA, 82(191), pp. 156-162, 2015. DOI: 10.15446/dyna.v82n191.43533.
- [14] Piña-Monárrez, M.R. and Ortiz-Yañez J.F., Weibull and lognormal Taguchi analysis using multiple linear regression. Reliability Engineering and System safety. 144, pp. 244-253, 2015. DOI: 10.1016/j.res.2015.08.004
- [15] Rinne, H., The Weibull distribution a handbook, CRC press, Boca Ratón, FL, USA, 2009.
- [16] Tobias, P. and Trindade, D., Applied reliability. Chapman and Hall/CRC, Boca Ratón, FL, USA, 2012
- [17] Vassiliou, P. and Metas, A., Understanding accelerated life-testing analysis, Reliability and Maintainability Symposium (RAMS), Tutorials CD, Seattle, WA, USA, 2002.
- [18] Weibull, W., A statistical theory of the strength of materials. Stockholm: Generalstabens litografiska anstalts förlag. 1939.

M.R. Piña-Monárrez, is a researcher-professor in the Industrial and Manufacturing Department at the Autonomous University of Ciudad Juárez, Mexico. He completed his PhD. in Science in Industrial Engineering in 2006 at the Technological Institute of Ciudad Juárez, Mexico. He had conducted research on system design methods including robust design, design of experiments, linear regression, reliability and multivariate process control. He is member of the National Research System (SNI-1), of the National Council of Science and Technology (CONACYT) in Mexico. ORCID: 0000-0002-2243-3400.

M.L. Ramos-López, is a PhD. student on the doctoral program in science and engineering (DOCI), at the Autonomous University of Ciudad Juárez, Mexico. She completed her MSc. degree in Industrial Engineering in 2012 at the Universidad Autónoma de Ciudad Juárez, Mexico. Her research is based on accelerated life time and weibull analysis. ORCID: 0000-0001-8614-1311.

A. Alvarado-Iniesta, is a researcher professor in the Industrial and Manufacturing Department at Universidad Autónoma de Ciudad Juárez, Chihuahua, Mexico. He completed his PhD. in Industrial Engineering in 2011 at New Mexico State University, Las Cruces, NM, USA. His research interests focus on Operations Research. He is member of the National Research System (SNI-1), of the National Council of Science and Technology (CONACYT) in Mexico. ORCID: 0000-0002-3349-4823

R.D. Molina-Arredondo, is a researcher-professor in the Industrial and Manufacturing Department at the Autonomous University of Ciudad Juárez, Mexico. He completed his PhD. in Science in Industrial Engineering in 2009 at the Technological Institute of Ciudad Juárez. He had conducted research on reliability and robust design. ORCID: 0000-0001-8482-4186.

A new hardening rule to model the effect of density on soil behavior

Márcio Muniz de Farias ^a, Robinson Andrés Giraldo-Zuluaga ^b & Teruo Nakai ^c

^a Geotechnical Research Group, Department of Civil and Environmental Engineering, University of Brasilia, Brasilia, Brazil, muniz@unb.br

^b Geotechnical Research Group, Department of Civil and Environmental Engineering, University of Brasilia, Brasilia, Brazil, roangizu@unb.br

^c Nagoya Institute of Technology, Geo-Research Institute, Nagoya, Japan, nakai.teruo@nitech.ac.jp

Received: November 7th, 2014. Received in revised form: September 14th, 2015. Accepted: January 20th, 2016.

Abstract

The authors propose a new hardening rule for constitutive models based on the subloading concept. This concept allows a smooth transition between unloading and reloading. Most importantly it characterizes the soil with a single set of parameters regardless of its initial density state. The new rule aims to improve the original model proposed by the last author. The original version included a linear evolution rule for the sake of simplicity, but did not include any empirical evidence for such law. This new version proposes an exponential law that provides a better fit to the experimental data. This law requires only two parameters with a clear physical interpretation. These parameters can be obtained from simple one-dimensional compression tests. The model was based on controlled tests performed with glass spheres, but was applied to real granular and clay soils. The results show that the model can very successfully reproduce the behavior of real soils.

Keywords: yield condition, constitutive relations, elastic-plastic material, laboratory tests.

Una nueva regla endurecimiento para modelar el efecto de la densidad en el comportamiento de los suelos

Resumen

Los autores proponen una nueva ley de endurecimiento para modelos constitutivos basados en el concepto *subloading*. Este concepto permite una transición suave entre la descarga y la recarga. Lo más importante es que caracteriza el suelo con un único grupo de parámetros, independientes de la densidad inicial. La nueva regla busca mejorar el modelo original propuesto por el último autor. La versión inicial incluía una regla de evolución lineal en aras de la simplicidad, sin incluir pruebas empíricas de dicha ley. Esta nueva versión propone una ley exponencial que se ajusta mejor los datos experimentales. Esta ley requiere sólo dos parámetros, con una interpretación física clara, obtenidos a partir de ensayos simples de compresión unidimensional. El modelo se basó en ensayos controlados con esferas de vidrio, sin embargo, fue aplicado a suelos granulares y arcillosos. Los resultados muestran que el modelo puede reproducir muy bien el comportamiento de suelos reales.

Palabras clave: condición de plastificación, relación constitutiva, material elastoplástico, ensayos de laboratorio.

1. Introduction

Constitutive models form a necessary set of equations in order to obtain analytical or numerical solutions for boundary value problems satisfying certain universal conservation laws at every point of a domain, in the context of continuum mechanics. Contrary to the universal conservation laws, the constitutive models are dependent on the material type and should reflect its internal constitution by means of macroscopic parameters. The constitutive laws are, therefore, mathematical

expressions that represent the views of researchers about how certain materials should behave, rather than the real behavior of the materials. Hence, the same set of experimental evidence may lead to many different and sometimes conflicting constitutive models depending on the general framework adopted (elastoplasticity, hyperelasticity, etc.) and on the particular views and background of the model proposer.

For the particular case of geomaterials, such as natural soils, an ever increasing number of models have been proposed. The development of these models generally follow two conflicting

How to cite: Farias, M.M., Zuluaga, R.A.G. and Nakai, T., A new hardening rule to model the effect of density on soil behavior. DYNA 83(197), pp. 58-67, 2016.

trends: some favor a more generalist approach trying to unify most materials and physical conditions under a single model; others consider that different models for each material and initial state conditions provide a simpler and more practical approach [1]. This latter view is shared by many practitioners who argue that the most adequate model should be employed in each case. This is a good point, but according to the authors of this paper, it is unfortunately, very common for geotechnical engineers to opt for over simplistic models that do not adequately represent the behavior of the materials involved in the analyses.

It is widely reported and accepted that the mechanical behavior of geomaterials is highly influenced by the initial conditions to which they are subjected. These initial conditions comprise both the stress level (confinement) and stress history (over-consolidation), as well as the initial physical indices (e.g. density) of the soil mass. Whereas the incorporation of the influence of the confinement level is quite straightforward in most constitutive models, the same does not happen with respect to the effect of stress history and most of all, the effects of the initial density and structure of the soil mass. Most models tend to bypass this problem by adopting different sets of parameters for different initial over-consolidation and density states, as if different materials were involved in the same analyses.

Nakai and Hinokio [2] proposed an original approach in which it was possible to incorporate the effects of both confinement and density into a reasonably simple constitutive model for three-dimensional loading conditions, using a single set of unified material parameters. Later, Nakai and collaborators working at Nagoya Institute of Technology proposed a theoretical scheme in which it was possible to take into consideration many other relevant variables which influence the mechanical and hydraulic behavior of soils and rocks [3,4]. The model in different stages can consider the effect of density (over-consolidation), structure (bonding), strain rate (creep), temperature and water content (suction).

The general framework summarizes years of research and is initially introduced in a very simple manner using only the effective stress versus void ratio relation, under one-dimensional compression condition. Later, the models are generalized for three-dimensional conditions by introducing the concept of a modified stress tensor called t_{ij} [5].

For the sake of simplicity and easy understanding of the theoretical background, only the one-dimensional condition is discussed in the present paper. The proposed formulation is compared initially with the experimental results obtained in laboratory by the authors, using very uniform glass beads in order to isolate other effects such as particle shape, structure and cementation. Finally, the model predictions are compared with experimental results for different types of real soils (cohesive and granular) published in the literature by other researchers.

2. Constitutive model

The model adopted in this paper uses a conceptual scheme based on isotropic strain-hardening elastoplasticity with the implicit introduction of the subloading concept proposed by Hashiguchi [6] and revised by Nakai and

Hinokio [2]. Using this framework it is possible to account for several important variables that influence soil behavior, such as density, structure, deformation rate, temperature and suction [3-5,7-10].

The modeling conception is divided into two main sets of variables. The first set accounts for internal variables, such as density and bonding, which modify the shape of the void ratio versus the effective stress compressibility curve. The second set includes external variables (rate of deformation, temperature and water content) whose main effect is to shift the position of the normal consolidation line (NCL). This paper will focus only on the influence of density, but a full description of other variables may be found in the above references.

Nakai *et al.* [3] initially show how to formulate models under one-dimensional conditions and later extend the analyses to full three-dimensional cases [4]. The extension to generalized 3D conditions is achieved by: (a) introducing a modified stress tensor t_{ij} [11]; and (b) assuming a decomposition of the plastic strain increment [12]. These two features, respectively, are responsible for accounting for the influence of the intermediate principal stress and the influence of stress path on the behavior. Alternative approaches using conventional stress invariants p , q and θ may be also effective, including applications to unsaturated soils [13].

Here, the simpler version will be explained with reference to the well-known void ratio (e) versus natural logarithm of the one-dimensional effective stress (σ) illustrated in Fig 1. For a normally consolidated material, the initial state is represented by point I (e_0, σ_0) = (e_{N0}, σ_0). Upon loading, the point follows a trajectory along the normal consolidation line (NCL), reaching a final point P (e, σ) = (e_N, σ), with $e = e_0 + d(-e)$ and $\sigma = \sigma_0 + d\sigma$. The subscript "N" denotes a point over the NCL. The inclination of the virgin NCL is denoted by the (one-dimensional) compression index λ , and the slope of the unloading-reloading line (URL) is given by the swelling index, κ . Under these conditions, the change in plastic void ratio, from simple geometric relations in Fig. 1, is given by:

$$-(\Delta e)^p = -(\Delta e) - (\Delta e)^e = (\lambda - \kappa) \ln \frac{\sigma}{\sigma_0} \quad (1)$$

For the sake of simplicity, let F and H denote the terms related to (logarithmic) stress change and plastic void ratio change in Eq. (1), respectively, so that:

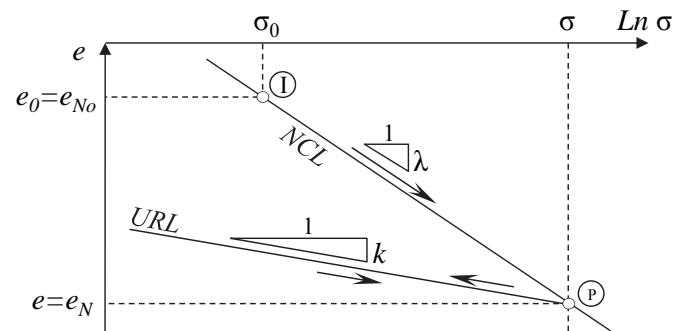


Figure 1. Changes in void ratio for normally consolidated clays. Source: The authors.

$$F = (\lambda - \kappa) \ln \frac{\sigma}{\sigma_0} \quad (2)$$

$$H = (-\Delta e)^p \quad (3)$$

Then the yield function (f) can be written generically as follows:

$$f = F - H = 0 \quad (4)$$

The consistency condition upon loading ($df=0$) implies:

$$df = dF - dH = (\lambda - \kappa) \frac{d\sigma}{\sigma} - d(-e)^p = 0 \quad (5)$$

$$d(-e)^p = (\lambda - \kappa) \frac{d\sigma}{\sigma} \quad (6)$$

Eq. (6) could also be directly obtained from differentiation of Eq. (1), but all steps were used only to illustrate the approach adopted in Nakai *et al.* [3] and later extended to other internal variables. For normally consolidated soils, Eq. (6) represents the incremental hardening law which relates changes of an internal strain-like hardening variable, represented by the plastic void ratio $d(-e)^p$, with changes of the internal stress-like hardening variable, given by $\sigma_o = \sigma$ or the maximum stress previously applied to the soil (pre-consolidation stress). This is akin to most conventional models based of the critical state theory, such as the Cam clay model.

Now, a new (strain-like) internal hardening variable, denoted by ρ , is introduced to account for the influence of density or pre-consolidation of the soil into the model. Referring to Fig. 2-b, this variable simply represents the deviation between the actual void ratio (e) in the present over-consolidated state (point I) and the corresponding void ratio (e_N) on the NCL for a normally consolidated soil under the same stress level (point I').

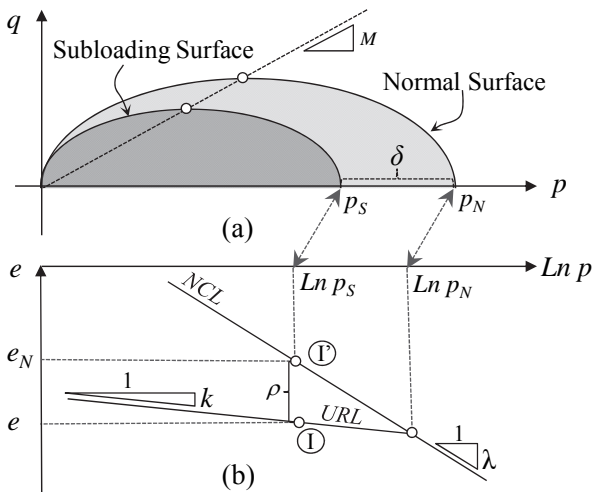


Figure 2. Internal variable δ and implicit subloading surface; (b) new strain-like internal variable ρ .
Source: The authors.

This new internal variable ρ is non-dimensional and represents a change in void ratio with respect to a reference normally consolidated state. This change may be due to pre-consolidation (single loading and unloading), but can also incorporate other effects such as cyclic loading. A null value for variable ρ means that the soil is normally consolidated and a high value of ρ denotes a very dense condition (much lower void ratio) with respect to the soil density if it were normally consolidated. So, variable ρ represents a densification effect due to pre-consolidation (and cyclic loading).

The relation between variable ρ and pre-consolidation can be better visualized using the p - q stress space of the Cam clay model as depicted in Fig. 2-a. Two yield surfaces are drawn: (i) the outer or “normal” surface whose “size”, given by its intersection with the average stress axis (p), corresponds to the maximum or pre-consolidation stress (p_N); (ii) the inner surface, called “subloading” surface, which always passes through the present stress state, and its size is denoted by (p_S). The distance ($\delta = p_N - p_S$) between the two surfaces is a measure of pre-consolidation and can be related to the over-consolidation ratio ($OCR = p_N / p_S$), $\delta = p_S (OCR - 1)$. From Fig. 2, it is possible to establish the following relation between ρ and OCR [14]:

$$\rho = (\lambda - \kappa) \ln \left(\frac{p_N}{p_S} \right) \text{ or } \rho = (\lambda - \kappa) \ln(OCR) \quad (7)$$

The use of two surfaces is the essence of modeling approaches such as subloading surface [6] and bounding surface [15], despite some basic differences in these approaches, especially regarding the influence of the confining pressure. While the inner surface tracks the present or actual stress state, the outer surface keeps a kind of memory of the loading history. The distance between the two surfaces can be measured by stress-like (δ) or strain-like (ρ) internal hardening variables, which are interrelated. In either case, it is necessary to devise an evolution law for the hardening variable, giving rise to different models and interpretations.

The authors favor the use of the strain-like density variable ρ . It should be emphasized that the subloading concept was evoked here only for interpretation purposes and that the modeling approach does not require the explicit definition of two surfaces.

Fig. 3 illustrates the changes in void ratio for an over-consolidated soil loaded from the initial state at point I (e_o, σ_o) to a final state at point P (e, σ), with $e = e_o + \Delta(-e)$ and $\sigma = \sigma_o + \Delta\sigma$. The change of plastic void ratio $(-\Delta e)^p$ upon loading from σ_o to σ can be expressed as:

$$\begin{aligned} -(\Delta e)^p &= -(\Delta e) - (-\Delta e)^e \\ &= \{(e_{N0} - e_N) - (\rho_0 - \rho)\} - (-\Delta e)^e \\ &= \lambda \ln \frac{\sigma}{\sigma_0} - (\rho_0 - \rho) - \kappa \ln \frac{\sigma}{\sigma_0} \end{aligned} \quad (8)$$

Hence the yield function can be written as:

$$f = F - \{H + (\rho_0 - \rho)\} = 0 \quad (9)$$

and the consistency condition ($df=0$) applied to Eq. (9), requires that:

$$\begin{aligned} dF - \{dH - d\rho\} \\ = (\lambda - \kappa) \frac{d\sigma}{\sigma} \\ - \{d(-e)^p - d\rho\} = 0 \end{aligned} \quad (10)$$

Variable ρ needs an evolution law (or hardening law). The research group of NIT suggested that its increment ($d\rho$) should be negative and proportional to the change of plastic void ratio $d(-e)^p$, and also dependent on the present value of density ρ by means of an arbitrary increasing function, $G(\rho)$. This can be expressed mathematically as follows:

$$d\rho = -G(\rho) \cdot d(-e)^p \quad (11)$$

Substituting the evolution law of Eq. (11) into Eq. (10), the change in plastic void ratio can be expressed as follows:

$$d(-e)^p = \left(\frac{\lambda - \kappa}{1 + G} \right) \frac{d\sigma}{\sigma} \quad (12)$$

There are only two conditions established for function $G(\rho)$: (i) it must be a monotonically increasing function of ρ and (ii) it should satisfy $G(0)=0$. The first condition assures that the value of variable ρ degrades faster for higher densities and the second condition assures that the evolution ceases when $\rho=0$, thus forcing the deformability curve to adhere to the NCL when the soil becomes normally consolidated. Following these conditions, Nakai *et al.* [8] suggested that linear ($G(\rho)=a\rho$) or quadratic rules ($G(\rho)=a\rho^2$) could be adopted. The main argument for the adoption of these functions is their simplicity and the fact that just one parameter (a) is added to the model. This parameter controls the decay rate between the pre-consolidated and normally consolidated stress states.

The absolute value of $G(\rho)$ can be interpreted as an increase in the volumetric plastic stiffness of the soil compared to the stiffness value that the soil would have if it were normally consolidated. This is easily achieved by expressing Eq. (12) in terms of volumetric strains (dividing by $1+e_o$) and inverting it:

$$d\sigma = \left(\frac{1 + e_o}{\lambda - \kappa} \sigma (1 + G) \right) d\varepsilon_v^p \quad (13)$$

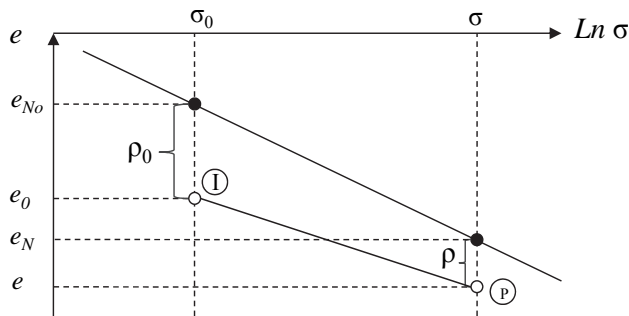


Figure 3. Changes in void ratio for over-consolidated soils. Source: The authors.

The term between brackets in Eq. (13) corresponds to the bulk or volumetric plastic stiffness (K^p) and it coincides exactly with the expression obtained for the Cam clay model when $G(\rho)=0$ (if the one-dimensional stress σ is substituted by the mean stress p), i.e., $K^p=p(1+e_o)/(\lambda-\kappa)$. The factor $(1+G(\rho))$ in K^p in Eq. (13) gives a scalar measure of this stiffness gain when the soil becomes over-consolidated [16]. This approach can be easily implemented in critical state based models for simulating geotechnical problems, mainly for applications where the soil undergoes severe densification, or in cases of cyclic loadings, including the simpler case of a single unloading-reloading cycle [17, 18, 19, 20].

3. Experimental program

The linear and quadratic hypotheses for the evolution function $G(\rho)$ were proposed based solely on their simplicity. However, these proposed shapes for $G(\rho)$ lack experimental backing, which is what is investigated in this paper. In order to check their validity, the authors devised a series of experimental tests with different initial void ratios (hence different initial values of ρ_o) and tracked their evolution.

The tests were initially performed with artificial sand comprised of almost perfect microspheres of glass beads. The idea was to test a well-behaved material, thus avoiding other interferences such as particle shape, natural structure and cementation. The particle diameters (D_p) and sample diameter (D_s) were chosen so as to minimize scale effects ($D_s/D_p > 100$). Conventional soil characterization tests, such as grain size distribution, density, minimum and maximum void ratios (e_{min} and e_{max}) etc., were performed with this material.

Table 1 summarizes the results of characterization tests on the glass beads. The package of microspheres can be classified as fine sand according to the ASTM standards. Photos of the glass beads amplified by 200x and 400x using an electronic magnifying glass (Digital USB Microscope) are shown Fig. 4 (a)-(b). The sphericity and uniformity of the microspheres can be appreciated in these photos.

Artificial soil samples were prepared by pouring the glass microspheres into a metallic cylinder with a diameter of 25 mm and a height of 50 mm (Fig. 4-c). The material was carefully placed in its loosest state and vibrated until the desired initial void ratio was reached by trial and error. Samples starting with seven initial void ratio states, from e_{max} down to close to e_{min} , were prepared: $e_o = 0.87-0.85-0.81-0.79-0.77-0.76-0.72$.

Table 1. Physical properties of the microspheres.

Property, symbol (unit)	Value
Maximum void ratio, e_{max}	0.87
Maximum void ratio, e_{min}	0.63
Specific gravity, G_s	2.51
Sphericity (% min)	80
Coefficient of uniformity, C_u	1.47
Coefficient of curvature, C_c	1.02
Average diameter D_{50} (mm)	0.063

Source: The authors.

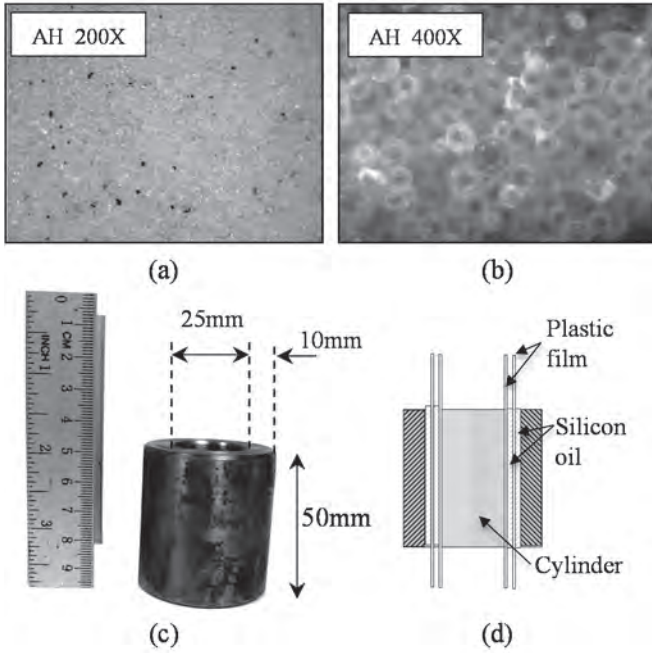


Figure 4. Details of material and cylindrical mold: (a) and (b) magnified images of the microspheres; (c) mold dimensions; (d) mechanisms to reduce lateral friction. Source: The authors.

The wall of the metallic mold was 10 mm thick to avoid radial deformation and ensure the desired one-dimensional strain compression state. In order to minimize wall friction, the interior of the cylinder was coated with silicon oil and two layers of plastic film also coated with oil in between them, as illustrated in Fig. 4-d.

The compressibility curves were determined under one-dimensional constant strain rate (CSR). The main advantages of this methodology are related to speed and precision [21]. The test equipment comprised a load frame with a reaction beam, a moving load plate, and sensors for load and displacement (LVDT). The strain rate was fixed at 1 mm/min and the test was carried until a maximum vertical force of 40 kN, due to limitations of the reaction system. The corresponding maximum vertical stress, around 80000 kPa, was high enough to observe particle crushing in the cases of spheres with larger diameters (not reported here).

Besides the artificial sand prepared with microspheres of glass, the constitutive model was also verified against the results of CSR one-dimensional tests on real soils compiled from the literature. The main idea was to test the hypotheses that void ratio deviation from the NCL, variable ρ , provides an appropriate measure of density which controls the soil behavior irrespective of particle shape and nature (sand or clay). The tests include two sands and two clays for which CSR one-dimensional test results were available for different initial densities. A quartz sand was tested by Nakata *et al.* [22] and Ganga sand was tested by Rahim [23]. Tests performed with London clay and Cambridge clay were also used [24].

4. Methodology for the calibration of model parameters

The compression (λ) and swelling (k) indices are quite easy to calibrate and may be obtained directly from the

inclinations of straight lines fitted to the initial and final sections of the compressibility curve, respectively. The straight line fitting is best achieved using least square error minimization for selected ranges of experimental points in the initial and final sections of the e - $\ln(\sigma)$ curve.

In order to find the actual evolution rule of the internal variable ρ , it is necessary to first determine the experimental values of $G(\rho)$ versus ρ directly from the test results. The following methodological steps are proposed:

- Generate an adjusted curve fitting the experimental e - $\ln(\sigma)$ data to obtain an expression for the stress σ as a function of the void ratio, $\sigma=f(e)$. This can be achieved with a high order polynomial, a spline curve, or piecewise linear fittings. This curve is used only as an intermediate stage in order to recover the data for equally spaced void ratio intervals and avoid the ragged nature of the experimental data;
- For equally spaced points ($\Delta e=\text{constant}$) in the compressibility curve, obtain the elastic and plastic void ratios and the internal variable ρ following the steps of the algorithm described in Table 2;
- Compute the changes in variable ρ and the corresponding changes in plastic void ratio, $d\rho$ and $d(-e)^p$, with respect to the previous values.
- Following Eq. 11, determine the value of $G(\rho)$ for each point of the compressibility curve as the ratio:

$$G(\rho) = \frac{d\rho}{d(-e)^p} \quad (14)$$

Following the steps described above, the authors were able to find the experimental relation for $G(\rho) \times \rho$, depicted by the open dots in Fig. 5, for a particular test with the densest sample ($e_0=0.72$). The experimental data, however, do not show the linear nor the quadratic trends proposed by Nakai *et al.* [8]. The same was observed for all tests performed by the authors.

It may be observed that the initial section of the data in Fig. 5 may be fitted with a straight line with a non-zero inclination, $G(\rho)=a\rho$. However if a straight line is fitted over the whole range of variable ρ , it will overestimate $G(\rho)$ for low values of ρ and underestimate the function for high values of ρ . The initial non-zero inclination of the experimental data cannot be captured with a simple quadratic curve of the type $G(\rho)=a\rho^2$, because its first derivative at the origin is zero. On the other hand, the experimental data show that the relation is clearly non-linear over the range of density tested, and that the inclination increases faster after a certain point. A hyperbolic function can meet these requirements

Table 2.
Algorithm to find values of variables ρ and e^p .

Variable	Expression
Starting from $e=e_0$, increment e	$e \leftarrow e + \Delta e$
Find stress (σ) using function f	$\sigma = f(e)$
Compute elastic void ratio $(-e)^e$	$(-e)^e = -k \ln(\sigma) + I_k$
Compute plastic void ratio $(-e)^p$	$(-e)^p = (-e) - (-e)^e$
Compute void at ratio NCL (e_N)	$e_N = -\lambda \ln(\sigma) + I_\lambda$
Compute variable ρ	$\rho = e_N - e$

Source: The authors.

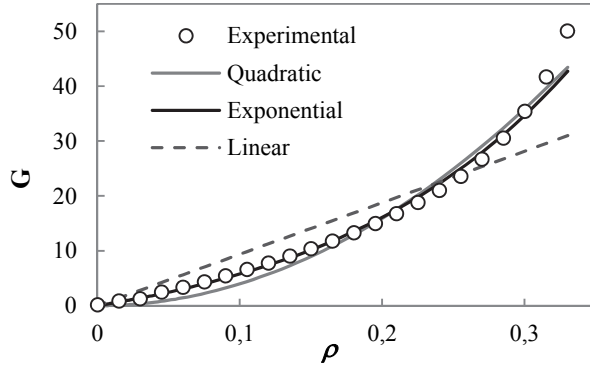


Figure 5. Experimental values of function $G(\rho)$ and alternative fitting curves.

Source: The authors.

[16] or maybe a quadratic function with two terms ($G(\rho)=a\rho^2+b\rho$), but this behavior can be better simulated with an exponential function, as follows:

$$G(\rho) = \frac{a}{b} (e^{b \cdot \rho} - 1) \quad (15)$$

in which a and b are material parameters. The derivative of Eq. (15) is $G'(\rho) = a \cdot e^{b \cdot \rho}$, and the parameter a mathematically represents the inclination when $\rho=0$, i.e., $a=G'(0)$, while parameter b gives the rate of increase of this inclination which is related to the second derivative of $G(\rho)$.

The proposed function is monotonically crescent, satisfies $G(0)=0$, and exhibits the trends of initial non-zero inclination and fast increase after a certain value. However, the exponential model requires two parameters (a and b) instead of just one (a) of the original linear or quadratic functions. Nevertheless, these parameters are easily calibrated with a simple procedure which starts by rewriting Eq. (15) as follows:

$$\chi = \ln\left(\frac{a}{b} G + 1\right) = b \cdot \rho \quad (16)$$

Eq. (16) represents a linear relation between an auxiliary variable χ and ρ . The best fitting is obtained by varying the ratio a/b until a good linear correlation ($R^2 \approx 1.0$) is obtained through the points (ρ, χ) . The inclination of this straight line is parameter b , and the other parameter a follows directly from the product between b and the ratio a/b .

Fig. 6 shows the results of parametric analyses to highlight the influences of parameters a and b on the overall compressibility curve, $e-\ln(\sigma)$. In these simulations, the initial void ratio was $e_0=0.78$ for the stress $\sigma_0=0.1$ kPa, and the conventional compressibility indices were $\lambda=0.120$ and $\kappa=0.03$. From these analyses, it can be observed that parameters a and b control the adherence between the actual compressibility curve and the straight sections represented by the normal consolidation line (NCL) and unloading-reloading line (URL), respectively.

The influence of parameter a can be appreciated in Fig. 6-a, in which the value of $b=10$ was kept constant and the value of a varied between 10 and 250. The parametric analyses show that the compressibility curve joins the NCL faster (i.e., at higher values

of void ratio, e) for lower values of parameter a . Parameter b , on the other hand, controls the adherence between the actual curve and the unloading-reloading line as illustrated in Fig. 6-b, where the value of $a=20$ was kept constant and the value of b varied between 1 and 50. Here it can be observed that the compressibility curve sticks to the URL section for lower values of b . In summary, low values of both a and b simulate a sharp bilinear transition as in the conventional Cam clay model.

The experimental results obtained by the authors using microspheres as well as results from tests on real soils published in the literature were used to calibrate the parameters of the model (λ , κ , a and b) following the procedures described in the previous section. These parameters are summarized in Table 3.

The first observation during the calibration process was that the swelling (k) was not exactly constant, but increased slightly for lower values of initial density of the sample ($k=0.014$ to 0.019). This is contrary to the basic assumption of conventional models, such as the Cam clay. In order to keep the model simple, this hypothesis was preserved and the values of these parameters were taken as the average for all the tests on a given material.

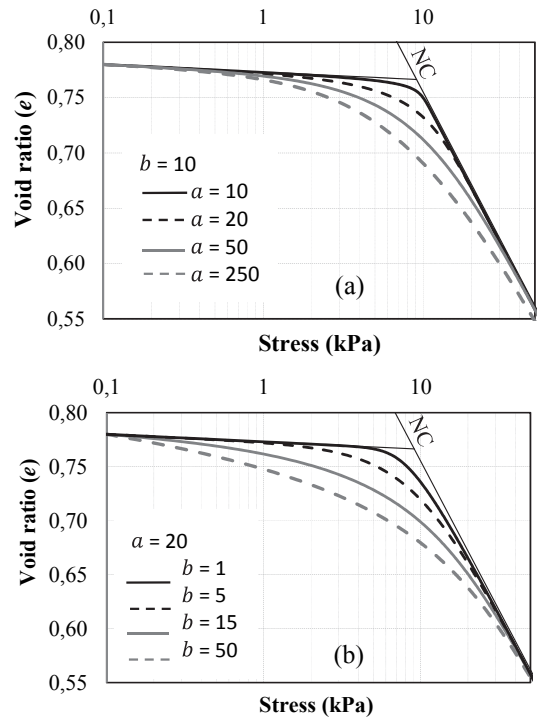


Figure 6. Parametric analyses: (a) Influence of a ; (b) Influence of b . Source: The authors.

Table 3. Model parameters.

Material	(λ)	(κ)	Exponential a	Exponential b	Linear a	Quadratic a
Microspheres	0.138	0.014	42.5	5.8	93.8	398
Ganga sand	0.327	0.006	6.4	1.9	16.7	19.8
Quartz sand	0.195	0.002	29.9	1.9	41.7	94.8
London clay	0.167	0.010	45.1	27.9	172.8	2432
Cambridge clay	0.210	0.007	12.8	8.4	36.1	213

Source: The authors.

Table 4.

Coefficients of correlation (R^2) for different functions $G(\rho)$.

Material	Exponential	Quadratic	Linear
Microspheres	0,998	0,993	0,959
Ganga sand	0,982	0,915	0,952
Quartz sand	0,982	0,915	0,958
London clay	0,948	0,926	0,795
Cambridge clay	0,944	0,835	0,741

Source: The authors.

The parameters for the evolution function $G(\rho)$ of the density variable were taken for the material at its densest state. The values parameters a (or a and b , for the exponential model) represent the best fit obtained by a least square error minimization process. The computed values of $G(\rho)$ were compared with the experimental ones and the coefficients of correlation (R^2) computed. These values are summarized in Table 4. The exponential model gave the best correlation for all materials, as expected. Compared with the linear model, the parabolic model yielded better correlations for natural clays, but slightly worse correlations for natural sands.

5. Model predictions and experimental results

A good fit of $G(\rho)$ is important, but the crucial point is the correlation between the complete model simulations with the whole experimental compressibility curve, $e-\ln(\sigma)$. These curves will be shown for all materials in this section, but a single case for the quartz sand is initially commented in more details.

A comparison between experimental $e-\ln(\sigma)$ data for the quartz sand and the simulations using three different functions for $G(\rho)$ is shown in Fig. 7-a. With the exponential function for $G(\rho)$, the overall simulation fitted the experimental data almost perfectly; while the linear function best fitted the last section of the curve (the NCL) and the quadratic function best fitted the initial section of the curve (the URL).

This result is explained when the actual values of $G(\rho)$ are compared to their simulations using the linear, quadratic and exponential models, as shown in Fig. 7-b. For this quartz sand, the experimental values of $G(\rho)$ showed a crescent but unexpected curved shape. This curve fitted reasonably well with a straight line for low values of ρ , hence the good overall agreement close to the NCL in Fig. 7-a for the model with a linear function for $G(\rho)$. The opposite is observed with the quadratic function for $G(\rho)$, which fits best for higher values of ρ , hence the good agreement close to the URL in the overall compressibility curve in Fig. 7-a.

The excellent agreement of the exponential model is due to its two parameters, one controlling the simulation in the URL range and the other in the NCL domain, as explained in the parametric analyses shown in the previous section. Therefore, this model will be used in all following simulations presented in this paper.

Next, further results of tests on some real soils described in the literature and the model simulations are described. Initially the results for quartz sand are completed by simulations with two initial densities, as shown in Fig. 8-a. The simulations can very successfully reproduce the compressibility curves over the whole domain. The same can be observed for the Ganga sand shown in Fig. 8-b.

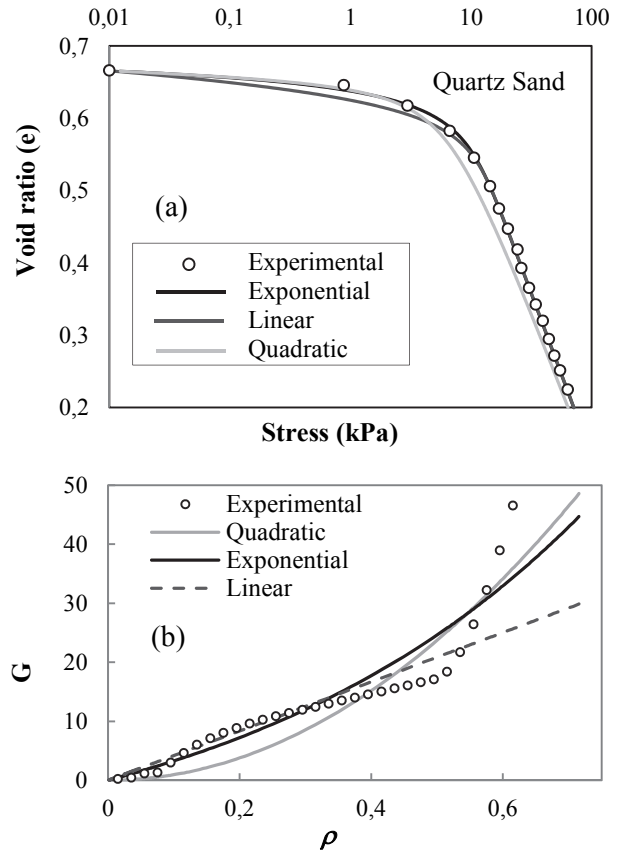


Figure 7. Simulations: (a) the overall compressibility curve using different functions for $G(\rho)$. (b) the evolution function $G(\rho)$ for the quartz sand. Source: The authors.

The experimental results and model simulations for London clay and Cambridge clay are shown in Fig. 9-a and Fig. 9-b, respectively. Three different initial conditions were available for London clay and two initial states for Cambridge clay. In both cases, the agreement between model simulations and experimental results was excellent for the entire range of stresses and densities analyzed.

The results for all tests using microspheres and the model simulation with exponential function $G(\rho)$ are shown in Fig. 10. For the average value of $\lambda=0.138$ in Table 3, and a reference stress equivalent to the atmospheric pressure ($\sigma_o = \sigma_{atm} = 100$ kPa), the following values for the initial density variable were computed, $\rho_o=0.51-0.54-0.58-0.60-0.62-0.64-0.66$. These values correspond to the whole range of void ratio values, between e_{min} and e_{max} presented previously. It is striking that the simulations can very successfully reproduce the compressibility curves over the whole domain for all range of initial density conditions with a single set of parameters.

In all the tests involving natural soils (sand and clay) and artificial sand (microspheres) a unique set of parameters was used for each material, regardless of the initial void ratio. Also, the reader can observe a smooth transition between the over-consolidated and normally consolidated states. This contrasts with the discontinuity in compressibility indices (λ

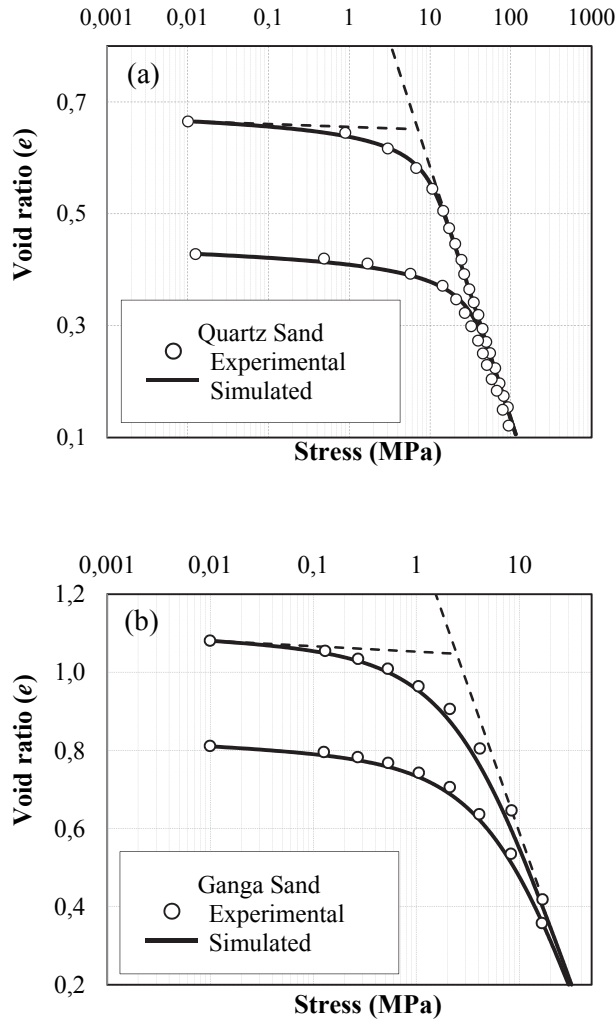


Figure 8. Simulations and experimental results of tests on quartz sand (a) and Ganga sand (b) with two initial density states.
Source: The authors.

and κ) postulated by the bilinear behavior assumed, for instance, in the Cam clay model. This sharp transition will be observed also in the simulation of conventional triaxial tests on over-consolidated soils.

The smoothness in compressibility is simulated by any model which implicitly or explicitly adopts the subloading concept. This smooth variation comes from the introduction of the scalar function $G(\rho)$ which reproduces plastic strains for any loading condition, irrespective of the previous stress history, i.e., regardless of whether the soil is normally consolidated or over-consolidated. This can be easily implemented in many models and all it takes is the addition of a scalar, related to $G(\rho)$, into the plastic multiplier of conventional elastoplastic models (see e.g. Pedroso *et al.*[14]).

6. Conclusions

This paper revises a simple and versatile model proposed

by Nakai *et al.* [3] to simulate the stress versus void ratio curve of geomaterials under one-dimensional compression conditions, as in oedometer tests.

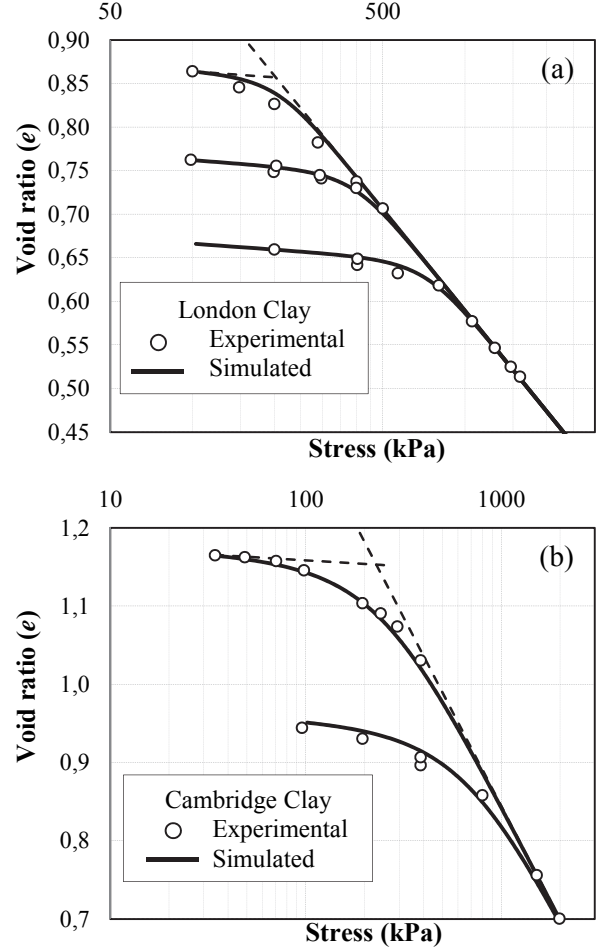


Figure 9. Simulations and experimental results of tests on London clay (a) and Cambridge clay (b) with different initial density states, respectively.
Source: The authors.

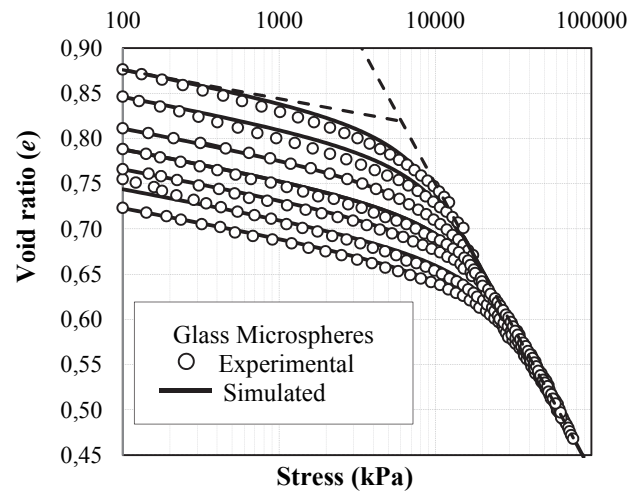


Figure 10. Simulations and experimental results of tests on artificial sand (glass microspheres) with seven initial density states.
Source: The authors.

The model implicitly incorporates the subloading concept which results in a smooth transition between the over-consolidated and normally consolidated stress states. This is achieved by means of an internal strain-like hardening variable, ρ , which is at the same time a measure of the density and over-consolidation ratio. This variable needs an evolution rule controlled by a smooth monotonic crescent function $G(\rho)$, for which the original model proposed a linear or a quadratic variation.

The above hypothesis was tested using the results of constant strain tests on artificial sand composed of glass microspheres and also tests on real soils (sand and clays) published in the literature. The authors proposed methodological steps to find the experimental values of $G(\rho)$ and the results showed that a better description of this function is given by an exponential relation with two model parameters (a and b).

These two parameters are easily calibrated from a single one-dimensional compression test and they also have a clear meaning: parameter a controls the behavior in the normally consolidated range and parameter b controls the behavior in the over-consolidated domain.

The simpler linear and quadratic relations, although not as accurate as the exponential, also produce good simulations of the overall compressibility curve under one-dimensional conditions. This may be enough for most practical applications and they have the advantage of using a single parameter (a). The limited experimental evidence shown in this paper indicates that the linear model best fits the results obtained with natural sands, while the quadratic model best reproduces the behavior of natural clays.

Irrespective of the shape of $G(\rho)$, the results support the use of the internal variable (ρ) as a powerful variable to incorporate density related effects on the stress-strain behavior of geomaterials, independently of grain shape or nature. The simulations could accurately reproduce the behavior of artificial glass microspheres, natural sands and clays, over a wide range of stress and void ratio conditions, using a single set of unified parameters.

In spite of presenting the formulation and experimental validation under one-dimensional compression conditions, the subloading concept described here can be easily incorporated in any model under general three-dimensional stress conditions. This has already been carried out by many others [4].

Acknowledgments

The authors acknowledge the financial support of the Brazilian National Research Council (CNPq).

References

- [1] Papamichos, E., Constitutive laws for geomaterials. *Oil Gas Science and Technology*, 54(6), pp. 759-771, 1999. DOI: 10.2516/ogst:1999064
- [2] Nakai, T. and Hinokio, M., A simple elastoplastic model for normally and over consolidated soils with unified material parameters. *Soil and Foundation*, 44(2), pp. 53-70, 2004. DOI: 10.3208/sandf.44.2_53
- [3] Nakai, T., Shahin, H., Kikumoto, M., Kyokawa, H., Zhang, F. and Farias, M.M., A simple and unified one-dimensional model to describe various characteristics of soils. *Soil and Foundation*, 51(6), pp. 1129-1148, 2011. DOI: 10.3208/sandf.51.1129
- [4] Nakai, T., Shahin, H., Kikumoto, M., Kyokawa, H., Zhang, F. and Farias, M.M., A simple and unified three-dimensional model to describe various characteristics of soils. *Soil and Foundation*, 51(6), pp. 1149-1168, 2011. DOI: 10.3208/sandf.51.1149
- [5] Nakai, T., Shahin, H.M., Kikumoto, M., Kyokawa, H. and Zhang F., Simple and unified method for describing various characteristics of geomaterials. *Journal of Applied Mechanics JSCE*, 19, pp. 371-382, 2009. (in Japanese).
- [6] Hashiguchi, K., Constitutive equation of elastoplastic materials with elasto-plastic transition. *Journal of Applied Mech. ASME*, 102(2), pp. 266-272, 1980. DOI: 10.1115/1.3153653
- [7] Nakai, T., Description of elastoplastic behavior of soil as a one-dimensional model. *Magazine of the Japanese Geotechnical Society*, 56(9), pp. 6-9, 2008 (In Japanese).
- [8] Nakai, T., Kyokawa, H., Kikumoto, M., Shahin, H.M. and Zhang, F., One-dimensional and three-dimensional descriptions of elastoplastic behavior in structured clays. *Proc. of 4th International Workshop on New Frontiers in Computational Geotechnics, IWS-Pittsburgh*, pp. 3-12, 2004.
- [9] Shahin, H.M., Nakai, T., Kikumoto, M., Kyokawa, H. and Miyahara, Y., Modeling of time dependent behavior of clay in one-dimensional consolidation. *Proc. of the 4th Sino-Japan Geotechnical Symposium*, pp. 54-61, 2010.
- [10] Kyokawa, H., Kikumoto, M., Nakai, T. and Shahin, H.M., Simple modeling of stress-strain relation for unsaturated soil. *GeoShanghai - International Conference, China*, pp. 17-25, 2010.
- [11] Nakai, T. and Mihara, Y., A new mechanical quantity for soils and its application to elastoplastic constitutive models. *Soil and Foundation*, 24(2), pp. 82-94, 1984. DOI: 10.3208/sandf1972.24.2_82
- [12] Nakai, T. and Matsuoka, H., A generalized elastoplastic constitutive model for clay in three-dimensional stresses. *Soil and Foundation*, 26(3), pp. 81-98, 1986. DOI: 10.3208/sandf1972.26.3_81
- [13] Pedroso, D.M. and Farias, M.M., Extended Barcelona basic model for unsaturated soils under cyclic loadings. *Computers and Geotechnics*, 38, pp. 731-740, 2011. DOI: 10.1016/j.compgeo.2011.02.004
- [14] Pedroso, D.M., Farias, M.M. and Nakai, T., An interpretation of subloading t_{ij} model in the context of conventional elastoplasticity theory. *Soil and Foundation*, 45(4), pp. 61-77, 2005. DOI: 10.3208/sandf.45.4_61
- [15] Dafalias, Y.F. and Popov, E.P., A model of nonlinearly hardening materials for complex loading. *Acta Mechanica*, 21, pp.173-192, 1975.
- [16] Zuluaga, R.A.G. and Farias, M.M., Experimental validation of a simple model for soils. *Proc. of VI Infogeo. Brazilian Symposium on Applications of Informatics to Geotechnics, Brasilia, V01*, pp. 11-20, 2011. (In Portuguese).
- [17] Farias, M.M., Moraes, A.H. and Assis, A.P., Displacement control in tunnels excavated by the NATM: 3-D numerical simulations. *Tunnelling and Underground Space Technology*, 19, pp. 282-293, 2004. DOI: 10.1016/j.tust.2003.11.006
- [18] Farias, M.M., Nakai, T., Shahin, H.M., Pedroso, D.M., Passos, P.G.O., Hinokio, M., Ground densification due to sand compaction piles. *Soil and Foundation*, 45(2), pp. 167-180, 2005. DOI: 10.3208/sandf.45.2_167
- [19] Nakai, T., Farias, M.M., Bastos, D. and Sato, Y., Simulation of conventional and inverted braced excavations using subloading t_{ij} model. *Soil and Foundation*, 47(3), pp. 597-612, 2007. DOI: 10.3208/sandf.47.597
- [20] Rebollo, J.F.R., Auvinet-Guichard, G.Y., Martínez-Carvajal, H.E., Settlement analysis of friction piles in consolidating soft soils, *DYNA*, 82(192), pp. 211-220, 2015. DOI: 10.15446/dyna.v82n192.47752
- [21] Wissa, A.E.Z., Christian, J.T., Davis, E.H. and Heiburg, S., Consolidation at constant rate of strain. *Journal of Soil Mechanics and Foundations Division*, 10, pp. 1393-1409, 1971.
- [22] Nakata, Y., Hyodo M., Hyde, A.F.L., Kato, Y. and Murata, H. Microscopic particle crushing of sand subjected to high pressure one-

- dimensional compression. *Soils and Foundations*; 41(1), pp. 69-82, 2001. DOI: 10.3208/sandf.41.69
- [23] Rahim, A., Effect of morphology and mineralogy on compressibility of sands. PhD thesis, Indian Institute of Technology Kanpur, Kanpur, India, 1989.
- [24] Butterfield, R. and Baligh, F., A new evaluation of loading cycles in an oedometer. *Geotechnique*; 46(3), pp. 547-553, 1996. DOI: 10.1680/geot.1996.46.3.547

M. Muniz de Farias, completed his BSc. in Civil Engineering in 1983 at the Federal University of Ceará, Brazil, his MSc. degree in Geotechnics in 1986, at the Pontifical Catholic University of Rio de Janeiro, Brazil, and his PhD. degree in Numerical Methods in 1993, at the University of Wales, Swansea. He spent his post-doctoral sabbatical at Nagoya Institute of Technology – NIT, Japan in 1998, Japan. He has been working at the University of Brasília (UnB) since 1986, and is a researcher for the Brazilian National Council for Scientific and Technological Development (CNPq). He has major teaching and research experience in the fields of Geotechnics and Pavements, and his main research interests include numerical and constitutive modeling applied to dams, tunnels and mechanistic pavement design.
ORCID: 0000-0002-5257-911X

R.A.G. Zuluaga, completed his BSc. in Civil Engineering in 2008 at the Faculty of Mines of the National University of Colombia, Medellín, Colombia, and his MSc. degree in Geotechnics in 2011, at the University of Brasília, Brazil. Currently, he is a doctoral student in Geotechnics at the University of Brasília, Brazil. His research interests include the behavior of geomaterials, constitutive modeling and micromechanics.
ORCID.ORG/0000-0001-6674-1977

T. Nakai, completed his BSc. and MSc. degrees in Civil Engineering in 1972 and 1974, respectively at Kyoto University, and the Dr. of Engineering degree in 1981 at Kyoto University, Japan. He became a professor of civil engineering at the Nagoya Institute of Technology in 1991. His research interests include, (a) laboratory testing of geomaterials and their constitutive modeling in general stress systems, (b) the application of the constitutive model to boundary value problems such as tunneling, braced excavation, bearing capacity of foundations, reinforced soils and other soil-structure interaction problems, and the corresponding model tests. He was awarded the prize for active young researcher in 1982, the prize for excelling research papers in 1991, and the prize for best papers in 2005, from the Japanese Geotechnical Society.
ORCID: 0000-0002-1346-3560



UNIVERSIDAD NACIONAL DE COLOMBIA

SEDE MEDELLÍN
FACULTAD DE MINAS

Área Curricular de Ingeniería Civil

Oferta de Posgrados

Especialización en Vías y Transportes
Especialización en Estructuras
Maestría en Ingeniería - Infraestructura y Sistemas
de Transporte
Maestría en Ingeniería – Geotecnia
Doctorado en Ingeniería - Ingeniería Civil

Mayor información:

E-mail: asisacic_med@unal.edu.co
Teléfono: (57-4) 425 5172

Monitoring and evaluation of energy use in oil treatment facilities

Laura Milena Corredor-Rojas ^a, Ariel Uribe-Rodríguez ^b, Juan Carlos Cobos-Gómez ^c

^aInstituto Colombiano del Petróleo, Ecopetrol S.A, Colombia. ^aLaura.corredor@ecopetrol.com.co, ^bAriel.uribe@ecopetrol.com.co,
^cJuan.cobos@ecopetrol.com.co

Received: March 11th, 2015. Received in revised form: January 23th, 2015. Accepted: May 25th, 2016.

Abstract

The paper presents the steps required to undertake an energy audit in oil treatment facilities and presents key energy performance indicators that can be applied to benchmarking and monitoring of this type of system. The proposed indicators are classified into the five following categories: production, consumption, efficiency, cost and reliability indicators. The paper also presents some general initiatives, short, medium or long term, that are classified according to their maturity and their economic benefits in order to increase energy efficiency.

Keywords: Production facilities, Key performance indicators, Energy performance.

Monitoreo y evaluación del uso de la energía en estaciones de tratamiento de crudo

Resumen

El artículo presenta los pasos a seguir para realizar una auditoría energética en estaciones de tratamiento de crudo y los indicadores claves de desempeño energético que pueden ser aplicados para el monitoreo y el desarrollo de estudios comparativos de éste tipo de sistemas. Los indicadores propuestos se agrupan en 5 categorías que son indicadores de producción, consumo, eficiencia, costo y confiabilidad. También se presentan algunas iniciativas generales orientadas a incrementar la eficiencia energética de las estaciones las cuales se clasifican de acuerdo a su plazo de amortización y a sus beneficios económicos en iniciativas de corto, mediano o largo plazo.

Palabras clave: Facilidades de producción, indicadores energéticos, gestión energética.

1. Introduction

Production Facilities include processes, equipment and required materials to collect, separate, handle, characterize and measure oil, gas and water from producing wells. Considering that energy use has a very important role in the process cost, the following general strategies should be taking in account to optimize energy consumption, starting from the design phase [1]:

- Evaluation and quantification of the thermodynamic constraints of treatment processes and the selection of the processes with minimal energy requirements.
- -Selection of the process units with operation conditions similar to the ones used in the reservoir, which allow the reduction of heat exchange processes.
- Avoid hot fluids transportation to minimize energy losses.

- Evaluation of the use of direct contact heat exchangers instead of indirect heat contact exchangers.
- Maximization of fuel efficiency through heat recovery by steam generation (if cogeneration is required).
- Maximization of heat integration between processes to minimize external heating or cooling.
- Selection of equipment such as heaters and pumps that are highly efficient.

When oil treatment facilities are in operation, energy optimization is performed by applying a methodology to systematically monitor the processed energy performance and identify technological alternatives for optimization. The general applied methodology steps for the evaluation of energy management are shown in Fig. 1.

How to cite: Corredor-Rojas, L.M., Uribe-Rodríguez, A. and Cobos-Gómez, J.C., Monitoring and evaluation of energy use in oil treatment facilities. DYNA 83(197), pp. 68-73, 2016.

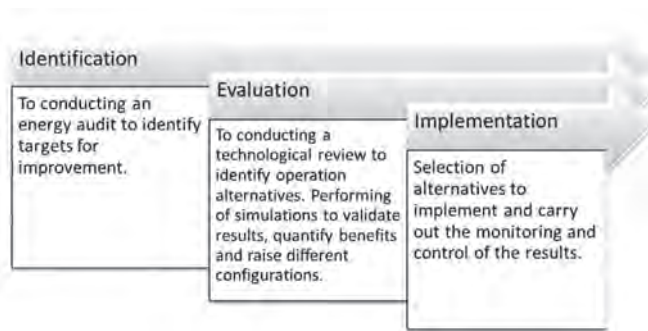


Figure 1. Methodology to evaluate energy management.
Source: Authors' presentation.

The energy audit is one of the most important steps in the methodology. The main objective of an audit is to study the current status of processes, cost analysis, and identification of areas, equipment or facilities that could be improved and the enumeration of possible measures that could be applied. In addition, the audit seeks to adapt the current energy consumption of the plant to nominal energy consumption, ensuring proper maintenance of the facilities, reducing nominal consumption with new technologies that increases the efficiency of energy usage and reduces demand of the processes, optimizing the operation of energy services. The results of the audits have shown increases in the efficiency of electricity consumption by 12% and savings in fuel consumption between 8 and 15 % per unit of product [2].

The state-of-the-art presents the general steps to conduct energy audits in industrial process. Despite this, neither the requirements of each step, in order to apply them in oil treatment facilities, nor key performance indicators to be used for analysis of energy use and to monitor the results of the proposed initiatives in the plan obtained from the energy audit are presented. For this reason, this article will focus on the steps that must be followed in order to ensure a proper energy audit, the requirements of each step, and the key performance indicators that may be applicable to oil treatment facilities.

2. Energy audit: Types and methodologies

The type of audit to be performed depends on several factors: on the industry and its principal function, on the depth of analysis to be performed, on the potential and the magnitude in cost reductions desired. There are two types of audits: preliminary and detailed [3].

2.1. Preliminary audit

This is the most economical type of audit and it allows preliminary energy savings to be identified. By means of a visual inspection, basic information is collected to identify opportunities for energy savings in operational and maintenance areas and it draws attention to whether any further analysis is required. To summarize, this type of audit is performed to [4]:

- Set the global energy consumption of the processes.
- Build the energy consumption baseline.

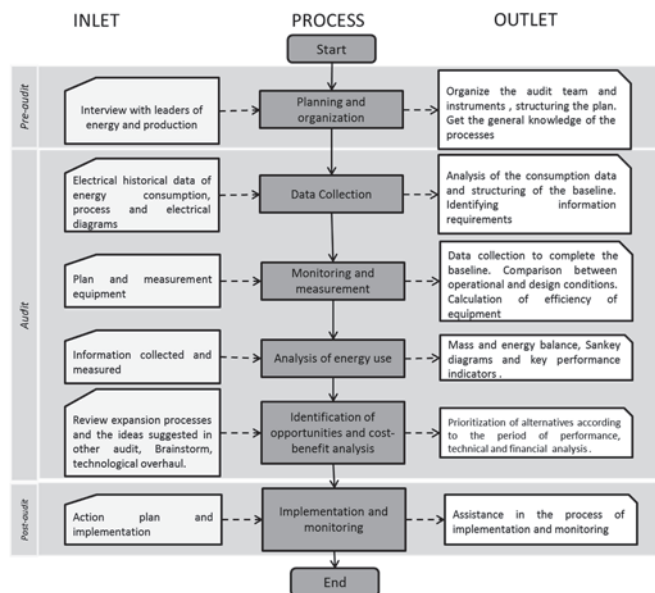


Figure 2. Stages of a detailed audit.
Source: Authors' presentation.

Table 1.
General information of oil treatment facilities. Source: Authors' presentation.

General information	Units
Design capacity	BOE
Current processing capacity	BOE
Streams input	
Inlet fluid (oil + water)	BOE
Water cut	%BSW
API gravity	°API
Diluent (if applicable)	%v
Streams output	
Crude	BOE
Water cut of the crude	% BSW
API gravity of the crude	°API
Gas (if applicable)	MMSCFD
Condensates	BOE
Specific gravity of gas (if applicable)	-

Source: Authors' presentation

- Identify energy losses that can be corrected through operative and maintenance actions.
- Estimate the savings
- Categorize areas depending on investments and savings.

2.2. Detailed audit

The detailed audit involves all the components of the preliminary audit, including tests and measurements to evaluate the energy use and its losses, as well as evaluating the economy for the proposed changes. It also incorporates the assessment of how much energy should be used in each process and operative function by using computer modeling simulations [5]. The detailed audit is divided into 3 stages which are: pre-audit, audit and post-audit. Fig. 2 presents the processes that are to be followed in each stage. Each process will be later explained for oil treatment facilities [6].

3. Information required for energy analysis of oil treatment facilities

Tables 1 and 2 present the required information to perform an energy audit in oil treatment facilities.

Table 2.

Information required for the energy analysis of oil treatment facilities.

Information	Source	Description
Basic information	Operation Manual	Process description and operating procedures.
	Block Flow Diagrams (BFD) and Process Flow Diagrams (PFD)	Conceptual approach to the processes' functions.
	Control and instrumentation diagrams (P&ID)	Graphic approach of measurement, monitoring and control systems.
	Industrial services diagrams (Water, steam)	Graphic approach of industrial services
	Line diagrams	Graphic approach to electrical installations.
	General list of equipment and data sheets of major equipment	Design data of equipment (diameter, height), Current operational status and optimum operating conditions. Design and operation capabilities. Maximum levels of fluids, pressure, temperature, other.
	Production	Total fluid (oil, water and gas)
	Properties and composition of the fluids	Fluid compositions, BSW, Calorific (gases and condensable) and properties of the diluent (if applicable).
	Chemical treatment	Chemical product (volume, concentration, injection, chemical cost per barrel of fluid).
	Historical consumption and cost (electricity and fuels)	Consumption and cost of electricity, gas, diesel and other fuels.
Statistical Data	Historical properties and composition of fuel	Characteristics and composition of the fuel gas: average molecular weight, specific gravity and net LHV.
	Power quality measurements of the main equipment	Current, voltage, power factor and frequency.
	Gas flare Tea	Gas flow, evaporated diluent (if applicable) and other fuels burned.
Reliability	Operation and performance indicators of equipment	Measurement of performance indicators
	Historical databases of failures	Equipment, date, type and cause of the failure and downtime
	Factory test equipment	Operating curves
	Periods of equipment	Lifetime of the process and equipment

General Data of equipment	operation	Programs and initiatives for management and consumption of Energy. Maintenance and reliability management, and operation and production management.
	Management systems and programs	Input and output flows of oil, water and gas (separated). Inlet and outlet temperatures. Flow of diluent (if applicable). Operation level fluid and retention times.
	Separators and tanks	Inflow and outflow, inlet and outlet temperatures, fuel flow, composition and characteristics of the fuel, stack temperature, % excess oxygen.
	Heaters and Coolers	Rotational speed, power and nominal efficiency, net positive suction, suction and discharge pressures, pump capacity, rated capacity, hydraulic power; voltage, current and power factor.
	Pumps	Process equipment, temperatures (input/output), pressures (input/output).
	Vapor Recovery Unit (VRU)	Regulations and environmental requirements. Delivery quality criteria for products output.
	Environmental management plan and other requirements	Elevation above sea level, minimum day and night atmospheric pressure, relative humidity, average temperatures (wet bulb and dry bulb).
	Climatic and geographical conditions	Exhaust gas analysis (temperature, flow and composition), composition (CO, CO ₂ , NO _x , SO _x , H ₂ O, etc.) and operational conditions at the time of gas analysis.
	Emissions	
	Environmental Requirements	

Source: Presentation of the authors.

4. Analysis of energy use

Energy use diagrams are presented in Fig. 3. These apply to a treatment facility without crude heating processes.

The analysis should include energy consumption per process, calculation of current and optimal yields of processes, the list of equipment with higher energy consumption and the key performance indicators' approaches (KPIs).

In this stage, KPIs play an important role because they are control tools that allow the planned goals to be monitored. The indicators should provide enough and relevant information to not bias the analysis, and in addition they should allow for decision-making and effective processes control. In order to do this, an indicator must at least comply with features that are internationally recognized by the acronym SMART (Specific, Measurable, Actionable, Relevance, Timely). The suggested steps to formulate or adjust the indicators are:

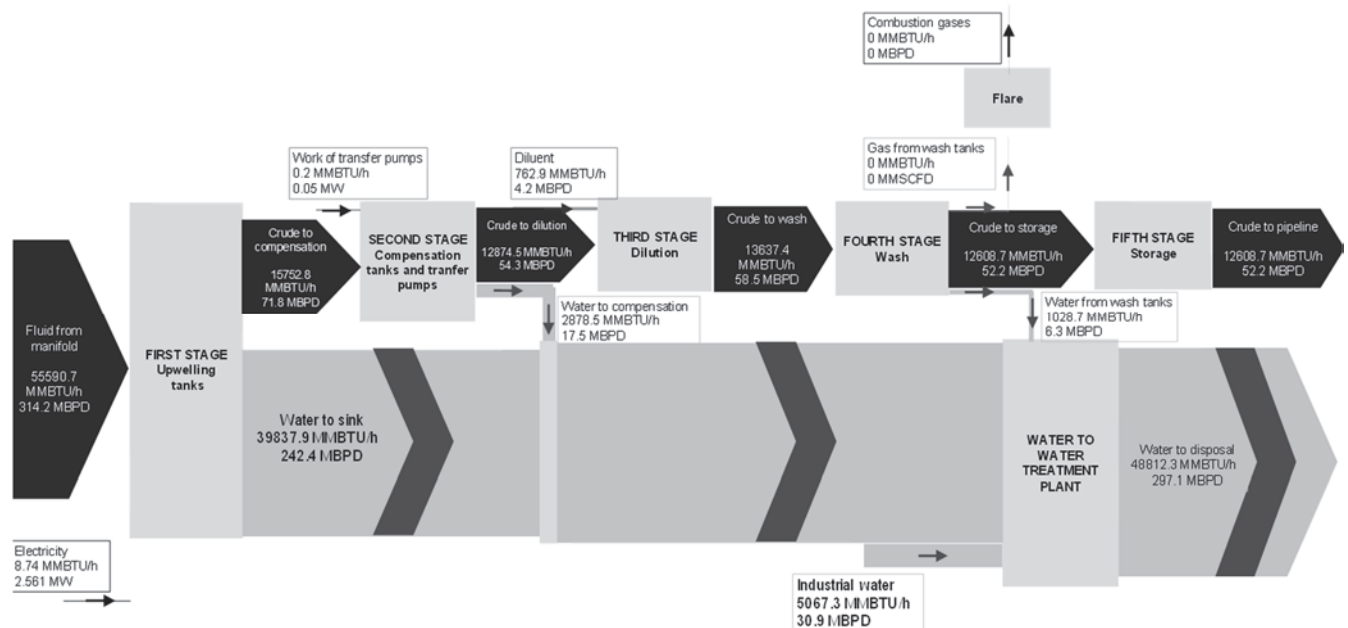


Figure 3. Sankey diagram for an oil treatment facility.
Source: Authors' presentation.

- Set the variables required for measurement.
- Identification of the benchmarks.
- Set the target to be measured.
- Identification over what information is required from the target.
- Identification of people for whom this information is useful.

There are two kinds of indicators: management and results. Management indicators measure how the process is performing in terms of resource optimization. This kind of indicator is used to detect what is happening and is used to take appropriate actions to improve the situation. Results indicators are associated with management goals as well as the efficiency and effectiveness of their achievement. This kind of indicator is used to measure situations when they have already occurred and to take appropriate measures to prevent reoccurrences. The proposed KPIs for oil treatment facilities are presented in Table 3.

The analysis of energy consumption can be performed based on both internal and external benchmarking performance studies. Internal studies include the analysis of historical data and trends, while external audits are undertaken based on data or indicators obtained from systems or similar processes. Comparative factors to be taken into account for external analysis are the maximum capacity of the station, the technologies used in the processes and the quality specifications for both raw and treated products. Benchmarking studies allow:

- Quantification of fixed and variable energy consumption according to production levels.
- Comparison of the usability factor with other facilities.
- Identification of gaps and best practices.
- Establishment of the basis for monitoring and goal setting.
- Quantification of the target variability margin for consumption and energy cost.

5. Identifying opportunities for optimization

The last step in the energy auditing process is to recommend an action plan to optimize the use of energy, the cost of operation and to reduce emissions. The plan should include the period of implementation for each initiative and the investments' payback period [7].

Initiatives for energy management can be classified as short, medium or long term, depending on their maturity and their economic benefits. Short-term projects are those that must be performed quickly to resolve a compliance issue regarding air quality, or that have a very short payback period (a high rate of return).

The projects that are to be carried out for environmental reasons are sometimes to be implemented in the future, but are still classified as short-term because the work should begin as soon as possible [8]. Initiatives with high rates of return should be classified according to their ease of implementation. If little or no capital is required, projects should be performed immediately. If projects require time and capital but have a high rate of return, they should be started as soon as possible.

The medium-term initiatives have periods of economic return that is usually between one to three years. Because they require capital investment, they should be retested to make sure they are economically feasible [8].

The long-term initiatives are characterized by simple economic returns over three years. If capital investment is large or the return is small, initiatives are classified as having a low priority. However, it is important not to dismiss such initiatives because they may have changes in energy costs, in technologies and environmental regulations: situations that make their implementation imperative.

Table 3.
KPIs for an oil treatment facility.

Category	Indicator	Description	Unit
Cost	Energy cost per product	Total cost of energy consumed by product yield	\$/BOE
	Energy cost per treated fluid volume	Total cost of energy consumed per fluid volume (water + oil treated)	\$/Bbl
	Relative energy cost	Ratio of energy cost and total cost of treatment	%
Consumption	Total energy consumed	Fuel consumption (burned, vented) and electricity	BTU, HP, J
	Energy intensity	Energy consumption per gross production	BTU/Bbl
Efficiency	Thermodynamic efficiency of equipment	Ratio of useful energy and the supplied energy to the equipment	%
Production	Usability Factor	Ratio of treatment capacity by the maximum capacity of the facility	%
Reliability	Failure index	Failure rate of the equipment or system	#failure/equipment/year
	Mean time between failures	Expected time between two failures for a repairable system	days
	Mean time to repair or replace	Average time needed to restore an asset to its full operational capabilities after a failure	days
	Carbon intensity	Emissions (venting, combustion)	t CO _{2eq} /BOE

Source: Authors' presentation

Short-Term

- Reducing excess air in tube heaters
- Maintenance of rotating equipment
- Change the flexible electrical demand to periods of low prices
- Installing a monitoring system for peak of demand electricity
- Installation of high efficiency motors
- Installing fuel and gas exhaust meters
- Review of price agreements for the purchase of fuel and electricity
- Monitor operating conditions and personnel training

Medium-term

- Installing waste heat recovery units
- Increase the load factors of the equipment (specially motors)
- Use variable speed driven pumps
- Installation of a control system to the operating conditions

Long-term

- Installation of electricity generation units
- Relocation processes or equipment to reduce transport costs.
- Energy integration in multi-unit installations.
- Power Factor Correction

Figure 4. Alternatives to optimize oil treatment facilities.
Source: Authors' presentation.

The reduction in energy costs associated with short-term initiatives is between 10 and 20%, while initiatives with a repayment period of two years or less offer additional reductions: between 20 and 30% [9].

Some initiatives to optimize treatment facilities are presented in Fig. 4. The size of the installation will determine whether an initiative is short, medium or long term [10-13].

6. Conclusions

An energy audit quantifies the consumption and energy costs of industrial processes, in order to identify gaps and establish a plan of initiatives to reduce losses, maximize efficiency, optimize energy supply and costs associated with processes. Energy auditing affects three key factors: profitability, through the optimization of energy consumption; productivity through optimization of equipment and processes; and performance, to use energy rationally.

Structuring an optimization plan is one of the most important steps in an energy audit. The initiatives proposed in this article are focused on improving models to contract the purchase of electricity and fuel, improve power quality, reduce energy loss from processes and equipment, include and track maintenance plans and improve operating conditions' control systems. Additionally, they can help to minimize emissions associated with the processes.

The sustainability of the audit's results will be ensured through monitoring and control of key performance indicators, accompanied by an energy management program. Additionally, these indicators will allow comparative studies with other treatment facilities that have similar characteristics (production capacity, used technologies in processes, crude and products quality).

References

- [1] Nardella, N.D. Improving energy efficiency in thermal oil recovery surface facilities [On line]. Montreal, 2010. [date of reference November 15th of 2014]. Available at: <http://www.worldenergy.org/documents/congresspapers/201.pdf>. World energy congress.
- [2] Negri, et al., Improving energy efficiency in upstream facilities. Conference paper. Offshore mediterranean conference and exhibition, march 23-25, 2011.
- [3] Tyers, B. and Franklin, N., Monitoreo y seguimiento del uso de la energía. Arpel, 2002.
- [4] Bureau of Energy Efficiency. General aspects of energy management and energy audit [On line]. [date of reference November 15th of 2014]. Available at: <http://www.bsr.org/reports/bsr-energy-management-handbook.pdf>. 2005.
- [5] Turner, W. and Doty, S., Energy management handbook. Atlanta: The Fairmont press, 2007.
- [6] Seai., Energy efficient design methodology [On line]. 2002. [date of reference November 16th of 2014]. Available at: www.seai.ie
- [7] Thumann, A. and Mehta, D., Handbook of energy engineering. The Fairmont press, 2001.
- [8] Tyers, B. and Franklin, N. Implementación de programas de gestión de energía en la industria del petróleo y el gas. Arpel, 1999.
- [9] Kreith, F. and Goswami, D., Handbook of energy efficiency and renewable energy. Taylor & Francis group, London, UK, 2007. DOI: 10.1201/9781420003482
- [10] New Jersey State Department of Environmental Protection. Establishing an energy management program and identifying energy

savings opportunities. The State University of New Jersey, New Jersey, USA, 2000.

- [11] Murray, D et al., Evaluación de proyectos relacionados con la energía. Arpel, 1999.
- [12] Thumann, A. and Woodroof, E., Handbook of financing energy projects. Atlanta: The Fairmont press, 2005.
- [13] Capehart, B., Turner, W. and Kennedy, W. Guide to energy management, third edition. The Fairmont press, Atlanta, USA, 2000.

L.M. Corredor-Rojas, gained her BSc in Chemical Engineering in 2004 and her MSc. degree in Chemical Engineering in 2008, all from the Industrial University of Santander, Colombia. From 2004 she worked in the Colombian Petroleum Institute (ICP) in the operational analysis of non-catalytic processes. Since 2009 she has worked at the same institute but in the implementation of alternatives to optimize the efficiency and cost of the processes associated with production and processing of crude oil.
ORCID: 0000-0001-7981-1460.

A. Uribe-Rodriguez, gained his BSc in Chemical Engineering in 2001 and his MSc. degree in Computer Science in 2004, both from the Universidad Industrial de Santander (UIS), Colombia. Since 2004, he has been working as a researcher at the Colombian Petroleum Institute (ICP) on the analysis of refining schemes, and energy planning and optimization of oil production processes. For the last three years, he focused on optimizing upstream costs and the possible inclusion of bioenergy as a renewable energy source for production processes.
ORCID: 0000-0001-5950-8923.

J.C. Cobos-Gómez, gained his BSc in Chemical Engineering in 1996 from the Universidad Industrial de Santander, Colombia. He was awarded the qualification of specialist manager in energy resources from the Autonomous University of Bucaramanga in 2006 and the qualification of specialist project manager (PFM) in 2010. He has worked on projects involving the optimization of industrial services, distillation, catalytic cracking of paraffins at a refinery in Barrancabermeja. He is currently leading the project "Emerging technologies to improve heavy oil crudes" at the Colombian Petroleum Institute (ICP).
ORCID: 0000-0002-0014-0754.



UNIVERSIDAD NACIONAL DE COLOMBIA

SEDE MEDELLÍN
FACULTAD DE MINAS

Área Curricular de Ingeniería
Química e Ingeniería de Petróleos

Oferta de Posgrados

Maestría en Ingeniería - Ingeniería Química
Maestría en Ingeniería - Ingeniería de Petróleos
Doctorado en Ingeniería - Sistemas Energéticos

Mayor información:

E-mail: qcaypet_med@unal.edu.co
Teléfono: (57-4) 425 5317

Differences between the use of ferric sulphate and ferric chloride on biodesulfurization of a large coal particle

Gerardo Andrés Caicedo-Pineda ^a & Marco Antonio Márquez-Godoy ^b

^a Facultad de Minas, Universidad Nacional de Colombia, Medellín, Colombia. gacaiced@unal.edu.co

^b Facultad de Minas, Universidad Nacional de Colombia, Medellín, Colombia. mmarquez@unal.edu.co

Received: April 29th, 2015. Received in revised form: November 05th, 2015. Accepted: February 20th, 2016.

Abstract

Three iron sources (FeSO_4 , $\text{Fe}_2(\text{SO}_4)_3$ and FeCl_3) at different concentrations (150, 700 and 1250 mg Fe/L) were evaluated on large coal particle biodesulfurization processes at Erlenmeyer level. A consortium of *Acidithiobacillus ferrooxidans* (ATCC 23270) and *Acidithiobacillus thiooxidans* (ATCC 15494) was used in all the assays. By using 150 mg Fe^{2+} /L (from FeSO_4), pyrite biooxidation was 31.14% after 12 days. When $\text{Fe}_2(\text{SO}_4)_3$ replaced FeSO_4 , oxidation improved by 21.16%. The assays using the highest concentrations of sulphate sources also obtained the same increase. However, $\text{Fe}_2(\text{SO}_4)_3$ assays had a better sulphate removal from coal. This suggests that using the smallest concentration of $\text{Fe}_2(\text{SO}_4)_3$ is a good alternative to boost the pyrite oxidation rate and avoid the formation of precipitates. Additionally, biooxidation in the FeCl_3 assays decreased, indicating, *a priori*, that the microorganisms were not able to adapt properly to Cl^- ions.

Keywords: coal biodesulfurization; pyrite oxidation; ferric chloride, ferric sulphate; *Acidithiobacillus ferrooxidans*; *Acidithiobacillus thiooxidans*.

Diferencias entre el uso de sulfato férrico y cloruro férrico en la biodesulfurización de un carbón con tamaño de partícula grueso

Resumen

Se evaluaron tres fuentes de hierro (FeSO_4 , $\text{Fe}_2(\text{SO}_4)_3$ y FeCl_3) a diferentes concentraciones (150, 700 y 1250 mg Fe/L) en procesos de biodesulfurización de un carbón con tamaño de partícula grueso, utilizando un consorcio de *Acidithiobacillus ferrooxidans* (ATCC 23270) and *Acidithiobacillus thiooxidans* (ATCC 15494). Al usar 150 mg Fe^{2+} /L (FeSO_4), luego de 12 días se obtuvo 31.14% de pirita oxidada. Al reemplazar FeSO_4 por $\text{Fe}_2(\text{SO}_4)_3$, la oxidación mejoró en un 21.16%. Aunque los ensayos con las mayores concentraciones de sulfatos obtuvieron un resultado similar, al utilizar $\text{Fe}_2(\text{SO}_4)_3$ permitió mejor remoción de sulfatos del carbón. Lo anterior sugiere que basta con utilizar la menor concentración de $\text{Fe}_2(\text{SO}_4)_3$ para mejorar el índice de oxidación de pirita y evitar formación de precipitados. Por otra parte, los ensayos con FeCl_3 tuvieron una disminución en la tasa de biooxidación, indicando *a priori*, que los microorganismos no fueron capaces de adaptarse adecuadamente a los iones Cl^- .

Palabras clave: biodesulfurización de carbón; oxidación de pirita; cloruro férrico; sulfato férrico; *Acidithiobacillus ferrooxidans*; *Acidithiobacillus thiooxidans*.

1. Introduction

Among the elements contained in coal, sulphur plays an important role in almost all systems that use the material, especially combustion processes. The generation of sulphuric-volatile compounds (SO_2 , SO_3) causes atmospheric pollution, contributing to acid rain generation [1]. In order to limit sulphur oxide emissions, some research proposes precombustion

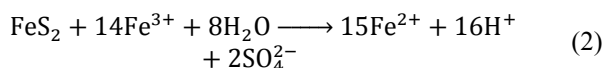
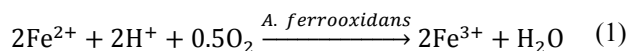
processes as good methods to reduce sulphur forms from coal. Sulphur on coal is present in two basic forms: organic, part of the coal structure, and inorganic, formed basically by sulphides (mainly pyrite) and sulphates [2].

High-sulphur coals ($\text{S} > 3\%$ w) generally present pyrite at a high rate (60-80% of total sulphur) [3,4]. For this reason, a lot research has focused on coal depyritization processes. Biological processes have many economic and

How to cite: Caicedo-Pineda, G.A. and Márquez-Godoy, M.A., Differences between the use of ferric sulphate and ferric chloride on biodesulfurization of a large coal particle. DYNA 83(197), pp. 74-80, 2016.

environmental advantages in comparison to chemical and physical options [5,6]. Nowadays, the research evaluates physical, chemical and biological parameters in order to seek alternatives with possible industrial applications [6].

Basically, biodepyritization consists of the oxidation of sulphides catalysed by acidophilic microorganisms, in an aqueous medium, generating soluble sulphates [1,7,8]. A typical culture medium contains Fe^{2+} from ferrous sulphate. Bacteria like *A. ferrooxidans*, oxidize Fe^{2+} producing Fe^{3+} by eq. (1). Then, Fe^{3+} attacks the pyrite, releasing more Fe^{2+} ions by eq. (2) which are taken by bacteria, creating a cycle of Fe^{3+} regeneration [8,9].



It is known that the use of fine-particle coal increases the exposed surface area of pyrite, improving its leaching. However, pyrite oxidation is not the only variable to consider in the design of the process. Precipitate generation is also important to define the best parameters for operation. Using fine particles increases the effect of coal compounds such as alkalinizing agents and carbonates. This induces the formation of undesirable compounds (i.e. jarosite) which could reprecipitate over the coal, reducing the efficiency of global desulfurization [4,6]. Recent studies showed low or null precipitate generation when large-particle coal is used, although pyrite oxidation decreased [10].

The use of large-particle coal also presents important advantages to the process. This minimizes risks to health, because volatile particles of fine coal could cause damage to the airways of humans and animals who come into contact with the material. Additionally, large-particle coal does not require heavy grinding and/or milling processes and it is easy to transport to regions far from the source. All of these are important factors when using a low-cost material. Besides, a wide range of cheap systems can treat coal at this size, from processes such as packed-bed bioreactors (heaps) [6] and even to stirred tank bioreactors. Previous research reports on a pilot plant level reactor using -1" Tyler mesh coal [11]. The low-specific gravity of the material (around 1.30) allowed it to be maintained in suspension without requiring a high agitation.

Multistage processes and/or long residence times would help to improve pyrite oxidation [5,12]. However, these also increase the costs of treatment. Previous research found that replacing ferrous sulphate by ferric salts also helps to accelerate pyrite oxidation, because it minimizes the *lag* phase, where the bacteria begin to oxidize ferrous ions [8,13]. This might be an alternative to boost the pyrite oxidation rate in processes using large-particle coal. Therefore, the research presented here evaluated how ferric ions from two different iron sources (sulphate and chloride) interacted with desulfurization processes of pyrite-rich coal mediated by a consortium of acidophilic bacteria. Chemical and mineralogical techniques allowed the behaviour of the process to be evaluated.

Table 1.

Proximate analyses, sulphur forms and iron composition in the coal sample.

Proximate analyses ^a	Value
Humidity (%)	4.6
Ash (%)	27.0
Volatile substance (%)	34.7
Fixed carbon (%)	33.7
Calorific value (Cal/g)	5106
Sulphur forms (%)	
Pyritic	4.00
Sulphate	0.79
Organic	1.48
Iron composition (%)	
Pyritic	3.50
Non-pyritic	0.65

^a All values calculated on gross basis.

Source: The authors.

2. Experimental Procedure

2.1. Coal

High-volatile bituminous coal samples were collected from "La Angostura" mine (Morales, Cauca, Colombia). The sample was ground to a particle size between -8 +30 Tyler mesh (2.38 mm < particle size < 0.50 mm). Table 1 shows proximate analyses, sulphur forms and iron composition in coal samples.

2.2. Microorganism

A consortium of *A. thiooxidans* (ATCC 15494) and *A. ferrooxidans* (ATCC 23270) were selected from the collection of Laboratorio de Biomineralogía y Biohidrometalurgia of the Universidad Nacional de Colombia - Sede Medellín. The microorganisms were previously adapted to coal biodesulfurization, according to an established protocol [3]. Inoculums preparation was carried out in 350 mL flasks, with a working volume of 150 mL, containing 1g of coal per 10 mL of solution. The solution was composed of 10% v of inoculum (5×10^8 cells/mL), 150 mg Fe^{2+} /L (from $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$) and modified T&K medium [14], defined as: 0.50 g/L of $(\text{NH}_4)_2\text{SO}_4$, 0.50 g/L of $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 0.50 g/L of KH_2PO_4 and 1.5 mL/L of H_2SO_4 98% . The microorganism was adapted to ferric sulphate (92% of total iron as Fe^{3+} ions) and ferric chloride, replacing ferrous sulphate. The cultures were incubated in a shaker for 12 days, at 30 ± 1 °C, using a mixing rate of 180 ± 2 rpm. The incubation was repeated several times and no variations of the biodesulfurization ratio among the three last processes were observed.

2.3. Coal biodesulfurization processes

The assays were prepared in 350 mL flasks, with a working volume of 150 mL, using a coal/culture medium ratio of 1:10 (g of coal: mL solution). The variables evaluated were: (i) proportion of iron added (mg/L): 150 (C1), 700 (C2) and 1250 (C3) and (ii) iron source: ferrous sulphate (SF(II)), ferric sulphate – with 8% Fe^{2+} ions – (SF(III)) and ferric chloride (CF(III)). All processes were

incubated for 12 days, under similar conditions to the inoculum preparation and with the respective replica and abiotic control.

All the experiments were monitored with measures of pH and redox potential (ORP), using a pH/ORP-meter SCHOTT Handylab. Total and ferrous iron in solution was determined in a spectrophotometer Thermo GENESYS UV 10, employing the method E 394-09, according to ASTM standard. At the end of the experiments, sulphur forms in coal samples were measured by the ASTM D 2492-02 method.

Additionally, the mineralogical composition of the original and treated samples was established by XRD (X-ray diffraction). XRD analysis were made to -200 Tyler mesh coal in a Rigaku Miniflex II X-ray Diffractometer, using a step by step method, with 5° start angle, 70° stop angle, 0.01°/step angle, and counting time of 1 s. The minerals present on the coal were quantified by XRD using a Rietveld refinement. All simulations and calculations were carried out in the program X'Pert HighScore Plus®, using the data base PDF2.

3. Results

3.1. Biodesulfurization process

Fig. 1 shows pH behaviour along the process. The initial value was the highest in the ferrous sulphate assays. In contrast, ferric chloride assays CF(III) C2 and CF(III) C3 started at the lowest pH. In the first days of the process, ferrous sulphate assays increased in value, where SF(II) C2 and SF(II) C3 assays reached the maximum (1.71 at day 3). Subsequently, SF(II) C1 assay did not have significant changes; however, SF(II) C2 and SF(II) C3 assays had a continuous pH decreasing up to day 12, especially by using 1250 mg Fe²⁺/L (1.51).

The assays using ferric salts also raised the pH value on the first day, although it was not as pronounced as the values observed in the ferrous sulphate assays. Later, all the assays had a pH decreasing, which was dependant on the iron source and concentration. CF(III) C3 assay showed the lowest value (1.37).

As regards the abiotic controls, they presented a pH increase on the first day (Fig. 1b) and then did not evidence significant changes during the process, except for SF(III) C3 and CF(III) C3 abiotic controls, whose values decreased after day 1, although they did not reach the same values as their corresponding bioassays.

Fig. 2 presents redox potential (Eh). In the same way as the pH values, initial Eh depended on iron source and the initial concentration. Ferrous sulphate assays had the lowest values (below 450 mV). In contrast, the values were over 530 mV in the assays using ferric salts, especially by using ferric chloride. Nonetheless, all the assays had a similar behaviour after day 1. Eh of the assays tended to stabilize between 550-560 mV after day 6 (day 9 for CF(III) C1 assay), except for CF(III) C2 and CF(III) C3 assays, where the value did not surpass 500 mV. On the other hand, all the abiotic controls tended to reach 413 mV after day three.

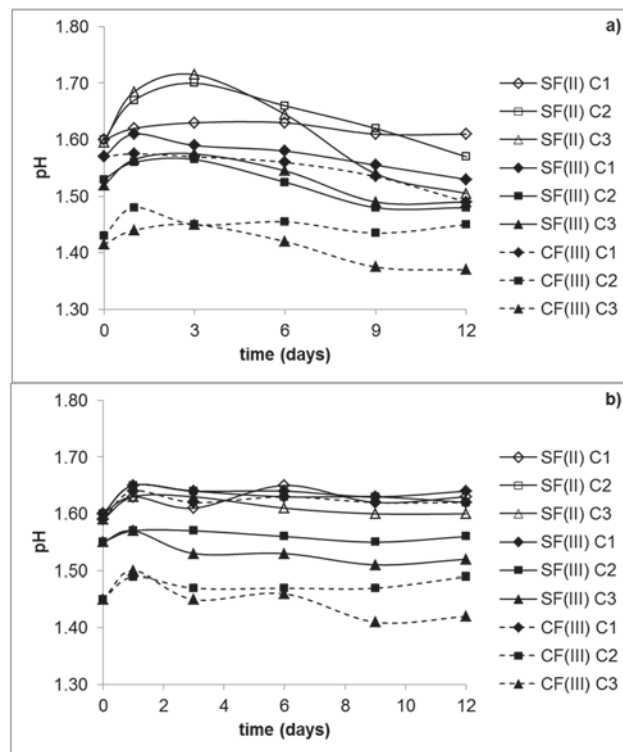


Figure 1. pH behaviour vs. time, a) Coal biodesulfurization assays to different iron salts and b) Abiotic controls. SF(II): ferrous sulphate, SF(III): ferric sulphate, CF(III): ferric chloride. Iron added were C1: 150 mg/L, C2: 700 mg/L, C3: 1250 mg/L.

Source: The authors.

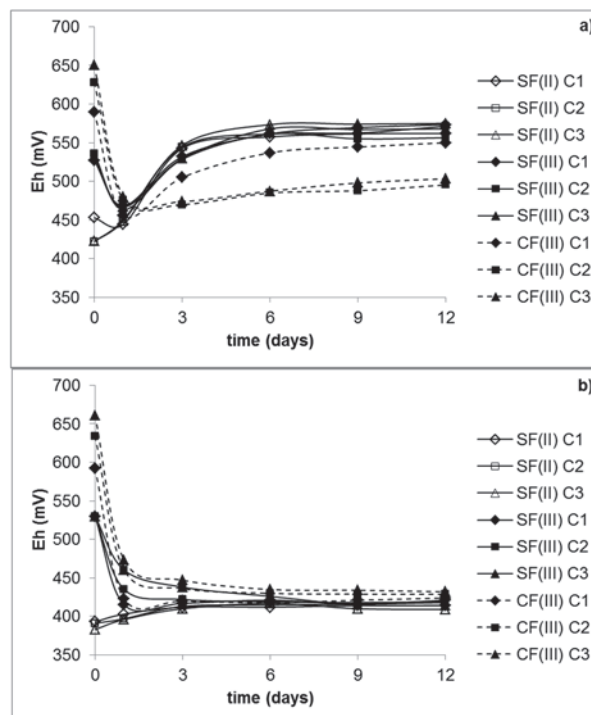


Figure 2. Redox potential behaviour vs. time, a) Coal biodesulfurization assays to different iron salts and b) Abiotic controls. SF(II): ferrous sulphate, SF(III): ferric sulphate, CF(III): ferric chloride. Iron added were C1: 150 mg/L, C2: 700 mg/L, C3: 1250 mg/L.

Source: The authors.

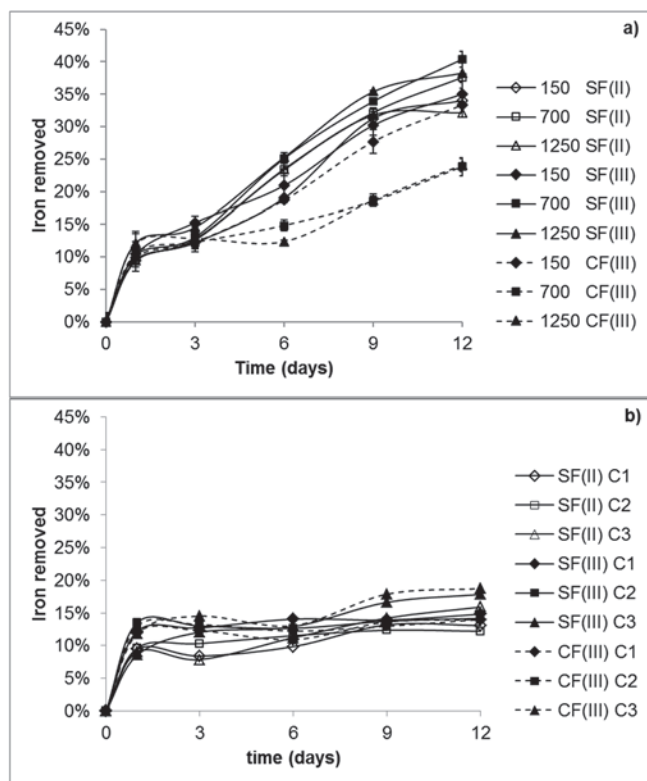


Figure 3. Iron removed behaviour vs. time, a) Coal biodesulfurization assays to different iron salts and b) Abiotic controls. SF(II): ferrous sulphate, SF(III): ferric sulphate, CF(III): ferric chloride. Iron added were C1: 150 mg/L, C2: 700 mg/L, C3: 1250 mg/L.

Source: The authors.

All assays and abiotic controls showed iron dissolution on day 1 (Fig. 3). The quantity removed was around 12% in SF(III) C3, CF(III) C3 assays and abiotic controls and 10% in the other assays. After day 2, the abiotic controls of SF(III) C3 and CF(III) C3 assays reached a maximum removal of around 19%, while the abiotic controls of the other assays were between 12-14% (Fig. 3b). Moreover, SF(II) C1, SF(III) C1 and CF(III) C1 assays reached a removal between 33-35% at the end of the process (Fig. 3a) and SF(III) C2 and SF(III) C3 assays a maximum around 40%. In contrast, CF(III) C2 and CF(III) C3 assays had a removal below 25%. This correlated with the low redox potential observed in Fig. 2.

Fig. 4 presents pyrite oxidized at the end of the process. All the assays using ferric sulphate and SF(II) C2 and SF(II) C3 assays reached the maximum of pyrite oxidation (37.80%). In contrast, CF(III) C2 and CF(III) C3 assays obtained the minimum values (18.00% and 14.00%), showing a decreased pattern when the iron concentration increased (similar to observed in the Fig. 2, 3). SF(III) C3 and CF(III) C3 abiotic controls reached a chemical pyrite oxidation of around 9.00%, while the others did not have a significant value (below 3%).

Although the assays using the two highest iron concentrations of ferrous and ferric sulphate reached the maximum of pyrite oxidized, SF(III) C2 and SF(III) C3 assays had the best sulphate removal of around 70.00% (Fig. 5). On the other hand, sulphate removed was low in CF(III) C2 and CF(III) C3 assays, similar to that observed in the parameters shown in Fig. 2, 3.

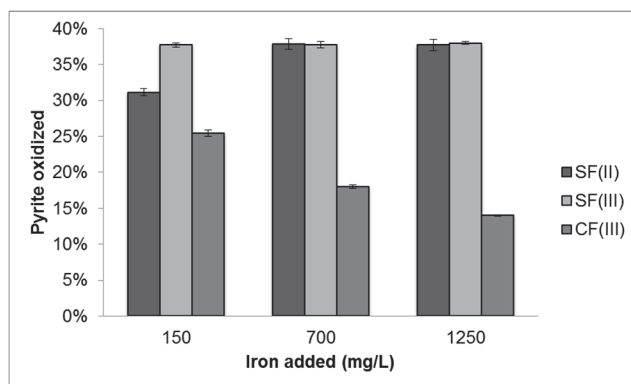


Figure 4. Percentage of pyrite oxidized after coal biodesulfurization processes. SF(II): ferrous sulphate, SF(III): ferric sulphate, CF(III): ferric chloride.

Source: The authors.

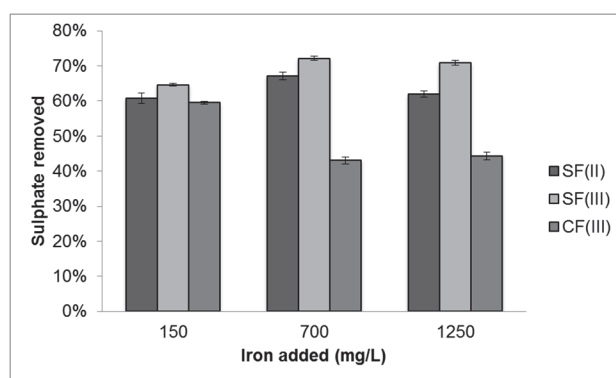


Figure 5. Percentage of sulphate removed after coal biodesulfurization processes. SF(II): ferrous sulphate, SF(III): ferric sulphate, CF(III): ferric chloride.

Source: The authors.

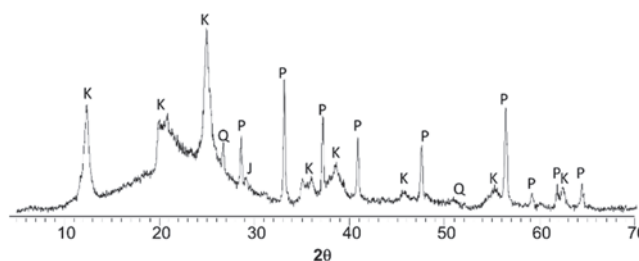


Figure 6. X-ray diffraction pattern of the coal sample before the biodesulfurization process. K: kaolinite, Q: quartz, P: pyrite and J: jarosite.

Source: The authors.

3.2. X-ray diffractogram analyses

The X-ray diffractogram of non-treated coal (Fig. 6) showed the presence of pyrite, kaolinite, quartz and jarosite. The curvature of baseline between $2\theta=13^{\circ}$ - 23° corresponds to the amorphous phase of coal [15].

The Rietveld refinement for the semi-quantification of XRD patterns, showed 7.5% w of pyrite, 28.86% w of kaolinite, 1.69% w of quartz and not significant amounts (below 0.5%) of jarosite. A goodness-of-fit below 1.1% and

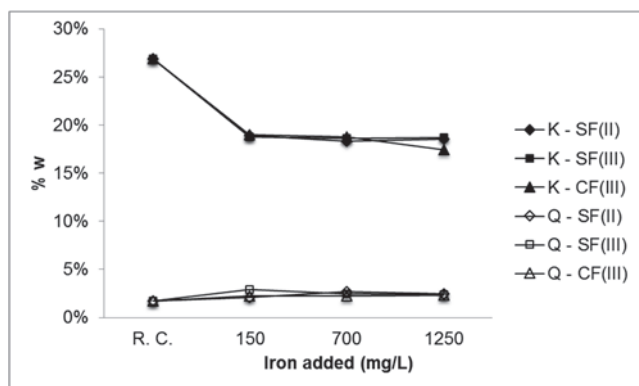


Figure 7. Silicates proportions in coal samples. SF(II): ferrous sulphate, SF(III): ferric sulphate, CF(III): ferric chloride. K: Kaolinite, Q: Quartz. R.C.: raw coal.

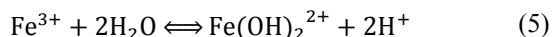
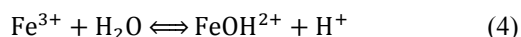
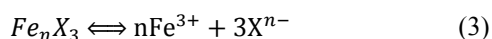
Source: The authors.

weighted residual profile below 20% were observed, considered good to accept this method [16,17].

Treated coal samples presented a decrease of kaolinite content (Fig. 7), reporting an average percentage of 18.57% ($\pm 0.45\%$) between all the assays without a specific correlation with respect to the evaluated variables. Quartz did not show significant changes. As additional data, jarosite content was not significant (less than 0.4% w).

4. Discussion

Ferric salts influenced pH at the beginning of the process. They ionise to Fe^{3+} and the respective anion (X^n) in water solution by eq. (3). Then, Fe^{3+} hydrolyses by eq. (4)-(6) and releases H^+ to the medium [18]. This explains why the assays using ferric salts started at lower values than assays using ferric sulphate. Besides, the equilibrium reaction of Cl^- and H^+ generates a stronger acid than SO_4^{2-} and H^+ , influencing low pH.



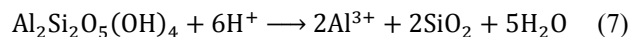
Additionally, the initial redox potential values depended on the concentration of Fe^{3+} ions, as this parameter is directly related to the $\text{Fe}^{3+}/\text{Fe}^{2+}$ ion exchange ratio [4,19–21]. Ferric sulphate assays had a lower Eh than ferric chloride assays, because the reagent contained 8% of Fe^{2+} ions.

The following sections provide information on how both initial pH and Eh variations affect assays and abiotic controls during the processes for each iron source.

4.1. Sulphate salts effect

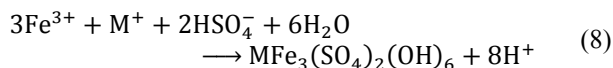
The increment of pH on day 1 (Fig. 1) indicated H^+ consumption by the *A. ferrooxidans* to biooxidize Fe^{2+} , as

shown in eq. (1). The results obtained agreed with the reaction mechanism, because the assays using the highest concentrations of Fe^{2+} consumed more H^+ , which explains why the highest pH values were similar to those reported by other authors [22]. Moreover, Fig. 7 indicates that kaolinite also might have consumed acid in its dissolution by eq. (7) [23,24].



The subsequent pH decrease indicated that the acid production came from pyrite oxidation by eq. (2) and Fe^{3+} hydroxylation by eq. (4)-(6) [18], which produced more H^+ after the lag phase. Therefore, assays using the two highest concentrations have a better pH decreasing, because they had a higher iron concentration in the culture medium. Since the coal did not present common alkalizing compounds such as carbonates [3,4,10], it did not have a significant participation on pH changes around the process.

The replacement of ferrous sulphate by ferric sulphate allowed the onset of sulphide oxidation to be faster by eq. (2). The extra acidity helped to avoid pH increase and improved iron and sulphate removal in relation to assays using ferrous sulphate (Fig. 3, 5), proving pH value affects solubility. Additionally, iron concentration also affected the process. Early research showed that values above 1200 mg/L induce the generation of precipitates as hydroxysulphates by eq. (8), compounds poorly insoluble at the conditions used [4,18,25,11]. This explains why SF(II) C3 and SF(III) C3 assays decreased iron removal after day 9 (Fig. 3).



Moreover, the similar pyrite oxidation reached by SF(II) C2, SF(II) C3, SF(III) C2 and SF(III) C3 (around 38%) indicate that Fe^{2+} concentrations above 800 mg/L produced sufficient Fe^{3+} to obtain a good $\text{Fe}^{3+}/\text{Fe}^{2+}$ ion exchange, regardless of the iron sulphate source. The only advantage of using high ferric sulphate concentrations was to reduce precipitates formation, because it did not affect redox potential behaviour after day 3 (Fig. 2). Ferric sulphate abiotic controls demonstrated that Fe^{3+} added oxidized pyrite on the first days, especially by using the highest concentration. However, Fig. 2b, 3b showed Fe^{3+} ions did not regenerate again due to the absence of the microorganism (stabilization of Eh to 413 mV and non-significant iron removal after day 3). This suggests that bacterial activity is the main controller of $\text{Fe}^{3+}/\text{Fe}^{2+}$ ratio.

Nevertheless, the assay SF(III) C1 also obtained the best pyrite oxidation, 20% higher than the assay SF(II) C1. It indicated that the biodepyritization process (under the evaluated conditions) might obtain an increase of pyrite oxidized and a sulphate removal (64.51%) by adding 150 mg Fe^{3+}/L .

4.2. Chloride salts effect

Cl^- ions had a negative impact on bacterial activity. These findings are similar to previous research which reported that

concentrations above 1000 mg/L of Cl^- inhibited bacterial growth, reducing the regeneration of Fe^{3+} ions, and therefore pyrite biooxidation rate [26,27]. 800 mg/L and 1350 mg/L of Fe^{3+} provided 1520 mg/L and 2565 mg/L of Cl^- respectively. Although this work evidenced microorganism activity, results showed the adaptation stages were not sufficient to counterattack all chloride toxicity for these concentrations.

Furthermore, although ferric chloride assays had the lowest pH values (Fig. 1), the extra acidification did not help to increase sulphate removed (actually, it diminished), contrasting with the pH-sulphate solubility relation observed in ferric sulphate assays. The commercial use of ferric chloride as a coagulant in water treatment, especially to remove phosphate [28], is well known. Ferric chloride probably attacked KH_2PO_4 from the culture medium and K^+ induced to the formation of precipitates as hydroxisulphates, reacting with the iron and sulphate product of oxidation of pyrite. The possible precipitation of ferric ions and some nutrients in the bacteria (K, P) might have affected sulphate-removed ratio, as observed in Fig. 5.

5. Conclusions

The main contribution to replace ferrous sulphate by ferric sulphate was to support acidity to the media, which helps to improve sulphate removal. Nonetheless, the present work found that replacing the common iron source of the culture medium by ferric sulphate in small quantities (150 mg/L of Fe^{3+}) was a good alternative to increase the pyrite oxidation rate within the same configuration process. However, this efficiency has to be tested in other scenarios, for example using high pulp concentration, because initial Fe^{3+} ions (from ferric sulphate) only acted as an accelerant and not as a catalyser.

Furthermore, ferric chloride was not efficient as an iron source of Fe^{3+} ions. Although Fe^{3+} promoted chemical oxidation of pyrite during the first days, Cl^- was toxic to the microorganism and furthermore acted as an inductor of the sulphate precipitates. Both factors considerably affected pyrite biooxidation and sulphate removal ratios.

Moreover, besides pyrite and sulphate removal, all the assays presented kaolinite dissolution by acid leachate. This improved raw coal, because as inorganic matter decreases, calorific capacity increases, obtaining a better quality product. However, this phenomenon only depended on the acidity of the medium and not on the variables assessed. However, although ferric chloride did not work, future studies might evaluate other kinds of ferric salts, in order to determine a positive influence of the anions on pyrite oxidation and sulphate removal ratios.

Acknowledges

To the Biomineralogy and Biohydrometallurgy Laboratory (National University of Colombia) where the investigation was carried out.

References

- [1] Juszczak, A., Domka, F., Kozłowski, M. and Wachowska, H., Microbial desulfurization of coal with *Thiobacillus ferrooxidans* bacteria, *Fuel*, 74, pp. 725-728, 1995. DOI: 10.1016/0016-2361(94)00019-N
- [2] Calkins, W.H., The chemical forms of sulfur in coal: A review, *Fuel*, 73, pp. 475-84, 1994. DOI: 10.1016/0016-2361(94)90028-0
- [3] Caicedo, G.A. and Márquez, M.A., Selection procedures of consortia bacterial with *A. ferrooxidans* like bacteria and *A. thiooxidans* like bacteria in coal biodesulfurization process. *Revista Facultad de Ingeniería Universidad de Antioquia* [online], 52, 2010. [date of reference April 28th of 2015]. Available at: <http://www.scielo.org.co/pdf/rfiua/n52/n52a08.pdf>
- [4] Cardona, I.C. and Márquez, M.A., Biodesulfurization of two Colombian coals with native microorganisms, *Fuel Process Technology*, 90, pp. 1099-1106, 2009. DOI: 10.1016/j.fuproc.2009.04.022
- [5] Klein, J., Technological and economic aspects of coal biodesulfurization. *Biodegradation* [online], 9, 1998. [date of reference April 28th of 2015]. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10022072>
- [6] Cara, J., Vargas, M., Morán, A., Gómez, E., Martínez, O. and García Frutos, F.J., Biodesulphurization of a coal by packed-column leaching. Simultaneous thermogravimetric and mass spectrometric analyses, *Fuel*, 85, pp. 1756-1762, 2006. DOI: 10.1016/j.fuel.2006.02.014
- [7] Kodali, B., Rao, M.B., Narasu, M.L. and Pogaku, R., Effect of biochemical reactions in enhancement of rate of leaching, *Chemical Engineering Science*, 59, pp. 5069-5073, 2004. DOI: 10.1016/j.ces.2004.09.004
- [8] Sand, W., Gehrke, T., Jozsa, P.G. and Schippers, A., (Bio) chemistry of bacterial leaching—direct vs. indirect bioleaching, *Hydrometallurgy*, 59, pp. 159-75, 2001. DOI: 10.1016/S0304-386X(00)00180-8
- [9] Sand, W. and Gehrke, T., Extracellular polymeric substances mediate bioleaching/biocorrosion via interfacial processes involving iron(III) ions and acidophilic bacteria, *Research in Microbiology*, 157, pp. 49-56, 2006. DOI: 10.1016/j.resmic.2005.07.012
- [10] Caicedo, G.A., Márquez, M.A. and Moreno, C.X., Particle size, iron concentration and pH influence in biodesulfurization process of coal – laboratory tests. *Revista Colombiana de Biotecnología* [online], 13, 2011. [date of reference April 28th of 2015]. Available at: <http://www.scielo.org.co/pdf/biote/v13n2/v13n2a20.pdf>
- [11] Caicedo, G.A., Prada, M., Peláez, H., Moreno, C.X. and Márquez, M.A., Evaluation of a coal biodesulfurization process (semi-continuous mode) on the pilot plant level. *DYNA* [online], 79(174), 2012. [date of reference April 28th of 2015]. Available at: <http://www.scielo.org.co/pdf/dyna/v79n174/v79n174a15.pdf>
- [12] Loi, G., Mura, A., Trois, P. and Rossi, G., Bioreactor performance versus solids concentration in coal biodepyritization, *Fuel Processing Technology*, 40, pp. 251-260, 1994. DOI: 10.1016/0378-3820(94)90147-3
- [13] Nemati, M., Harrison, S.T.L., Hansford, G.S. and Webb, C., Biological oxidation of ferrous sulphate by *Thiobacillus ferrooxidans*: A review on the kinetic aspects, *Biochemical Engineering Journal*, 1, pp. 171-190, 1998. DOI: 10.1016/S1369-703X(98)00006-0
- [14] Tuovinen, O.H. and Kelly D.P., Studies on growth of *thiobacillus ferrooxidans*. I. Use of membrane filters and ferrous iron agar to determine viable numbers, and comparison with $^{14}\text{CO}_2$ -fixation and iron oxidation as measures of growth, *Archiv für Mikrobiologie*, 88, pp. 285-298, 1973. DOI: 10.1007/BF00409941
- [15] Lu, L., Sahajwalla, V., Kong, C. and Harris D., Quantitative X-ray diffraction analysis and its application to various coals, *Carbon*, 39, pp. 1821-1833, 2001. DOI: 10.1016/S0008-6223(00)00318-3
- [16] McCusker, L.B., Von Dreele, R.B., Cox, D.E., Loue, D. and Scardi, P., Rietveld refinement guidelines, *Journal of Applied Crystallography*, 32, pp. 36-50, 1999. DOI: 10.1107/S0021889898009856
- [17] Kniess, C.T., Cardoso de Lima, J. and Prates, P.B., The quantification of crystalline phases in materials: Applications of rietveld method. in Shatkhya, V., Ed. *Sintering - Methods and Products* [online], Intech, 2012, pp. 293-316. [date of reference April 28th of 2015]. Available at: <http://goo.gl/NWols4>

- [18] Daoud, J. and Karamanev, D., Formation of jarosite during Fe^{2+} oxidation by *Acidithiobacillus ferrooxidans*, *Minerals Engineering*, 19, pp. 960-967, 2006. DOI: 10.1016/j.mineng.2005.10.024
- [19] Bhatnagar, A., Dastidar, M.G. and Roychoudhury, P.K., Integrated technique for pyritic sulphur reduction in Indian (ASSAM) coal, *Energy Conversion Management*, 39, pp. 1121-1126, 1998. DOI: 10.1016/S0196-8904(98)00005-3
- [20] Malik, A., Dastidar, M.G. and Roychoudhury, P., Biodesulphurization of coal: Effect of pulse feeding and leachate recycle, *Enzyme Microbial Technology*, 28, pp. 49-56, 2001. DOI: 10.1016/S0141-0229(00)00283-0
- [21] Meruane, G., Oxidación bacteriana de sulfato ferroso, PhD dissertation Thesis, Universidad de Chile, Santiago, Chile, 2002.
- [22] Acharya, C., Kar, R.N. and Sukla, L.B., Bacterial removal of sulphur from three different coals. *Fuel* [online], 80, 2001. [date of reference April 28th of 2015]. Available at: <http://goo.gl/QmA9TG>
- [23] Ford, K.J.R., Leaching of fine and pelletised Natal kaolin using sulphuric acid. *Hydrometallurgy* [online], 29, 1992. [date of reference April 28th of 2015]. Available at: <http://www.sciencedirect.com/science/article/pii/0304386X9290008N>
- [24] Panda, A.K., Mishra, B.G., Mishra, D.K. and Singh, R.K., Effect of sulphuric acid treatment on the physico-chemical characteristics of kaolin clay, *Colloids Surfaces A Physicochemical Engineering Aspects*, 363, pp. 98-104, 2010. DOI: 10.1016/j.colsurfa.2010.04.022
- [25] Peláez, H., Prada, M., Caicedo, G.A., Moreno, C.X. y Márquez, M.A., Influencia de la relación inicial de $\text{Fe}^{3+}/\text{Fe}^{2+}$, en un proceso de biodesulfurización de carbones en suspensión. *Revista Internacional de Contaminación Ambiental* [online], 29, 2013. [date of reference April 28th of 2015]. Available at: <http://www.journals.unam.mx/index.php/rica/article/view/25160>
- [26] Deveci, H., Jordan, M.A., Powell, N. and Alp, I., Effect of salinity and acidity on bioleaching activity of mesophilic and extremely thermophilic bacteria, *Transactions of Nonferrous Metals Society of China*, 18, pp. 714-721, 2007. DOI: 10.1016/S1003-6326(08)60123-5
- [27] Gahan, C.S., Sundkvist, J.E. and Sandström A., A study on the toxic effects of chloride on the biooxidation efficiency of pyrite, *Journal of hazardous materials*, 172, pp. 1273-1281, 2009. DOI: 10.1016/j.jhazmat.2009.07.133
- [28] Caravelli, A.H., de Gregorio, C. and Zaritzky, N.E., Effect of operating conditions on the chemical phosphorus removal using ferric chloride by evaluating orthophosphate precipitation and sedimentation of formed precipitates in batch and continuous systems, *Chemical Engineering Journal*, 209, pp. 469-477, 2012. DOI: 10.1016/j.cej.2012.08.039

G.A. Caicedo-Pineda, received his BSc. Eng. in Chemical Engineering in 2005, his MSc. degree in Biotechnology in 2008, and his PhD. degree in Engineering - Science and Technology of the Materials in 2015, all of them from the Universidad Nacional de Colombia, Manizales (Bs. Eng.) and Medellín (MSc. and PhD. degrees). From 2006 to 2010, he worked as a researcher in the Universidad Nacional de Colombia, Medellín. From 2011 to the present, he has worked as a Process Engineering Researcher in Biofenix S.A.S. and Inverdesarrollo S.A.S. His research interests include biomining, bioleaching, nanobiotechnology, bioremediation, biological wastewater treatment, green technology and raw materials bioimprovement. ORCID: 0000-0003-2320-0980

M.A. Márquez-Godoy, received his BSc. Eng. in Geological Engineering in 1989, his MSc. degree in Geology in 1995, and his PhD. degree in Geology in 1999. The BSc. Eng. degree was from the Universidad Nacional de Colombia, Medellín and both the MSc. and PhD. degrees from Fundação Universidade de Brasília, Brazil. From 1990 to 1995, he worked as a geologist administrator in Ministerio de Minas y Energía. From 1996 to 1997, he worked as a geologist in Ingeominas. Currently, he works as Full Professor in the Universidad Nacional de Colombia, Medellín. Additionally, he has worked as a professor in EAFIT (1984 to 1986) and the Universidad de Antioquia (2000-2001). His research interests include engineering, microbial biotechnology, mining engineering, biogeochemistry, geomicrobiology, mineral processing, ore deposits, and minerals. ORCID: 0000-0002-7462-2430



UNIVERSIDAD NACIONAL DE COLOMBIA

SEDE MEDELLÍN
FACULTAD DE MINAS

Área Curricular de Ingeniería
Geológica e Ingeniería de Minas y Metalurgia

Oferta de Posgrados

Especialización en Materiales y Procesos
Maestría en Ingeniería - Materiales y Procesos
Maestría en Ingeniería - Recursos Minerales
Doctorado en Ingeniería - Ciencia y Tecnología de
Materiales

Mayor información:

E-mail: acgeomin_med@unal.edu.co
Teléfono: (57-4) 425 53 68

Human error and response to alarms in process safety

Beata Mrugalska ^a, Salman Nazir ^b, Edwin Tytyk ^c & Kjell Ivar Øvergård ^d

^a Faculty of Engineering Management, Poznan University of Technology, Poznan, Poland, beata.mrugalska@put.poznan.pl

^b Training and Assessment Research Group (TARG), Department of Maritime Technology and Innovation, Buskerud and Vestfold University College, Borre, Norway, salman.nazir@hbv.no

^c Faculty of Engineering Management, Poznan University of Technology, Poznan, Poland, edwin.tytyk@put.poznan.pl

^d Training and Assessment Research Group (TARG), Department of Maritime Technology and Innovation, Buskerud and Vestfold University College, Borre, Norway, Kjell.Oevergaard@hbv.no

Received: November 30th, 2015. Received in revised form: February 25th, 2016. Accepted: April 25th, 2016.

Abstract

Human operators are required to respond to alarms in normal conditions, and also to find solutions to unexpected situations in real time. The aim of this study is to increase the knowledge of human responses to alarms in the context of occupational situations. It describes how humans contribute to accidents, and pays special attention to the assurance of process safety assurance, which is in part realized by timely reactions to system alarms. An experiment involving an operator's reaction times to alarm signals was undertaken to investigate whether there are differential responses to visual as opposed to auditory alarms. The findings in the research show that visual alarm indicators are perceived faster than auditory signals. Furthermore, there was a significant negative correlation between the number of errors and reaction time, indicating an individual difference in error-proneness when reacting to visual alarms in a supervisory task.

Keywords: alarms; auditory signal; visual signal; human error; process safety; reaction time.

Error humano y reacción a alarmas en seguridad de procesos

Resumen

Operadores humanos responden a alarmas en condiciones normales y también en situaciones inesperadas y anormales en tiempo real. El objetivo de este estudio consiste en aumentar el conocimiento de las reacciones humanas a alarmas en el ámbito laboral. Describe como los humanos pueden contribuir a accidentes y presta atención a la seguridad del proceso que se efectúa en el tiempo de reacción a las alarmas de sistema. Este experimento fue ejecutado para investigar si hay diferencias en respuestas diferenciales a alarmas visuales comparándolas a alarmas auditivas. Los resultados indican que las alarmas visuales son percibidas de manera más eficiente que las alarmas auditivas. Incluso, se pudo ver una correlación negativa entre la cantidad de errores y el tiempo de reacción, lo que indica una diferencia individual en su disposición a errores cuando reaccionan a alarmas visuales en operaciones de supervisión.

Palabras Clave: alarmas; señales auditivas; señales visuales; error humano; seguridad de procesos; tiempo de reacción.

1. Introduction

Technological progress has contributed to more advanced human-machine systems. However, in complex systems, human error is still perceived as a cause or contributing factor to in approximately 90% of accidents. Most of these accidents could have been prevented by safety management measures [1]. This is particularly visible in the case of accidents in construction and manufacturing, in which errors

are viewed as inappropriate human decisions that have a negative impact on the operation of technical systems. Thereby they impact system performance by reducing safety and effectiveness [2]. Management of human errors is commonly focused on training, motivation, hardware design, and management systems, all of which can lead to improved safety performance [3,4].

In this paper, the role of humans in system failures in an industrial setting is discussed, and the particular element of alarm perception and reaction is tested in an experiment. The

How to cite: Mrugalska, B., Nazir, S., Tytyk, E. and Øvergård, K.I., Human error and response to alarms in process safety. DYNA 83(197), pp. 81-86, 2016.

study aims to illustrate the effect of different alarm signals on human operators' reaction times and identify alarms for which the response was defined according to a pre-defined code. Reaction time and the correctness of responses for compound reactions were assessed for auditory and visual signals in a laboratory simulation situation. The final part of the paper formulates recommendations for overall system performance in machinery and equipment design.

2. Human contribution to occupational accidents

Theories on the causes of occupational accidents have changed over the years. However, two basic perspectives on human error and human contribution to accidents can be identified [5,6]. The first approach, hereafter named the "old" approach began at the start of the 20th century. This approach states that "...*certain individuals are always more likely than others to sustain accidents, even though exposed to equal risk...*" [7, p.1]. The basic assumption in the 'old' approach was to see engineered systems as fundamentally safe, while unreliable humans were the primary threats to safety. According to this approach, the employer was not responsible for bad working conditions, and employees were accused of causing accidents [8]. The way to improve safety was to protect the systems through the training of operators, by keeping discipline, and through the application of procedures and through the selection of appropriate workers (e.g. by employing the less accident prone workers). In more recent times, this approach involved introducing automation in industrial processes [5]. It was not until the second half of the twentieth century that a system was seen as a factor that contributed to accidents. Employers were also required to provide safe working conditions for the first time. At the same time, it was realised that a person's accident proneness may differ from one period to another [7,9]. Moreover, research has shown that dismissing workers due to their accident proneness does not positively affect the ratio of occupational accidents [10]. The problem of accident proneness remains a widely discussed topic [11] and studies show that operators' accident proneness only refers to a small amount of accidents, for which the application of preventive safety barriers can reduce the involved risks [12].

Alternatively, according to the new approach, human error is no longer seen as a cause, but rather as a symptom of an underlying systemic weakness or failure. Hence, human error is not a cause in itself but it rather is something that must be explained. In contrast to the assumptions of the old approach, the new approach holds that safety is not intrinsic in systems as systems are created to fulfil multiple incompatible human goals simultaneously. Many of these goals can be conflicting, as is exemplified by the wish to improve safety and efficiency simultaneously (e.g. reducing relative number of errors while also increasing production). Human error is clearly associated with features of tools, tasks and operating environments. Therefore, human beings and their interaction with technology are important for creating safety and progress in terms of safety levels [5]. Thus, there is need for further investigation of human error and organizational deficiencies, design problems, and procedural shortcomings. Following on from this, it can be claimed that

Table 1.

Accident models: a review.

Author	Model	Accident causation
Heinrich (1931)	Domino theory	Ancestry and social environment, worker fault, unsafe act together with mechanical and physical hazard, accident, damage or injury
Bird & Germain (1985)	ILCI model	Lack of control: inadequate program, program standard, compliance to standard; basic causes: personal factors, job factors; immediate causes: substandard acts, substandard conditions; incident: contact with energy or substance; loss: people, property, environment, progress
Reason (1990)	Swiss cheese accident causation model	Fallible board decisions and policy, line management problems, psychological precursors of unsafe acts, unsafe acts, accident
Rasmussen (1997)	Risk management framework	Government, regulators associations, company, management, staff, work
Leveson (2004)	STAMP model	Inadequate control or enforcement of safety-related constraints, safety: control problem; accidents appear when component failures, external disturbances, dysfunctional interactions between system components are not adequately handled

Source: Adapted from [13-17].

the new approach to human error shows a significant movement towards human factors and points to new ways of managing organizational safety.

As there are different approaches to the causes of occupational accidents, the literature review allowed us to make a distinction between different accident theories, as are summarised Table 1.

These models provide an insight into how and why accidents happen in organization. However, it should be stressed that there are only a few accident models that are based on human factors research [11]. They refer to different factors that appear in many forms; however, generally they reflect human and technical aspects, when a human being can be recognized as assurance of safety. It should also be mentioned that the technical system design can be controlled/designed to reduce the occurrence of human errors, whereas the control of environmental factors and manner of human operation of the system are somewhat less controllable. Predominantly, in large organizations and compound technologies, more than one single liability can be distinguished [5,8,15,18].

3. Process safety performance

Over the last three decades, automation has changed the role of operators from manual control to supervision [10]. A great effort has been put into equipment design and interface, operation, operator selection and training and behavioral enforcement. These efforts have resulted in a drop of process

safety accidents. There is still a need to operate complex systems with the recognition of operation breakdowns that occur as a result of human error [18]. It should be emphasized that the main category of human error is operating errors. For example, the following situations can result in operating errors [19]:

- lack of appropriate procedures
- task complexity and overload conditions
- poor personnel selection and training
- operator carelessness and lack of interest
- poor environmental conditions
- departure from the correct operating procedures

In order to monitor the performance of systems the operators have to do many tasks that can rely on [20]:

- monitoring sensor data from components (voltages, status of the capacitors, transmission lines, transformers, circuit breakers, etc.)
- monitoring the grid as a holistic system
- communicating information among several sub-systems
- responding to alarms

Alarm systems play a significant role in industrial production as they assure process quality and reliability. Industrial production requires human operators who observe, and, if necessary, intervene in process malfunction or abnormal system performance *i.e.* when the values of the system parameters exceed predefined tolerance limits, or approach critical levels [21,22]. However, nowadays a large number of notifications, which are configured inside such a system, lead to the operator receiving much more alerts than s/he can physically and cognitively perceive. Due to the material, energy and information flow in a plant, particular disturbances can cause multiple subsequent alarm messages, which may overload the operator with redundant alarms. This results in an “alarm flood”, when an operator is not able to react within a recommended response time and find the root cause of disturbances. Finally, the industrial applications of alarm systems become unreliable due to their complexity [23].

In order to increase knowledge of human reactions to unreliable signals, researchers refer to theories of learning and human cognition such as:

- Probability matching - the approximation of the perceived true alarm rate on the basis of the observed alarm responses
- Signal Detection Theory - signal detection and response is more rapid for historically reliable alarm systems
- Automation trust - the exhibited trust relates to behavioral patterns of system use and misuse

Regardless of what the operation task is, alarm reaction patterns are unquestionably influenced by perceived alarm system reliability [24].

According to Lee and See [25] reliability of the alarm system is described as the percentage of critical events that are correctly identified by the alarm system. It is determined by the alarm's sensitivity, setting of thresholds and defined probability of dangerous conditions [26]. The higher the alarm system's reliability is, the more that operators can rely on the alarm and the less monitoring the data is required. On the other hand, when reliability is low, the chance of the occurrence of false alarms or misses is higher and the relevant

process data have to be monitored more frequently to be able to keep monitoring performance at a high level [27,28]. Unreliable systems can contribute to sensor failure or their inadequate performances, data processing limitations or output failure (*i.e.* electrical problems or component malfunction, etc.). All these states can lead to potentially hazardous situations as they cause more deviations in other parts of the plant [23].

4. Human reaction to alarms

Displays and controls have been used for a long time in human-machine interfaces. They range from simple to complex manufacturing systems and processes. There are also distinct sets of human factor principles for choosing frequency, duration, and intensity of signal presentation. In order to obtain attention and a response, the two most common modalities can be differentiated, which are visual and auditory signals.

In everyday life, most information is perceived with the eyes; however, in complex systems, where the object of the task is out of visual view, the use of auditory stimuli has become more common. Therefore, in industrial control visual and auditory signals are often presented together rather than as alternatives; however, they may also be separate or simultaneous, synchronous or asynchronous [29,30]. They can be used in hand and/or foot controlled communication devices. The foot controls are mainly designed for seated operators when there is a need to reduce the workload on the hands, or when free hands are required for tasks with greater levels of precision and skill. However, it should be emphasized that even in the case of equipment or systems that have foot controls (*i.e.* automobiles, airplanes), nearly all controls are assigned for hand control [31].

The literature studies revealed that the timing of auditory stimuli is generally longer than the timing of visual stimuli for the same signals presented separately. Furthermore, duration discrimination is more precise for auditory than for visual stimuli in simultaneous presentations [32]. It was also identified that the time perception relies on the prefrontal and parietal regions, and on the auditory and visual cortices. However, the neural mechanism responsible for these modalities remains unclear [33]. Moreover, in recognition of visual and auditory signals, the phenomenon called visual dominance [34] must be taken into account. Based on this, auditory signals may seem not to be responded to in simultaneous presentation with visual signals, which indicates that their use can be ineffective in control situations. Nevertheless, in many situations, it was recognized that they might stimulate situational awareness and thereby improve the effectiveness of visual displays. They are superior to visual signals for warning in monitoring tasks in which there is no need to observe a particular location all the time [35].

The level of urgency of alarms is correlated with the level of urgency of the indicated danger [36]. Furthermore, the perception of alarms is subject to their frequency, intensity, and repetition. For instance, a low frequency alarm, which is repeated slowly, shows low urgency, whereas a high frequency alarm repeated quickly implies high urgency [37]. In simulated conditions it was shown that the reduction of

luminous intensity from 42 to 5 cd increases the reaction time by 84 ms [38]. In the case of auditory signals, increased signal intensity causes faster reaction time, whereas its high-intensity, which is synchronous with a visual signal, reduces choice reaction time [39]. Thus, the growth of the intensity of auditory signal weakens the visual supremacy. For synchronous auditory and visual tasks, sound source location does not have a facilitating effect. It was concluded that sound not only warns, but also has an overall alerting effect. Auditory signals are quite useful when the spatial-visual information is limited or chaotic. For instance, they can be applied in control console tasks in which they alert operators or request action. While listening to critical auditory messages, visual signals are monitored. To optimize their reaction rate and reduce their performance lag, a fore-period warning (at least 2s and no more than 4s) is required to draw attention to incoming signals. Both signals should be simultaneously presented from the same direction to ensure quick responses. Nevertheless, in the case of asynchronous signals, a short delay (200 ms) between them may give a negative result, giving a poor response speed and accuracy [29].

Base on the literature study, it was hypothesized that reaction times for visual alarms would be significantly shorter than reaction times for audible alarms. It was also postulated that audible alarms would have a higher percentage of errors than visual alarms.

5. Materials and methods

Eighty-seven second-year students from engineering management and safety engineering courses in Poznan University of Technology participated in the experiment. The sample consisted of 36 males and 51 females, with a mean age of 20.16 years ($SD = 0.37$). They did not report having hearing loss or color vision deficiency. In order to have free movements they were asked to wear casual and comfortable clothing. As the experiment was conducted during their laboratory classes, all the students were awarded a course credit.

At the beginning of the experiment the task was clearly explained, and then participants were allowed to practice until they became familiarized with it. Their individual training lasted for about 5 minutes. Afterwards, they experienced 3 visual and 1 auditory signals while operating the simulator (MCR-2001E). The experimental task was to react to 50 signals according to the previously defined code, which is shown in Table 2. When the experimental task

started, a single auditory or visual target signal appeared on its own for each test trial. No feedback on the accuracy of responses was given to the participants.

Only reaction times for correct responses were recorded. Incorrect responses were given a reaction time of zero (0) and treated as a missing variable in the statistical analysis. Mean reaction time was calculated by summing up all the correct response times for each participant and then dividing this sum by the number of correct responses for visual and auditory alarms respectively.

IBM SPSS 22 was used for statistical analysis. All confidence intervals for r (Pearson correlation) were calculated using the bootstrap function in SPSS, and there were 1000 resamplings that had replacement.

6. Results and discussion

Visual alarms ($M = 524.67$ ms, 99% CI [506.39, 540.73]) had a significantly shorter reaction time than audible alarms ($M = 579.27$ ms, 99% CI [552.88, 608.64]) as measured by a paired samples t -test ($t_{(86)} = -6.336$, $M_{diff} = -54.61$ ms, 99% CI of M_{diff} [-77.31, -31.90], $p < .001$). This is in accordance with the literature review and the stated expectations of this study.

Following this difference in mean reaction time, a difference in delayed (but correct) responses was expected. Delayed responses were defined as responses with more than 500 ms reaction time. Visual alarms had a much lower ratio of delayed responses ($M = 42.78\%$, 99% CI [37.37%, 49.47%]) than audible alarms ($M = 57.12\%$, 99% CI [49.01%, 65.57%]). This difference was statistically significant as measured by a paired samples t -test ($t_{(86)} = -4.869$, $M_{diff} = -14.33\%$, 99% CI of M_{diff} [-22.09%, -6.58%], $p < .001$).

Regarding the number of errors, a paired samples t -test showed that there was no difference in the ratio of errors for visual alarms ($M = 11.30\%$, 99% CI [9.35%, 13.49%]) as compared to audible alarms ($M = 10.49\%$, 99% CI [7.61%, 13.36%]). The difference was not significant ($t_{(86)} = 0.616$, $M_{diff} = -0.008\%$, 99% CI of M_{diff} [-0.027, 0.043], $p = .54$). This is not in accordance with the hypothesized difference that audible alarms would be more prone to errors.

When the relationship between the rate of errors and reaction time was further investigated, a statistically significant negative correlation was found between reaction time for visual alarms and the ratio of errors to correct responses ($r_{(87)} = -.301$, 95% CI [-.484, -.085]). This indicated that the participants who on average responded faster to visual alarms also tended to make a larger number of errors; therefore, this indicated an individual difference in error-proneness when reacting to visual alarms in a supervisory task.

For audible alarms, the same negative correlation between reaction times and ratio of errors was observed, but this relationship was not statistically significant ($r_{(87)} = -.095$, 95% CI [-.223, .117]). The lack of a significant negative correlation in this study could be due to the relatively lower number of auditory alarms (8 in total) compared to visual alarms (42 in total).

This study employed a non-randomized convenience

Table 2.

Alarm sequence representation

Element number	Modality	Element number	Modality
1	visual	26	audible
2	audible	27-31	visual
3-5	visual	32	audible
6	audible	33-35	visual
7-9	visual	36	audible
10	audible	37-39	visual
11-22	visual	40	audible
23	audible	41-50	visual
24-25	visual		

Source: The authors.

sample, hence limiting the ability to make causal inferences outside the current sample. However, our findings are still in accordance with the existing literature on this matter.

The limited number of audible alarms has restricted our ability to identify real differences in error rates when responding to audible alarms. This limitation can explain our lack of ability to statistically identify differences between error rates as well as the lack of a significant negative correlation between reaction time and error rates for audible alarms. This limitation can be amended in future studies by balancing the number of visual to audible alarms.

7. Conclusions

In any process industry, safety issues are of primary importance as any accident can have cataclysmic consequences for both people and the environment. Therefore, efforts must be made to prevent or limit the hazards of the plant. To achieve this aim, an increasing number of industrial quality and reliability systems is applied. In these systems, operators pay attention to present process states and predict their future states using information from alarm systems. Alarms are frequently performed via display terminals in which information is presented in a variety of formats. In this paper, particular attention has been paid to visual and audible signals. The results of the present study revealed that the participants could react faster to visual signals than audible ones when the signals are predefined according to a code. No difference was found in the reaction of errors compared with the total number of alarms.

These results aim to be helpful in formulating recommendations to work towards a more optimal design of control consoles when auditory and visual displays are presented. They may be also effective in improving system performance in machinery and equipment design.

References

- [1] Leva, M.C., Human errors analysis and safety management systems in hazardous activities. [Online]. (Interim Report IR-05e003), 2005. Available at: <http://web.archive.iiasa.ac.at/Admin/PUB/Documents/IR-05-003.pdf>.
- [2] Sanders, M. and McCormick, E., Human factors in engineering and design. New York: McGraw-Hill 1993.
- [3] Górný, A., Ergonomics in the formation of work condition quality. Work. A Journal of Prevention, Assessment and Rehabilitation, 41, pp. 1708-1711, 2013.
- [4] Belu, N., Misztal, A., Butlewski, M. and Rachieru, N., The directions for improvement in Polish and Romanian enterprises in applying the principles of Social Responsibility, Proceedings of the 4th Review of Management and Economic Engineering International Management Conference, pp. 339-344, 2014.
- [5] Dekker, S.W.A., Reconstructing human contributions to accidents: The new view on error and performance. Journal of Safety Research, 33, pp. 371-385, 2002. DOI: 10.1016/S0022-4375(02)00032-4
- [6] Sanmiquel-Pera, L., Vintro, C. and Freijo, M., Characteristics of the 3 most common types of occupational accident in Spanish sub-surface and surface mining, from 2003-2008. DYNA, 79(172), pp. 118-125, 2012.
- [7] Froggatt, P. and Smiley, J.A., The concept of accident proneness: A review. British Journal of Industrial Medicine, 21(1), pp. 1-12, 1964. DOI: 10.1136/oem.21.1.1
- [8] Mrugalska, B., Nazir, S., Tytyk, E. and Øvergård, K.I., Process safety control for occupational accident prevention, in: Arezes, P.M., Baptista, J.S., Barroso, M.P. et al. (eds.) Occupational Safety and Hygiene III, (SHO), London: Taylor and Francis Group, 2015. pp. 365-369. DOI: 10.1201/b18042-74
- [9] Heinrich, H.W., Industrial accident. London: McGraw-Hill, 1959.
- [10] Hämäläinen, P., Global estimates of occupational accidents and fatal work-related diseases. PhD Thesis, Tampere University of Technology, Tampere, Finland, 2010.
- [11] Nazir, S. and Manca, D., How a plant simulator can improve industrial safety. Process Safety Progress, 34(3), pp. 237-243, 2015. DOI: 10.1002/prs.11714. DOI: 10.1002/prs.11714
- [12] Kjellen, U., Prevention of accidents through experience feedback. London: Taylor & Francis Group, 2000. DOI: 10.1201/b17206
- [13] Heinrich, H.W., Industrial accident prevention. London: McGraw-Hill, 1931.
- [14] Bird, F.E.J., and Germain, G.L., Practical loss control leadership. Loganville, Georgia: International Loss Control Institute, Inc., 1985.
- [15] Reason, J., Human error. Cambridge: Cambridge University Press, 1990. DOI: 10.1017/CBO9781139062367
- [16] Rasmussen, J., Risk management in a dynamic society. Safety Science, 27, pp. 183-213, 1997. DOI: 10.1016/S0925-7535(97)00052-0
- [17] Leveson, N., A new accident model for engineering safer systems. Safety Science, 42, pp. 237-270, 2004. DOI: 10.1016/S0925-7535(03)00047-X
- [18] Grigoras, G. and Barbulescu, C., Human errors monitoring in electrical transmission networks based on a partitioning algorithm. Electrical Power and Energy Systems, 49, pp. 128-136, 2013. DOI: 10.1016/j.ijepes.2012.12.016
- [19] McLeod, J.E.N. and Rivera, S., Human error management optimization in CAREM NPP. Proceedings of the World Congress on Engineering (WCE), pp. 618-622, 2009.
- [20] Su, R. and Yurcik, W., A survey and comparison of human monitoring of complex network. National Center for Computing Applications (NCSA), [Online]. University of Illinois at Urbana-Champaign. Available at: http://www.dodccrp.org/events/10th_ICCRTS/CD/papers/342.pdf.
- [21] Manca, D., Nazir, S., Colombo, S. and Kluge, A., Procedure for automated assessment of industrial operators. Chemical Engineering Transactions, 36, pp. 391-396, 2014.
- [22] Mrugalska, B., Akielaszek-Witczak, A. and Stetter, R., Robust quality control of products with experimental design. In D. Popescu (ed.), International Conference on Production Research – Regional Conference Africa, Europe and the Middle East and 3rd International Conference on Quality and Innovation in Engineering and Management, Cluj-Napoca, Romania, pp. 343-348, 2014.
- [23] Schlegel, M., Christiansen, L., Thornhill, N.F. and Fay, A., A combined analysis of plant connectivity and alarm logs to reduce the number of alerts in an automation system. Journal of Process Control, 23, pp. 839-851, 2013. DOI: 10.1016/j.jprocont.2013.03.010
- [24] Bliss, J.P. and Chancey, E.T., An investigation of training strategies to improve alarm reactions. Applied Ergonomics, 45, pp. 1278-1284, 2014. DOI: 10.1016/j.apergo.2013.06.007
- [25] Lee, J.D. and See, K.A., Trust in automation: designing for appropriate reliance. Hum. Factors, 46, pp. 50-80, 2004. DOI: 10.1518/hfes.46.1.50.30392
- [26] Bustamante, E.A., Bliss, J.P. and Anderson B.L., Effects of varying the threshold of alarm systems and workload on human performance. Ergonomics, 50(7), pp. 1127-1147, 2007. DOI: 10.1080/00140130701237345
- [27] Velez, J.C., Lishak, M.Y. and Antonov-Antipov, Y.N., False alarm and detection probabilities calculation of a digital unified detector. DYNA, 77(162), pp. 271-280, 2010.
- [28] Onnasch, L., Ruff, S. and Manzey, D., Operators' adaptation to imperfect automation – Impact of miss-prone alarm systems on attention allocation and performance. International Journal of Human-Computer Studies, 72(10-11), pp. 772-782, 2014. DOI: 10.1016/j.ijhcs.2014.05.001
- [29] Lee, F.C.H. and Chan, A.H.S., Attending visual and auditory signals: Ergonomics recommendations with consideration of signal modality and spatial stimulus-response (S-R) compatibility. International

- Journal of Industrial Ergonomics, 37, pp. 197-206, 2007. DOI: 10.1016/j.ergon.2006.10.017
- [30] Tsang, S.N.H., Kang S.X.Q. and Chan, A.H.S., Spatial S-R compatibility effect with head rotation. Proceedings of the International MultiConference of Engineers and Computer Scientists (IMECS), pp. 1115-1119, 2014.
- [31] Chan, A.H.S. and Chan, K.W.L., Three-dimensional spatial stimulus-response (S-R) compatibility for visual signals with hand and foot controls. Applied Ergonomics, 41, pp. 840-848, 2010. DOI: 10.1016/j.apergo.2010.03.005
- [32] Wearden, J.H., Todd, N.P.M. and Jones, I.A., When do auditory/visual differences in duration judgements occur? Q. J. Exp. Psychol., 59, pp. 1709-1724, 2006. DOI: 10.1080/17470210500314729
- [33] Chen, Y., Huang, X., Luo, Y., Peng, C. and Liu, C., Differences in the neural basis of automatic auditory and visual time perception: ERP evidence from an across-modal delayed response oddball task. Brain Research, 1325, pp. 100-111, 2010. DOI: 10.1016/j.brainres.2010.02.040
- [34] Posner, M.I., Orienting of attention. Quarterly Journal of Experimental Psychology, 32, pp. 3-25, 1980. DOI: 10.1080/00335558008248231
- [35] Lee, F.C.H. and Chan, A.H.S., Ergonomics recommendations for simultaneous and delayed presentation of visual and auditory signals. Displays, 29, pp. 124-131, 2008. DOI: 10.1016/j.displa.2007.09.006
- [36] Arrabito, G.R., Mondor, T.A. and Kent K.J., Judging the urgency of non-verbal auditory alarms: a case study. Ergonomics, 47(8), pp. 821-840, 2004. DOI: 10.1080/0014013042000193282
- [37] Parasuraman, R., Hancock, P.A. and Olofinboba, O., Alarm effectiveness in driver-centred collision-warning systems. Ergonomics, 40, pp. 390-399, 1997. DOI: 10.1080/001401397188224
- [38] Sivak, M., Flannagan M.J., Sato, T., Traube E.C. and Aoki, M., Reaction times to neon, LED, and fast incandescent brake lamps. Ergonomics, 37(6), 1994. DOI: 10.1080/00140139408963712
- [39] Keuss, P.J.G., Van Der Zee, F. and Van Der Bree, M.B.M., Auditory accessory effects on visual processing. Acta Psychologica, 75(1), pp. 41-54, 1990. DOI: 10.1016/0001-6918(90)90065-N

B. Mrugalska, received her MSc. Eng. in Corporate Management in 2001, her PhD. in Machines Building and Operations in 2008, from Poznan University of Technology, Poland. Since 2009, she has been working in the Chair of Ergonomics and Quality Engineering at the Faculty of Engineering Management in Poznan University of Technology. She is a Centre for Registration of European Ergonomists (CREE) Council representative in her capacity as a member of the Polish National Assessment Board. Her research interests include: robust design and control, parameter optimization techniques, system faults and failures, quality control, machinery safety, process safety and reliability.
ORCID: 0000-0001-9827-9449

S. Nazir, is a researcher at Human Factors Research Group at Buskerud and Vestfold University College, where he works in the Faculty of Technology and Maritime Sciences. He received his MSc. and PhD in Chemical Engineering (focuses on System Engineering and Human Factors respectively) from Hanyang University, South Korea and Politecnico di Milano, Italy, respectively. His research interests include: human factors, training methods, virtual reality training simulators, performance indices, complex socio-technical systems, system safety, and distributed situation awareness. Dr. Nazir has won several awards, scholarships and some research grants during his career. He was awarded with Cum Laude (outstanding honor) when awarded his PhD.
ORCID: 0000-0002-2058-6147

E. Tytyk, received his MSc. Eng. in Plastic Working in 1972, his PhD. degree in Machines Building in 1979, and then his Dr. in Machines Building and Ergonomics in 1991. He has worked as a full professor in Design Methodology and Ergonomics since 2004 in the Poznan University of Technology, Poland. Currently, he is a head of the Chair of Ergonomics and Quality Engineering at the Faculty of Engineering Management in Poznan University of Technology. He works as Full Professor both in the Poznan

University of Technology and in the Komenski's High School Academy in Leszno, Poland. His research interests include: ergonomic design methodology, ergonomic design for disabled and senior citizens, design of technical and organizational solutions to improve working conditions in different plants and technologies.
ORCID: 0000-0001-9892-4380

K.I. Øvergård, received his BSc. in Welfare Nursing/Developmental Psychology from the University College of Akershus in 2001. He was awarded his MSc. in Psychology with a specialization in Ergonomics in 2004, and his PhD in Psychology – Cognitive Systems Engineering in 2008, from the Norwegian University of Science and Technology in Trondheim, Norway. Øvergård worked at the Accident Investigation Board of Norway and later at the Norwegian Railroad. In 2011 he was employed as Full Professor in Work and Organizational Psychology–Maritime Human Factors at Buskerud and Vestfold University College, Norway – where he is now the Head of Maritime Research at the Faculty of Technology and Maritime Sciences. In 2012 he became Chair of Cognitive Systems Engineering and Human Factors at the Global Maritime Knowledge Hub. His research interests include: maritime human factors, identifying characteristics and causes of organizational failures/resilience, applied statistics and involvement of human error in accidents.
ORCID: 0000-0002-4029-4344



UNIVERSIDAD NACIONAL DE COLOMBIA

SEDE MEDELLÍN

FACULTAD DE MINAS

Área Curricular de Ingeniería Administrativa e
Ingeniería Industrial

Oferta de Posgrados

Especialización en Gestión Empresarial
Especialización en Ingeniería Financiera
Maestría en Ingeniería Administrativa
Maestría en Ingeniería Industrial
Doctorado en Ingeniería - Industria y Organizaciones

Mayor información:

E-mail: acia_med@unal.edu.co
Teléfono: (57-4) 425 52 02

Changes in EEG amplitude (*Alpha* and *Beta* waves) with Thermal environment

Emanuel Tiago-Costa ^a, Emília Quelhas-Costa ^b & João Santos-Baptista ^c

^{a,b,c} Research Laboratory on Prevention of Occupational and Environmental Risks (PROA), Porto, Portugal

^{b,c} Associated Laboratory for Energy, Transports and Aeronautics (LAETA), Faculty of Engineering, University of Porto, Porto, Portugal.

^a ega08027@fe.up.pt; ^b eqc@fe.up.pt; ^c jsbap@fe.up.pt

Received: November 30th, 2015. Received in revised form: March 17th, 2016. Accepted: April 25th, 2016.

Abstract

When executing a task, brain activity can be observed through electric waves recorded in different frequency bands. The main ones are the *Alpha*, *Beta*, *Theta* and *Delta* waves. According to some authors, *Alpha* and *Beta* waves are related with cognitive capacity and, their presence and evolution, with an individual's alertness. In this study, the intention was to evaluate how brain activity is directly influenced by temperature and humidity while performing a task. Therefore, a set of laboratory tests were undertaken to simulate sedentary work in two different thermal environments: 18°C - 40% RH (cold) and 35°C - 80% RH (hot). In the hot thermal environment, oscillation on *Alpha* waves amplitude suggests an alternation between periods of tiredness with periods in which concentration increases. Regarding *Beta* waves in the cold thermal environment, most of the time *Beta* activity is typically higher than that verified in the hot thermal environment.

Keywords: *Alpha* waves, *Beta* waves, thermal environmental, cognitive capacity.

Cambios en la amplitud del EEG (ondas *Alfa* y *Beta*) con el medio ambiente térmico

Resumen

Cuando se ejecuta una tarea, puede observarse actividad cerebral a través de las ondas eléctricas en diversas bandas de frecuencia. Los principales son las ondas Alfa, Beta, Theta y Delta. Según algunos autores, las ondas Beta y Alfa están relacionados con la capacidad cognitiva y, su presencia y evolución, la aparición y evolución, con el estado de alerta de un individuo. En este estudio, la intención fue evaluar cómo actividad cerebral está directamente influenciada por la temperatura y humedad mientras realiza una tarea. Por lo tanto, se llevaron a cabo un conjunto de pruebas de laboratorio para simular trabajo sedentario en dos diferentes ambientes térmicos: 18 °C - 40% RH (en frío) y 35 °C - 80% RH (caliente). En el ambiente térmico caliente, la oscilación de amplitud de las ondas alfa sugiere una alternancia entre periodos de cansancio con los periodos en que aumenta la concentración. Con respecto a las ondas Beta en el ambiente térmico frío, la mayoría de la actividad Beta es típicamente mayor que el verificado en el ambiente térmico caliente.

Palabras clave: ondas Alfa, ondas Beta, ambiente térmico, capacidad cognitiva.

1. Introduction

1.1. Occupational risk and thermal environmental

The economic and social development that has taken place throughout the 20th century has been accompanied by a greater awareness of work safety and health problems and

also by the development of protective measures for workers. The objective has been to minimize risk. Despite these efforts, there are still working situations that can be very harmful to workers' safety and health. One of these situations is related to workers who perform activities in "extreme" thermal environments, even if these are sedentary activities.

It is pertinent to assess the relationship between thermal

How to cite: Tiago-Costa, E., Quelhas-Costa, E. and Santos-Baptista, J., Changes in EEG amplitude (*Alpha* and *Beta* waves) with Thermal environment. DYNA 83(197), pp. 87-93, 2016.

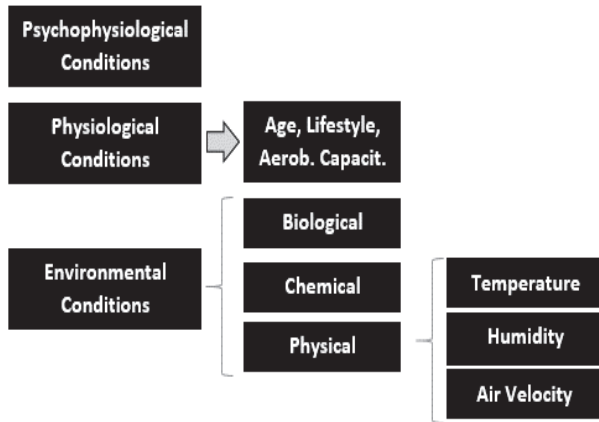


Figure 1. Influence of different conditions in human ability. Source: [10].

environment and cognitive capacity since all human activity is influenced by the surrounding environment, including working activities. This allows the question of how thermal environment influences human being's responsiveness to be clarified. Over the course of undertaking this research, we found several studies on the subject that have been performed in controlled environments, such as climatic chambers [1-5]. Furthermore some of the relevant literature [6,7] states that thermal environment can have a negative impact on performance, particularly when combined with other variables such as exposition time or task duration.

With increased technological development the workplace assumes a greater complexity. This increased complexity often brings additional risks for workers. Increasing technological complexity requires, from the worker, an increased cognitive performance, which may be jeopardized due to an inadequate thermal environment.

1.2. Influence of environment in human ability

The way that each individual feels the surrounding environment is influenced by intrinsic and extrinsic factors from the individual himself [8,9]. Tolerance to environment depends directly on, among others factors, age, aerobic capacity, hours of rest, lifestyle and consumption of medicines. These factors can be classified as shown in Fig. 1.

All of these strongly influence the feeling of well-being. The psychophysiological and physiological factors mean that different individuals in similar environmental conditions feel the conditions to which they are exposed differently.

1.3. Characterization and interpretation of brain waves

Monitorization and interpretation of brain waves can clarify how cognitive capacity is influenced by the thermal environment. The different types of brain waves / activities are:

- *Delta Activity*– is referred to as a "slow wave"; it is in the frequency range between 0.5 Hz and 4 Hz. This type of activity is recorded in the transitional phase for sleep and during sleep [11].
- *Teta Frequency*– this lies between 4 Hz and 7 Hz. It is

associated with different psychological states, which include low levels of alertness and, consequently, is associated with the decrease in the capacity to process information. Theta frequency can also be considered a "slow wave" [12].

- *Alpha waves*– This type is dominant in adult human brains, and has a frequency of between 8 Hz and 13 Hz. The waves' presence is common in the occipital cortex, during the period of wakefulness. Whenever the eyes are closed, the presence of this type of waves is common. When head wavers due to fatigue occurs, and when trying to be aware, Alpha wave activity decreases [11]. The presence of these waves during when the individual is awake and in a state of relaxation, leads to a slow response to stimuli [13].
- *Beta waves*– These have a frequency of between 13 Hz and 30 Hz, which indicates a high propagation speed. Their presence is associated with a state of arousal/alert. Beta activity can occur in humans while performing a motor task that requires high reaction time [14].

This study aims to contribute by establishing a relationship between thermal environment, and the signal amplitude of brain waves (Alpha and Beta), in particular the cognitive performance of individuals under uncomfortable environmental conditions.

2. Materials and Method

2.1. Main equipment

Monitoring and data collection was done using specific equipment. Table 1 presents the equipment that was used and its respective functions. After all the equipment [15] to be used was selected, each piece was tested and validated under trial protocol [16].

2.2. General trial conditions

The relationship between cognitive capacity, thermal environment and brain waves was evaluated by performing experimental trials with 10 male volunteers. Their ages were between 18 and 36 years old, with an average of 26±6 years.

In these trials, each volunteer performed two cognitive tests. The first was held at 18°C - 40% RH, and the second was done seven days later at 35°C - 80% RH. All the trials were performed between 6:00 pm and 8:00 pm. The operating conditions were simulated within a Climate Chamber with an error of ±1°C and ±2% RH.

Table 1.
Equipment used in laboratory tests.

Equipment		Function
Brain-sensors	(Emotiv Epoc – EEG*)	Monetarization of brain activity
Climatic Chamber	(Fito-Clima 2500EC20)	Simulation of thermal environments
Internal Thermal Sensor	(Vital Sense)	Monetarization of core temperature
Skin Thermal Sensors	(BioPlux)	Monetarization of skin temperature

* electroencephalogram.

Source: [17].

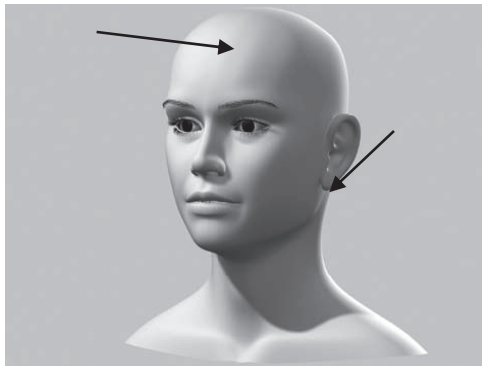


Figure 2. Neck and forehead points to control the skin temperature.
Source: [18]

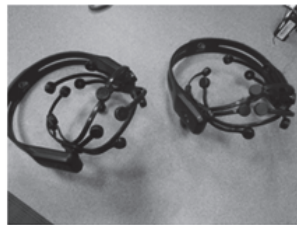


Figure 3. Emotiv equipment and the respective signal.
Source: www.engr.ucr.edu

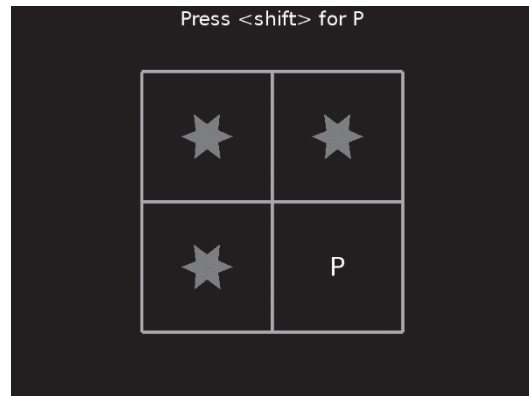
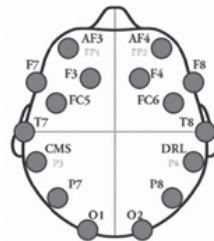


Figure 4. Go/No-Go test image.
Source: [20]

given instructions. Each volunteer undertook a one hour task, in this case, the Go / No-Go test [20] (Fig. 4) and completed it as successfully as possible. Go/No-Go is a selective attention test that does not have learning effects over the results as is totally random.

Before starting the test, we explained what we intended to happen and the risks to the volunteer and an informed consent was signed. The study was approved by the Ethics Committee of the University of Porto (nº04 CEUP/2012).

2.3. Questionnaires

Before entering the Climate Chamber the volunteer answered 4 questions in order to assess their lifestyle, namely: **a** -Ingested caffeinated drinks in the last 12 hours?; **b** -Ingested alcoholic drinks in the last 12 hours?; **c** - Any medicine taken?; **d** - Good overnight rest?

These questions are relevant as they allow us to know if the volunteer is tired at the beginning of the test and, therefore, if task performance could be affected by other factors in addition to temperature and humidity.

Once the test was completed (1 hour), all the equipment was turned off, and the volunteer leaves the Climatic Chamber.

3. Results and discussion

3.1. Lifestyle

The answers given by the volunteers about their lifestyle before starting the tests helped to interpret the results (Table 2). Lifestyle can interfere with the results, for example, disguising fatigue (coffee) or reducing concentration power (alcohol).

Although some volunteers had a lifestyle that may influence the results (gray in Table 2), i.e. inserting biases in the results, we decided to include them if they are usual consumers of small doses.

3.2. Skin and core body temperature

Fig. 5 shows the evolution of the skin temperature average in the forehead and neck at 18°C - 40% RH for the ten volunteers. Fig. 6 shows the evolution of skin temperature

Each volunteer was submitted to each one of the environmental conditions for 60 minutes after a stabilization period of 20 minutes outside the chamber (in the laboratory) at a temperature between 20°C and 22°C and a relative humidity between 55% RH and 65% RH. After this period, the volunteer who entered the climatic chamber waited for 10 minutes under environmental test conditions before beginning the essay. These 10 minutes were used to install an EEG device on the volunteer's head. The total time for each experiment was 90 minutes. Control of skin temperature was undertaken at two points, one on the neck and the other on the forehead of the volunteer, as shown in Fig. 2. In all the tests, all volunteers were dressed with 0,7 clo.

For the trials at 35°C - 80% RH, core temperature was also controlled. In total were conducted 20 tests. Skin temperature allowed individual stabilization to be controlled. Core temperature allows signs of thermal stress to be identified, especially when the organism was exposed to high temperatures (35°C - 80% RH) [19]. This monitoring also allows the risks of excessive exposure to heat to be minimized.

To evaluate brain activity we used a specific interface (brain computer interface) that allowed Alpha and Beta waves to be registered. Only brain waves from the left and right occipital were studied. The occipital lobe is related to the visual sense and to the ability to process information and understanding its contents.

All selected volunteers were male. To obtain a high-quality signal, all electrodes must be connected (in green), as is shown in Fig. 3.

Volunteers' cognitive capacity was stimulated by performing a "game/task". Each individual had to follow the

average at 35°C - 80% RH at the same skin points. In both environments, skin temperature stabilizes. This level is clearer on the forehead and for hot temperatures. It is also interesting to notice that in hot environments, on both control points, neck and forehead, the temperature value has almost the same value after stabilization.

The core body temperature that was measured only in the hot environment (35°C - 80% RH) shows a great stability over time (Fig. 7).

Table 2.
Lifestyle Assessment for each volunteer.

Volunteer	35°C - 80% RH				18°C - 40% RH			
	a	b	c	d	a	b	c	d
1	Yes	No	No	Yes	Yes	No	No	Yes
2	No	No	No	Yes	No	No	Yes	Yes
3	Yes	No	No	Yes	Yes	No	No	Yes
4	No	No	No	Yes	No	No	No	No
5	No	Yes	No	Yes	No	No	Yes	No
6	Yes	No	No	Yes	No	No	No	Yes
7	No	No	No	No	No	No	No	Yes
8	Yes	No	No	Yes	Yes	No	No	No
9	No	No	No	Yes	Yes	Yes	No	No
10	No	No	No	No	No	No	No	Yes
Yes	40%	10%	0%	80%	40%	10%	20%	60%
No	60%	90%	100%	20%	60%	90%	80%	40%

Source: The authors

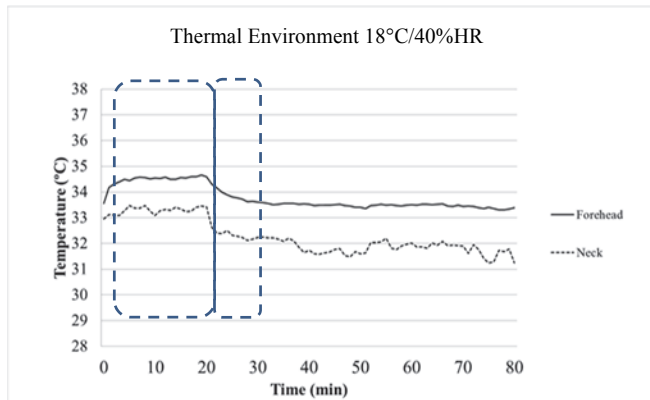


Figure 5 Forehead and neck mean temperature
Source: The authors

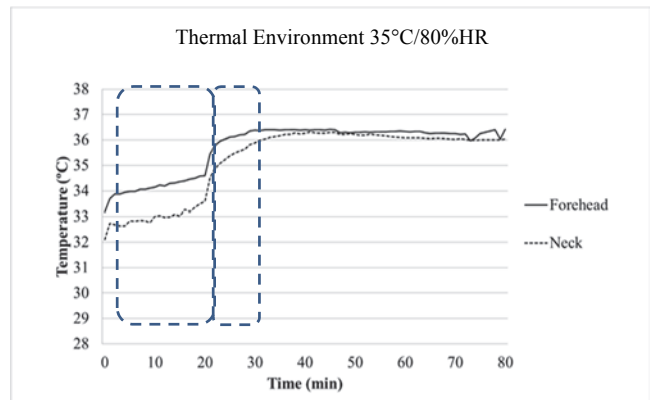


Figure 6 Forehead and neck mean temperature
Source: The authors

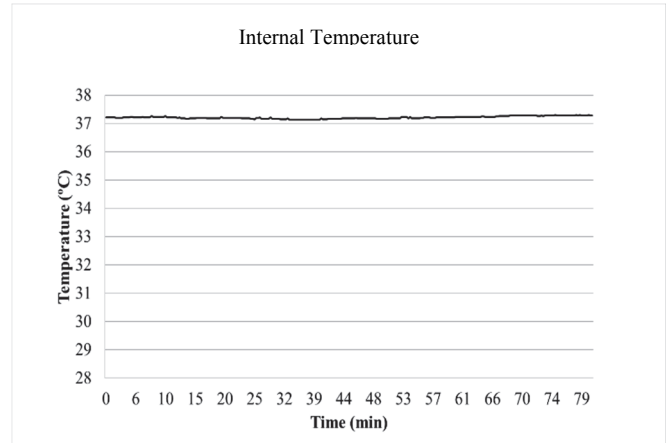


Figure 7 Core body temperature average at 35°C/80%RH
Source: The authors

3.3. Brain activity

Figs. 8 to 11 represent the average results for recorded brain activity. When Figs. 8 and 9 are compared with the Figs. 10 and 11, it is possible to confirm thermal environmental influence on brain activity (Alpha and Beta waves' amplitude). For the colder environment (18°C - 40% RH) Alpha activity is far superior to that obtained in the hot environment (35°C - 80% RH). This excludes the occasionally-raised hypothesis that the perspiration is the main cause of the occurrence of high values of the EEG signal for hot environments.

In a state of wakefulness in which the volunteers are concentrated (eyes open), Alpha activity is usually small. As individuals become fatigued Alpha activity is expected to increase. However, when time goes on, in the case of the trials at 35°C - 80% RH (Fig. 8), only a small decrease in Alpha activity can be observed, which does not verify the expected trend. For the trials at 18°C and 40% RH (Fig. 9), there is a decrease in Alpha activity in the first part of the test (20-30 min); this then increases until Alpha activity maintains values that are close to the initial values. The increase in Alpha activity suggests an increase of tiredness/fatigue in the volunteers.

Another relevant fact is the discrepancy of activity between the right and left brain hemispheres for hot thermal environments. In the left area of the brain (occipital 1) the range values activity is between 4-9 μ V while in right area (occipital 2) this range rises to 11-19 μ V (Fig. 8). It seems that in a hot thermal environment the temperature and relative humidity affect the left and right hemispheres differently.

According to several studies [11], Beta activity is considered the one that best describes alertness/arousal.

In a cold thermal environment, Beta activity is typically higher than the activity verified for the hot thermal environment.

It is usually considered that a decrease in Beta activity is related to the loss of the ability to concentration (reduction of surveillance), which was verified throughout the trial. It is therefore possible to consider that cold thermal environments (Fig. 10) are conducive to higher levels of concentration/alertness than hot environments (Fig. 11).

When comparing the activity between left side and right side of the brain, the activity in the occipital 1 (left) is lower than the one verified in occipital 2 (right).

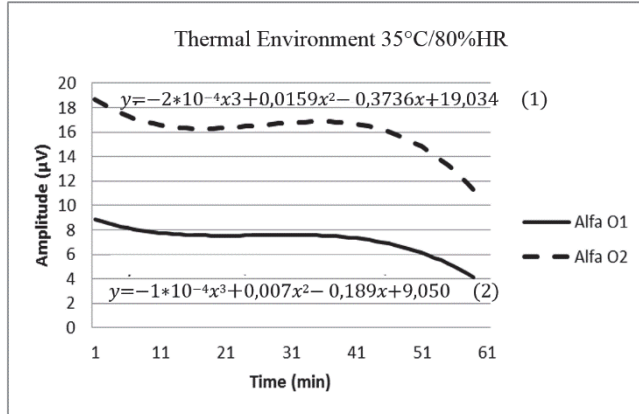


Figure 8. Alpha Brain Activity
Source: The authors

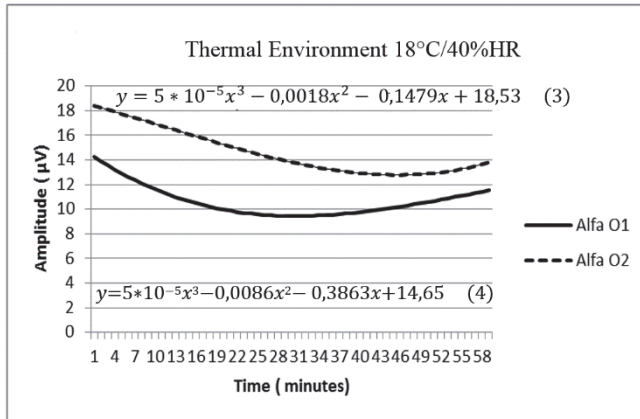


Figure 9. Alpha Brain Activity.
Source: The authors

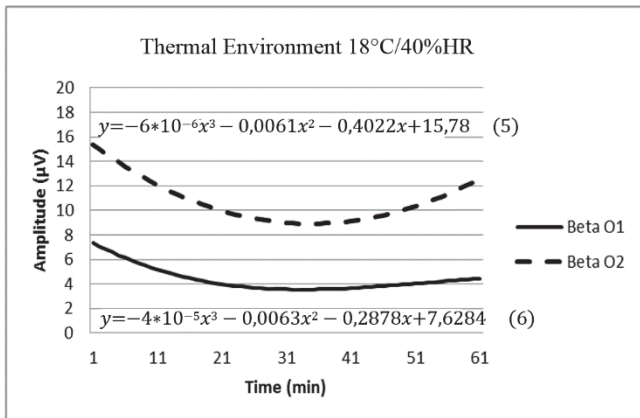


Figure 10. Beta Brain Activity
Source: The authors

After analyzing the data of brain activity for the two frequency bands (*Alpha* and *Beta*) it is possible to calculate the *Alpha/Beta* ratio. This ratio is developed to relate the two

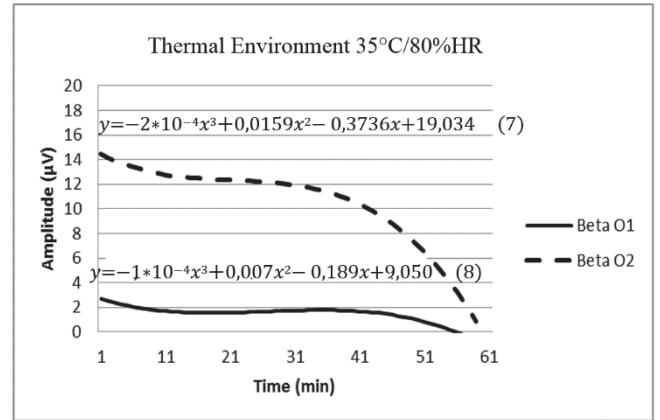


Figure 11. Beta Brain Activity
Source: The authors

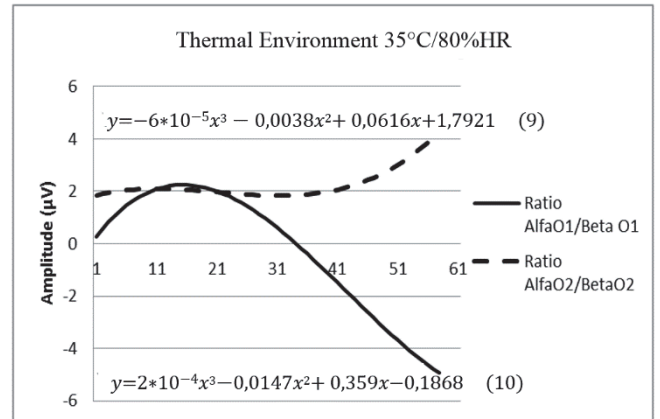


Figure 12. Ratio Alpha /Beta
Source: The authors

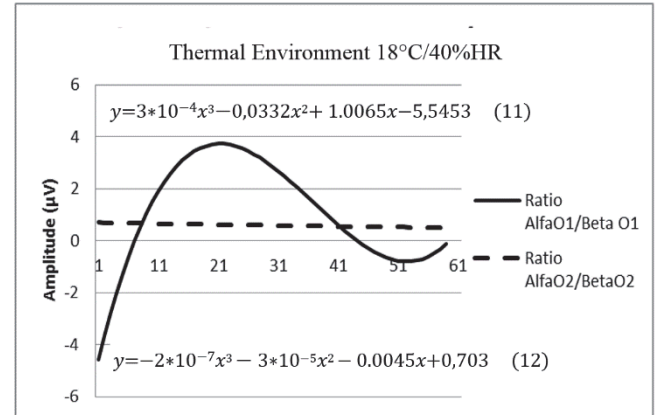


Figure 13. Ratio Alpha /Beta
Source: The authors

most relevant frequencies, which facilitates understanding of the evolution of individual's cognitive state.

With the analysis of the *Alpha /Beta* ratio, it is possible to observe the evolution of fatigue/tiredness over time. Theoretically an increase in fatigue levels over the trial time is expected, and the consequent tendency is to increase the *Alpha/Beta* values ratio.

Table 3

Values of the amplitude averages for the different waves over time in each hemisphere

	<i>Alfa O1</i>	<i>Alfa O2</i>	<i>Beta O1</i>	<i>Beta O2</i>	<i>AlfaO1/BetaO1</i>	<i>AlfaO2/Beta O2</i>
18°C40%RH	10,8 µV	14,6 µV	4,9 µV	11,5 µV	2,1 µV	0,6 µV
35°C80%RH	7,4 µV	16,7 µV	1,5 µV	12,5 µV	1,5 µV	2,1 µV

Source: The authors

However, this expectation was only applied to the occipital 2 for 35°C - 80% RH (Fig. 12) and very dimly at the end of the test in the cold thermal environment in occipital 1 (Fig. 13).

Interpreting *Alpha* and *Beta* wave activity still lacks scientific unanimity. It is commonplace that different studies reveal completely diverse results and interpretations. Some authors report that Alpha waves show changes at different temperature [21], and others report that alpha and beta increase at either 17°C or 28°C [4].

A comparison of both hemispheres can be seen in Table 3; when the average amplitude of *Alpha* waves increases, there is a decrease in the amplitude of *Beta* waves. This concurs with Eoh et al.'s results [22].

4. Conclusions

In hot thermal environment *Alpha* activity tends to decrease over the course of the task/test, for both areas of the occipital lobe. In the cold thermal environment, *Alpha* activity decreases for half the time of the test (30 min) and then rises, showing an increase for both areas of the occipital lobe. This means that a bigger effort was made.

Beta activity for the hot thermal environment is quite small on the left side of the occipital lobe (Fig. 10 and Table 3); a downward activity is shown as the task was being developed. In tests undertaken in a cold thermal environment, *Beta* activity decreases for approximately half of the test time (30 min), and then it tends to increase at the end of the trials.

The results support the idea that the different areas of the brain, in this case the left and right occipital lobes, are affected differently by the same environmental conditions.

Acknowledgements

The authors acknowledge the Master's program in Occupational Safety and Hygiene Engineering at the Faculty of Engineering, University of Porto, and the Research Laboratory on Prevention of Occupational and Environmental Risks (PROA) for their financial support.

References

- [1] Gaoua, N., Racinais, S., Grantham, J. and El Massioui, F., Alterations in cognitive performance during passive hyperthermia are task dependent. *International Journal of Hyperthermia*, 27(1), pp 1-9. 2011. DOI: 10.3109/02656736.2010.516305
- [2] Hashiguchi, N., Feng, Y., and Tochiara, Y., Gender differences in thermal comfort and mental performance at different vertical air temperatures. *European Journal of Applied Physiology*, 109(1), pp 41-48, 2010. DOI: 10.1007/s00421-009-1158-7
- [3] Lan, L., Lian, Z., Pan, L. and Ye, Q., Neurobehavioral approach for evaluation of office workers' productivity: The effects of room

- temperature. *Building and Environment*, 44(8), pp. 1578-1588 2009. DOI: 10.1016/j.buildenv.2008.10.004
- [4] Lan, L., Lian, Z. and Pan, L., The effects of air temperature on office workers' wellbeing, workload and productivity-evaluated with subjective ratings. *Appl Ergon*, 42(1), pp. 29-36 2010. DOI: 10.1016/j.apergo.2010.04.003
- [5] Racinais, S., Gaoua, N. and Grantham, J., Hyperthermia impairs short-term memory and peripheral motor drive transmission. *J Physiol*, 586, pp. 4751-4762, 2008. DOI: 10.1113/jphysiol.2008.157420
- [6] Pilcher, J.J., Nadler, E. and Busch, C., Effects of hot and cold temperature exposure on performance: a meta-analytic review. *Ergonomics*, 45, pp.682-698, 2002. DOI: 10.1080/00140130210158419
- [7] Gaoua, N., Cognitive function in hot environments: A question of methodology. *Scand J Med Sci Sports*, 20, pp. 60-70, 2010. DOI: 10.1111/j.1600-0838.2010.01210.x
- [8] Guedes, J.C., Costa, E.Q. and Baptista, J.S., EAT 2012 Book of Proceedings – Appendix 1 of Thermology International pp. 22-23. 2012.
- [9] Costa, E.Q., Baptista, J.S. and Tato, M.D., Effects of thermal environment on cognitive response in sedentary activities. A revision. *Faculdade de Engenharia da Universidade do Porto, Portugal*, 2012.
- [10] Costa, E., Correlação entre sinais das ondas cerebrais e a carga cognitiva em ambientes térmicos extremos controlados, MSc. Thesis, Departamento de Segurança e Higiene Ocupacionais, Universidade do Porto, Porto, Portugal, 2014.
- [11] Lal, S.K.L. and Craig, A., A critical review of the psychophysiology of driver fatigue, *Biological Psychology* 55, pp. 173-194, 2001. DOI: 10.1016/S0301-0511(00)00085-5
- [12] Fisch, B.J., Fisch and Spehlmann's EEG primer: Basic principles of digital and analog EEG. 3rd ed., Elsevier, USA, 2000.
- [13] Grandjean, E., Fitting the task to the man. Taylor and Francis, London, UK, 1988.
- [14] Sheer, D.E., A working cognitive model of attention to fit in the brain and in the clinic. Academic Press, USA, 1988.
- [15] Costa, E.Q. and Baptista, J.S., Thermal environment and cognitive performance: Parameters and equipment, in: Arezes et al., eds. *Occupational Safety and Hygiene –Taylor & Francis Group*, London, UK, 2013. DOI: 0.1201/b14391-55
- [16] Costa, EQ, Santos-Baptista, J., Couto1, D., Torres-Costa, J., Diogo, M.T., Almeida, P., Chaves, P., Carvalho, J. and Ponce L., Test protocol validation for thermal environment and cognitive performance. IRFMS, 2013.
- [17] Costa, E.T., Costa, E.Q. and Baptista, J.S., Changes in the brain waves' sign with thermal environment, in: Arezes et al., eds., *Occupational Safety and Hygiene –Taylor & Francis Group*, London, UK, 2015.
- [18] http://www.secondpicture.com/tutorials/3d/3d_modeling_of_a_human_head_3ds_max_01.html
- [19] Costa, E.Q. and Baptista J.S., Thermal environment and cognitive performance: Parameters and Equipment, n: *Occupational Safety and Hygiene*. CRC Press, pp.267-272, 2013.
- [20] Mueller, S.T., The PEBL Manual programming and usage guide for the Psychology Experiment Building Language PEBL Version 0.14. [Online]. ISBN 978-0-557-65817-6. Available at: <http://pebl.sourceforge.net>.
- [21] Yao, Y., Lian, Z., Liu, W. and Shen, Q., Experimental study on physiological responses and thermal comfort under various ambient temperatures. *Physiology & Behavior*, 93(1-2), pp. 310-321. 2008. DOI: 10.1016/j.physbeh.2007.09.012
- [22] Eoh, H.J., Chung, M.K. and Kim, S.-H., Electroencephalographic study of drowsiness in simulated driving with sleep deprivation. *International Journal of Industrial Ergonomics*, pp. 35(4), pp. 307-320, 2005.

E. Tiago-Costa, has two MSc. degrees, one in Environmental Engineering and the other in Occupational Safety and Hygiene Engineering, both from the Faculty of Engineering at the University of Porto, Portugal. He is currently a researcher in the Faculty of Engineering of the University of Porto; Portugal.
ORCID: 0000-0002-4996-1668

E. Quelhas-Costa, is a Chemical Engineer and graduated from the Porto Superior Institute of Engineering. She has a MSc. in Occupational Safety and Hygiene Engineering from the Faculty of Engineering at the University of Porto, and a PhD in Occupational Safety and Health from U. Porto. She is currently a researcher at the Research Laboratory on Prevention of Occupational and Environmental Risks (PROA).
ORCID: 0000-0002-3122-1127

J. Santos-Baptista, has a PhD in Mining Engineering from U. Porto. He is currently an associate professor and Director of the MSc. in Occupational Safety and Hygiene Engineering at the Faculty of Engineering at the University of Porto, and the coordinator of the engineering PhD program Safety and Health. He also leads the Research Laboratory on Prevention of Occupational and Environmental Risks (PROA).
ORCID: 0000-0002-8524-5503



UNIVERSIDAD NACIONAL DE COLOMBIA

SEDE MEDELLÍN
FACULTAD DE MINAS

Área Curricular de Medio Ambiente

Oferta de Posgrados

Especialización en Aprovechamiento de
Recursos Hidráulicos

Especialización en Gestión Ambiental

Maestría en Ingeniería Recursos Hidráulicos

Maestría en Medio Ambiente y Desarrollo

Doctorado en Ingeniería - Recursos Hidráulicos

Doctorado Interinstitucional en Ciencias del Mar

Mayor información:

E-mail: acia_med@unal.edu.co
Teléfono: (57-4) 425 5105

Effects of joints and their waterstops on pressures spread over a slab subject to turbulent flow on a horizontal apron

Mauricio González-Betancourt ^a & Lilian Posada-García ^b

^a Grupo del Posgrado en Aprovechamiento de Recursos Hidráulicos, Universidad Nacional de Colombia. Investigador financiado por el Fondo Francisco Jose de Caldas (COLCIENCIAS) y SENA, Colombia. magonzalezb@sena.edu.co

^b Facultad de Minas, Universidad Nacional de Colombia, Colombia. lilianposadagarcia3@gmail.com

Received: November 29th, 2014. Received in revised form: July 14th, 2015. Accepted: January 25th, 2016.

Abstract

Analyses of measured pressures using multiple sensors above and below a slab subject to supercritical flow and hydraulic jump are presented. Pressures were measured in a physical model constructed in the Hydraulic Laboratory of the Universidad del Valle in Cali, Colombia. Details such as expansion joints, waterstops, slab thickness, and soil-slab separation were simulated. Inflow was full and partially developed, with Froude numbers ranging from 2.84 to 11.6. The role of the expansion joints and waterstops in the generation of uplift pressure and hydraulic gradient below the slab were identified using pressure fields above and below the slab, in multiple tests. The results of this research are critical for the design of bottom slabs in hydraulic structures, which are required to withstand uplift forces and prevent internal erosion problems.

Keywords: uplift pressure, hydrodynamics uplift, slabs, hydraulic model, stilling basins.

Efectos de las juntas y sus sellos en las presiones distribuidas en una losa sujeta a flujo turbulento en un canal horizontal

Resumen

Se presentan los análisis de las presiones medidas con múltiples sensores sobre y debajo de una losa bajo flujo supercrítico y salto hidráulico. Las presiones fueron medidas en un modelo físico construido en el laboratorio de Hidráulica de la Universidad del Valle en Cali, Colombia. Se simuló detalles como las juntas de expansión, los sellos, el espesor de la losa y la separación suelo-losa. El flujo incidente fue total y parcialmente desarrollado con números de Froude entre 2,84 y 11,6. A través de los campos de presión sobre y debajo de la losa de múltiples pruebas, se identificó el papel de las juntas de dilatación y sellos en la generación de presiones de levantamiento y el gradiente hidráulico debajo de la losa. Los resultados de esta investigación son fundamentales para el diseño de losas de fondo en las estructuras hidráulicas que deben soportar las fuerzas de levantamiento y prevenir problemas de erosión interna.

Palabras clave: Presiones de levantamiento, levantamiento hidrodinámico, losas, modelo hidráulico, tanques de amortiguación.

1. Introduction

In interaction with turbulent flow, cavitation, abrasion, uplift force and internal erosion may singly, or jointly, cause the failure of floor slabs in hydraulic structures. The first two processes attack the slab on the face that is exposed to flow. The impact of vapor bubbles and sediments causes random damage to waterstops, allowing static and dynamic pressures to propagate

below the slab through cracks or joints. This propagation creates a pressure differential between the top and underside of the slab, which in turn produces uplift force. Other effects of pressure propagation are that the hydraulic gradient below the slab generates internal erosion that manifests through the migration of soil particles and the consequent creation of foundation voids. The fluctuation of uplift pressure and internal erosion induces vibrational movement and offset between slabs. These

How to cite: González-Betancourt, M. and Posada-García, L., Effects of joints and their waterstops on pressures spread over a slab subject to turbulent flow on a horizontal apron. DYNA 83(197), pp. 94-103, 2016.

combined factors bring the slab to failure by causing fatigue of the anchor [1], stagnation pressures [2] and displacement of the slab by the flow.

Uplift pressures has attracted the attention of engineers since the early 1960s, because hydraulic structures subjected to supercritical flow (SF) or to hydraulic jump (HJ) have failed with a flow rate much lower than the design flow rate. For example, damage in spillways, stilling basins or discharge channels was reported in the Big Sandy, Dickinson (USA; [3]), Tarbela (Pakistan; [4]), Karnafulli (Bangladesh; [5]) and Malpaso dams (Mexico; [6]). Other cases of vulnerability to failure were identified in the Scofield (USA; [7]), Santa Elena (Brazil; [8]), Bhakra (India; [9]), Liu-Xia Jia and Qiang Wu-xi dams (China; [10]), cases in which the importance and relevance of uplift force eventually became apparent as a structural design problem.

The role of uplift force in HJ has been studied by taking force or pressure measurements in physical models in order to determine the expected uplift force on the prototype, using Froude scaling relationships ([11-16]). However, as a result of the complexity involved in studying uplift force phenomena, up to now the criteria have failed to take into account many of the variables that affect the uplift pressure and slab thicknesses into the physical and conceptual models used. One influencing factor was that prototype parts were not reproduced or simulated in their true scale in the models; this includes expansion joints, waterstops, slab thickness (s) and slab-soil separation (δ). Moreover, in some cases the study of the uplift force below the slab has been limited to taking measurements using pressure taps at the floor of the flume ([11-13, 15]) or force sensors in the central part, below a thin plate which simulates the slab ([14, 16]). These cases do not take into consideration the effects of friction in the joint and below the slab. The flaws inherent in these conceptual models make it difficult to choose criteria that guarantee the stability of the slab at lowest cost. Consequently, designers generally choose the most conservative criteria, which in turn increases costs.

In recognition of these limitations in previous studies, investigations were initiated at the Fluids and Hydraulics Laboratory of the Universidad del Valle, Colombia, in order to verify del Risco's hypothesis [17], which deals with the conversion of kinetic energy into dynamic pressure in expansion joints. For this study a physical model of a slab with multiple sensors and joints was designed [18], and preliminary explorations of the mean pressure below the slab with undeveloped inflow were performed [18-20]. Specifically, with SF, it was considered essential to analyze the effect of transverse and longitudinal joints separately. The results showed that with SF the measured pressures were higher than the static pressures below the slab, which corroborated the conversion of kinetic energy into dynamic pressure in the joint. Moreover, a pattern of pressure in flow direction below the slab was identified: increasing with open longitudinal joints [20] and decreasing with the open transverse joints [19]. It was also found that δ inversely affects the value for pressure below the slab [19, 20].

Building on the findings of previous studies, this paper provides a more detailed analysis that presents the pressures measured using multiple sensors above and below a simulated slab with SF and HJ. The pressures spread over a

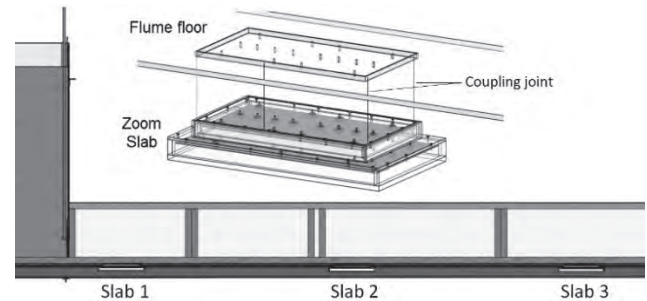


Figure 1. Physical model.

Source: The authors

slab were analyzed to find the effect of joints and their waterstops on uplift pressure and the hydraulic gradient below the slab. The work was characterized by: 1) analyzing pressure fields spatially in the time domain, 2) situating sensors below the slab and the joints in order to achieve a higher resolution of the pressure field, 3) considering multiple tests including physical and hydrodynamic variations, 4) identifying the critical hydraulic gradient below the slab to help in assessing internal erosion risks.

2. Materials and methods

The study was conducted in a horizontal flume (0.5 m high, 8m long and 0.35 m wide) at the Universidad del Valle, using a model that simulated a floor slab and joints. The expansion joints were simulated using an array of acrylic boxes with dimensions from largest to smallest (Fig. 1).

In the model, pressure was measured using 32 circular pressure taps (S1-S32), 2 mm in diameter, above and below the slab. There were two distributions of sensors. The first (D1) selected 8 pressure taps located above the slab (Fig. 2a) and 24 pressure taps below it, as well as the floor of the joints (Fig. 2b). The second (D2) selected 16 pressure taps above and below the slab with equal distribution (Fig. 2c). D1 located sensors below the slab and the joints so that a higher resolution for the pressure field could be achieved.

The experimental design included three slabs that were placed in the flume at different distances from the load tank, in order to vary the state of development of the boundary layer of the inflow. Thus, slab 1 (S1) was located at a distance in which Reynolds number in the boundary layer (Re_x) was between 300,000 and 660,000 (partially developed flow), slab 2 (S2) was located in the area where Re_x was between 4,150,000 and 9,130,000, while slab 3 (S3) was the farthest from the load tank, with Re_x between 7,900,000 and 17,380,000.

The length (L) of individual slabs ranged between 6 and 12 times the depth of the inflow (y_1); slab width (B) was approximately half L . Geometric variations in the model of δ and joint width (ϵ) were also performed (Table 1), in order to consider their influence on uplift pressure and hydraulic gradient below the slab. In the prototype, these were changed by rearranging the slabs, which could be linked to internal erosion or other sources of natural movement. In the model, δ could be adjusted by interposing aluminium sheet rings of 1 mm diameter with the thickness required to achieve the desired separation between slab and background. Joint width was changed by varying the size of the acrylic boxes.

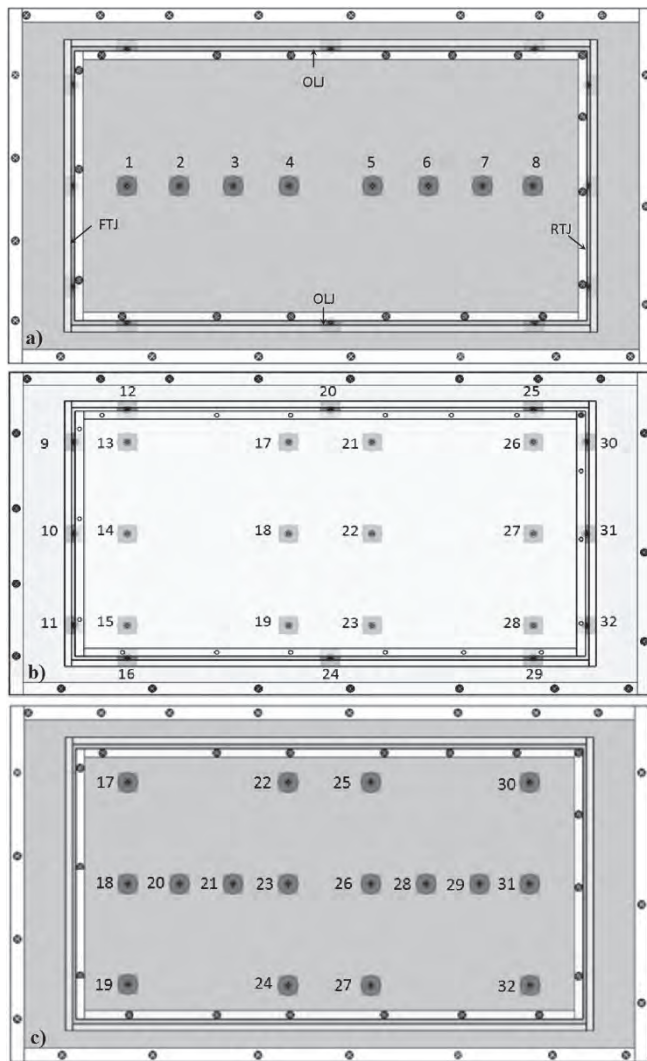


Figure 2 Distribution of expansion joints and sensors. a) D1, sensors above the slab; b) D1, sensors below the slab and expansion joints. c) D2, sensors above and below the slab.

Source: The authors

Table 1
Slab type, dimensions, sensor distribution (D) and tests.

Slab type	L (mm)	B (mm)	s (mm)	ϵ (mm)	δ (mm)	D (#)	Tests HJ (#)	Tests SF (#)
S1A	280	150	30	2	0, 0.2, 0.5, 1	D2	238	124
S1B	294	156	34	0.5	0.5	D1	70	86
S2	299	160	37	1	1	D1	56	65
S3	299	160	38	1.5	1	D1	56	65

Source: The authors

Pressure was measured using Motorola sensors (MPXV 4006GC7U, range 0-6 kPa and accuracy $\pm 5\%$) in a range of different tests, which included physical and hydrodynamic variations. In each test, the slab type (S1A, S1B, S2 and S3; Table 1), number of open joint(s), δ , flow type (SF or HJ) and Froude number of the inflow (Fr_1), was selected. The open joint(s) in S1B, S2 and S3 were simulated by detaching the waterstop(s) in: a) one of the four joints, that is, drawn from either the front transverse joint (FTJ), the rear transverse joint

(RTJ) or the longitudinal joint (OLJ); b) any two joints simultaneously, that is, drawn from either the longitudinal joints (LJ) or the transverse (TJ); c) in all joints (AJ). In S1A TJ, LJ or AJ were open. Hydrodynamic variations included a minimum seven different Fr_1 between 2.84 and 11.6 for each physical variation.

The discharge was regulated to between 8.15 and 14.1 Gallon/min and vertical gate of the load tank (1.8m high, 1m long, 0.35 m wide) was regulated to between 2.5 and 5.5 cm. The velocities of the inflow (V_1) ranged between 1.65 and 5.76 m/s with a turbulent flow regime (Reynolds numbers of 90,000 to 200,000). The flow rate was measured with a Prandtl tube at a point located 0.05 m upstream of the FTJ, with an accuracy of 4%. Discharge was measured using an Omega brand flow meter (FMG-901). Rectangular weirs of heights ranging from 5 to 20 cm placed in the end of the flume were used to generate HJ. Some tests of S1 were formed submerged hydraulic jump (SHJ) and others a free hydraulic jump (FHJ). Tests with FHJ were located with 30% of their length on the slab, because in general the highest pressure fluctuations are reported in the first third of the length of the HJ [13,15,21-27].

The problem of alignment, acknowledged by some researchers [16,28-30], was controlled by maintaining a static slab in the model without the possibility of failure or displacement. Offset between slabs was avoided since that would have increased the uplift pressure below the slab by stagnation points [2,3]. Thus, the slab was part of the flume floor, which was drilled and slotted to provide continuity to pressure taps and joints. The coupling between elements of the system was monitored and the offset was of the order 10^{-5} m. In the prototype, this might have occurred as a result of imperfections in the finishes of the slabs and because of natural movement.

Flow depth was measured with depth gages of range 300 mm, accuracy of ± 0.2 mm on a) FTJ, b) LJ in the middle of the slab; c) RTJ. At each point the minimum and maximum depths detected in 30 second were measured. The average static pressure below the slab (p_e) was calculated as the average of the pressure equivalent of three minimum depths measured on FTJ, LJ, RTJ with baseline below the slab.

The sensors were calibrated to relate their voltage signs with the pressures, in accordance with the reference [31]. The acquired signs with a data acquisition system (DAQ National Instruments, NI SCXI: 1000, 1102B, 1600, 1300) were sent to a laptop. The sampling frequency (f_s) was 200 Hz limited by the data acquisition system available; in addition, this complied with the sampling theorem (avoiding aliasing) and improves the resolution in the time of the digitized signal (5 ms). Following digitization and analysis of the signal, a digital filter was used to remove frequency components lying outside the phenomenon. Responding to the analysis of the frequency signal and the dynamic characteristics of the pressure measurement system, the cutoff frequency of the digital filter was selected at 10.58 Hz.

Macroturbulent natural phenomena do not display periodicity, so each segment of data acquired makes its own unique contribution. Thus, the degree of accuracy is proportional along the length of test run [32]. The minimum test run time stipulated by authors like Lopardo and Henning is one minute [33]. Fiorotto and Rinaldo [25] and Toso and

Bowers [13], however, showed how the pressure coefficient increased over time. In the case reported here, the length of the test run was 15 minutes, data acquisition being performed during the last five minutes. The duration of the data acquisition time was associated principally with extensive data and the number of tests explored (760; Table 1). One test had 60,000 pressure fields.

3. Results and analysis

Figs. 3 and 4 represent 3D snapshots of the measured pressure fields below the slab with maximum uplift pressure for SF (Fig. 3) and HJ (Fig. 4), respectively, where, " P_i/p_e " is the relationship between the measured pressure (P_i ; static and dynamic pressure) and the average static pressure (z axis, Figs. 3 and 4). " $P_{h_{min}}/p_e$ " and " $P_{h_{max}}/p_e$ " represents the relationship between the pressure equivalent to minimum and maximum depth measured at FTJ, LJ or RTJ and p_e (z axis, Figs. 3 and 4). The pressures were calculated with a baseline that lay below the slab. The positions below the slab are represented on the x axis (length) and y axis (width), which are dimensionless with length (L).

For SF (Fig. 3) a linear correlation of pressure in the direction of flow was observed when waterstops were detached from joints. The negative gradient (Fig. 3c) was favored with TJ or AJ, while the positive gradient (Fig. 3d) was favored with LJ. When there was only one open joint (FTJ, OLJ, RTJ) uniform

pressure below the slab was observed (Figs. 3a and 3b). FTJ amplified the static pressure (Fig. 3a) and RTJ decreases it (Fig. 3b).

Under HJ (Fig. 4), the positive pressure correlation is given by the flow; however pressure below the slab can vary depending on the detached waterstops. During tests it was found that, with only one waterstop detached on FTJ, the pressure generated below the slab was uniform (Fig. 4b), while with two open transverse joints (TJ) the positive gradient (Fig. 4a, Fig. 4d) might become altered, in some instances leading to negative pressure gradients (Fig. 4c).

The pressures measured simultaneously above and below the slab with HJ and with detached waterstops on joints, were analyzed. It was observed that the joints acted as a filter of the pressures generated in the flume. Pressure only propagated below the slab at those points located on joints without waterstops. The pressures in direction of flow in one snapshot, and in four other tests, are shown in Fig. 5. It is apparent that some pressure fluctuations above the slab (dotted line) were not perceived below it (continuous line). Furthermore, pressure was not uniform above and below the slab.

In the tests with HJ, open FTJ and open TJ, a travel time delay was found between the entry of the pressure waves at the joint and their arrival below the slab (Δt). This led to pressure differential (Δp) above and below the slab, which generated high uplift force. To make this evident, in Fig. 6,

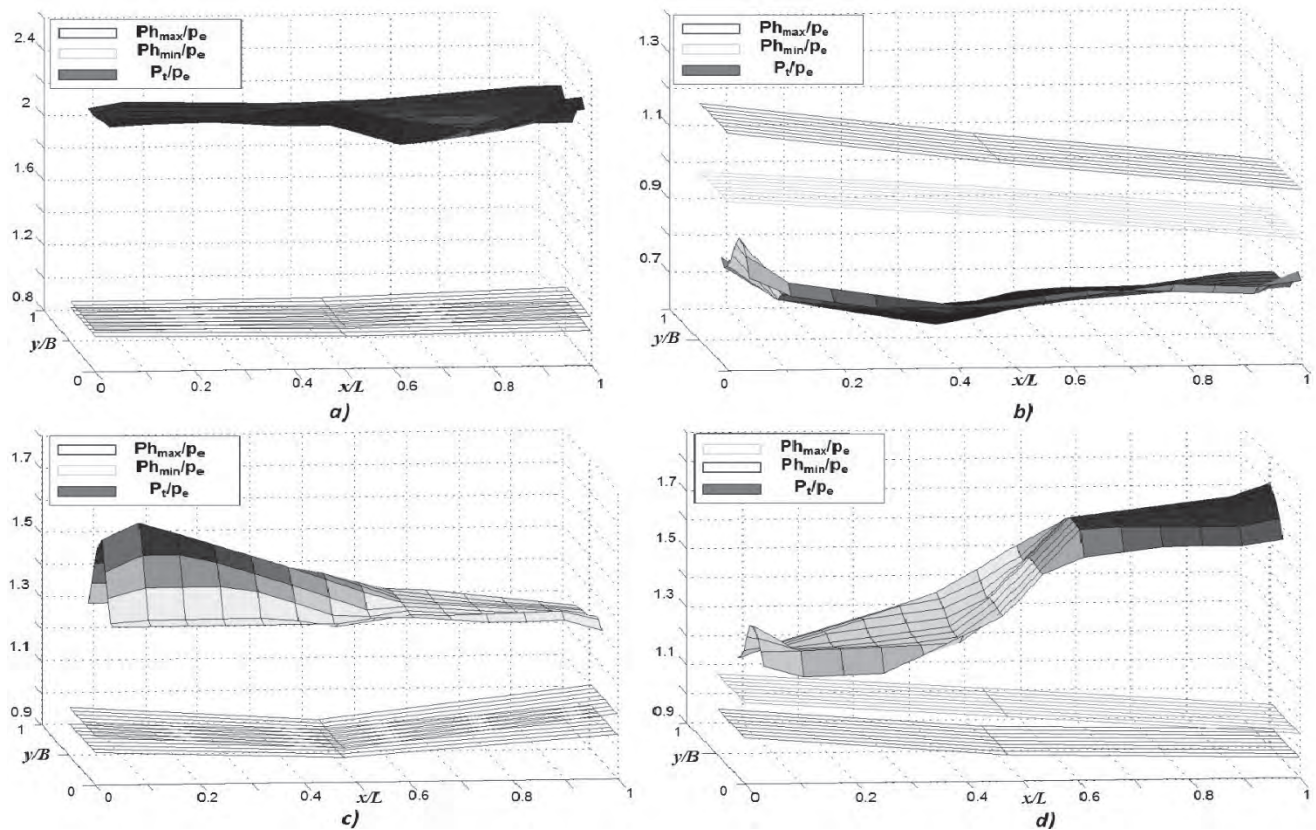


Figure 3 Measured pressure fields below the slab with maximum uplift pressure for SF and detached waterstops on: a) FTJ; test with Fr_1 of 11.5 and S1B. b) RTJ; test with Fr_1 of 8.28 and S3. c) TJ; test with Fr_1 of 8.84 and S2. d) LJ; test with Fr_1 of 9.66 and S2.

Source: The authors

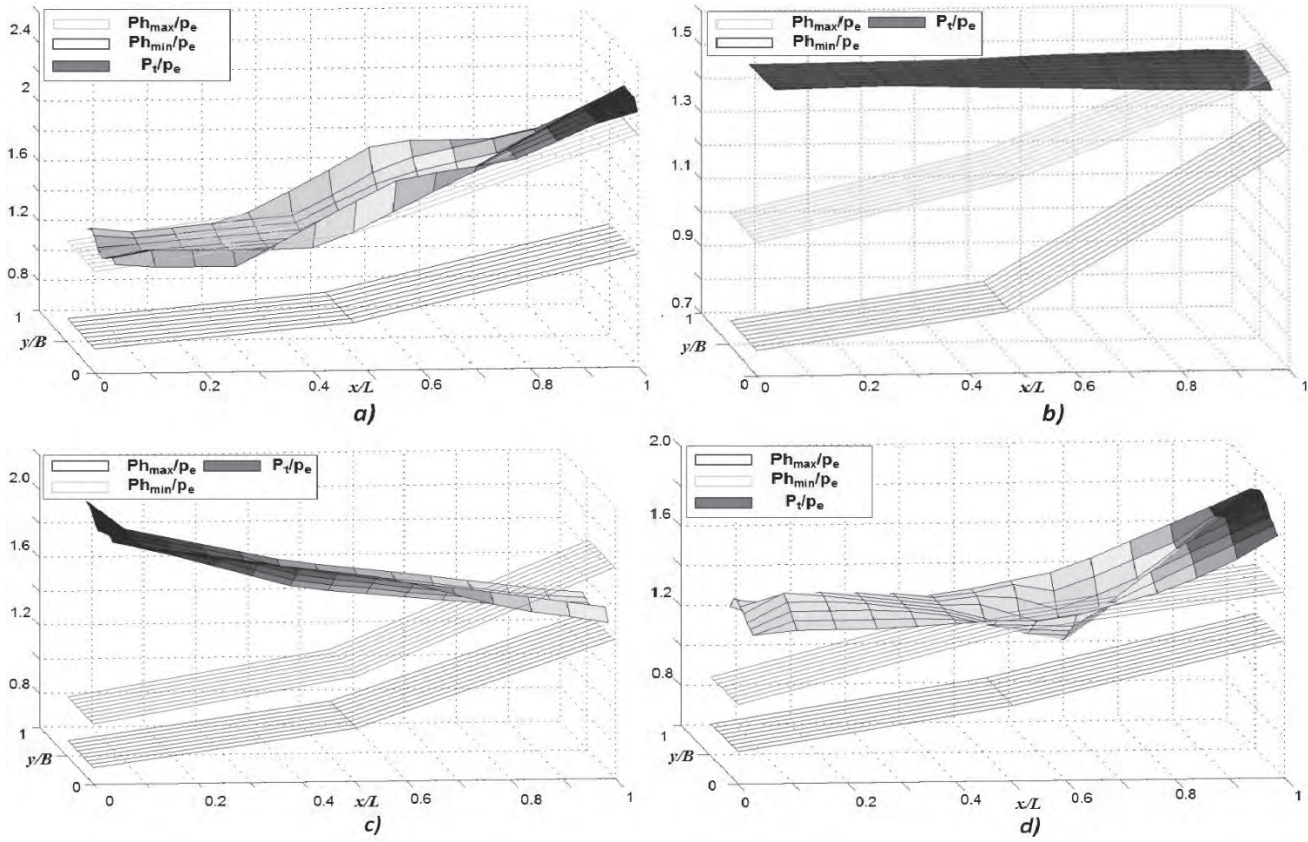


Figure 4 Measured pressure fields below the slab with maximum uplift pressure for HJ and detached waterstops on: a) LJ; test with Fr_1 of 8.63 and S3. b) RTJ; test with Fr_1 of 9.82 and S2. c) AJ; test with Fr_1 of 6.8 and S1B d) TJ; test with Fr_1 of 5.51 and S1B.

Source: The authors

two tests are presented to show the unfiltered signal from pressure sensors close to the open transversal joints above and below the slab. The time delay ranged from 69 to 100ms between the arrival of the pressure pulse at a sensor above the slab (S1) and the one below it (S10, S11, S28, S31). Interchannel delay between S1 and S32 of data acquisition system was estimated at 0.093 ms. This was depreciable against the sampling period (5ms). With open FTJ, a time delay smaller than 5 ms was observed between the unfiltered signals of the sensors below the slab at opposite ends (S10 and S28).

The role of the joints as frequency filters was observed in the comparative analysis of the spectrum of fluctuating pressures between the sensors located above (S2, S6) and below (S9-S32) the slab with SF and HJ (Fig. 7).

The sensors below the slab measured a significant drop in frequency amplitude ($|Y(f)|$) on frequencies higher than 6 Hz (Fig. 7). The frequency cutoff of the pressure fluctuation in the flume with HJ has been defined by some authors as 25 Hz [5, 13, and 24]; however this only applies above the slab, because the cutoff frequency would be less than 10 Hz below it. Furthermore, it was observed with SF that joint-flow interaction may increase $|Y(f)|$ for frequencies lower than 4 Hz (Fig. 7d), when measured below the slab.

The system of irregular cavities integrated by the joints, waterstops, and δ may be understood as a labyrinth of networked cavities with unpredictable passages, reductions, changes of direction, and dead ends [34]. Thus, when the

pressure wave enters the network of cavities through an open joint or missing waterstop, it spreads below the slab and experiences diffraction, interference, reflection and refraction that may alter its propagation velocity, amplitude and the transmitted energy. Furthermore, waterstops, joints and flow determine the configuration of vortices within the joint and in the δ , which vary the pressure that is transmitted below the slab.

In HJ, the pressure pulses on the slab were uncorrelated since the slab length is larger than the integral scale of the pressure fluctuations. The integral scale was calculated by Fiorotto and Rinaldo [25] as ranging from $0.4y_1$ to $1.5y_1$. The results of the current study showed that extreme pulses recorded by a single sensor in the large slab do not provide a representative sample of the pressure fields above and below the slab and cannot, therefore, be used to calculate the uplift force accurately. Above the large slab, a positive or negative pulse is a local effect. The compensation of the pressure pulses on the upper face of the slab was recognized by Bellin and Fiorotto [16].

After studying the pressure fields with different open joints, it was shown that pressure distribution was not uniform below the slab. In other words, a hydraulic gradient was formed.

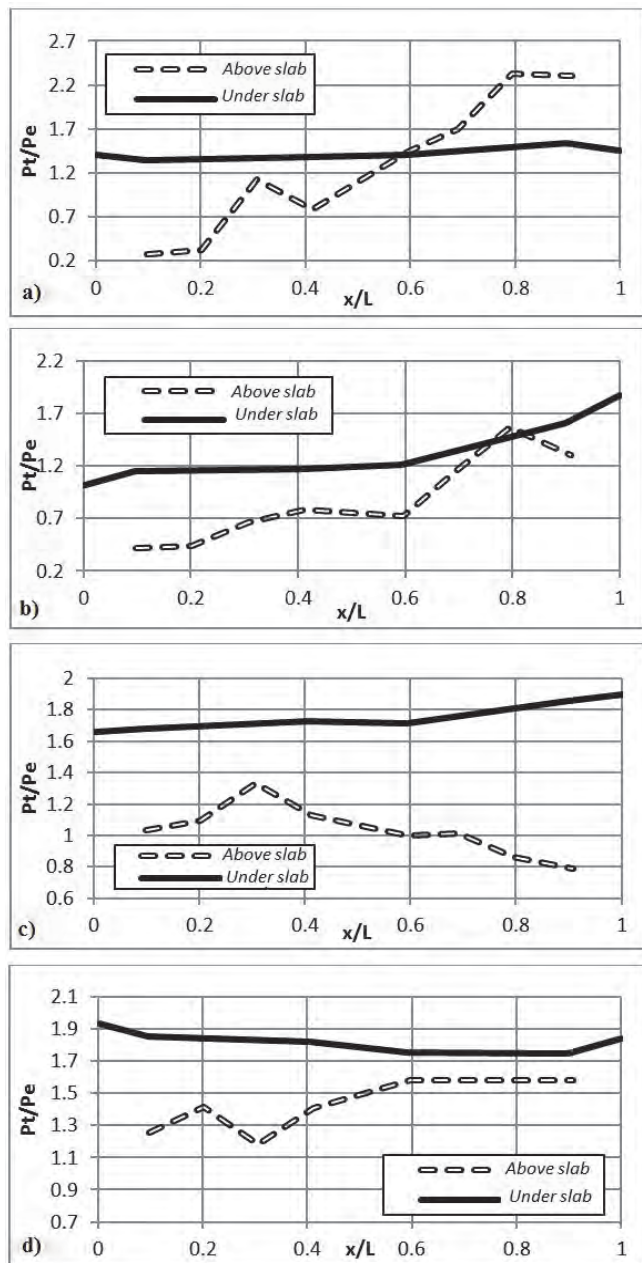


Figure 5 Pressure in longitudinal direction above and below the slab in snapshots. Test with: a) Open FTJ, Fr_1 of 4.56 and S1B. b) Open OLJ, Fr_1 of 8.71 and S3. c) Open LJ, Fr_1 of 9.17 and S2. d) Open TJ, Fr_1 of 5.85 and S2. Source: The authors

The hydraulic gradient below the slab generates a drag force on soil particles at δ . These particles dislodge from the soil and move in the direction of flow below the slab when the drag force is greater than the interlocking of the soil particles [35,36]. In the tests this dislodgement of particles and one unfiltered exit generated piping along the conduit and a loss of support for the slab [37].

The foundation materials may be eroded into drains and unfiltered systems over long periods of time before they are detected. Therefore, the slab and foundation materials must be designed for a critical hydraulic gradient. Unfortunately at

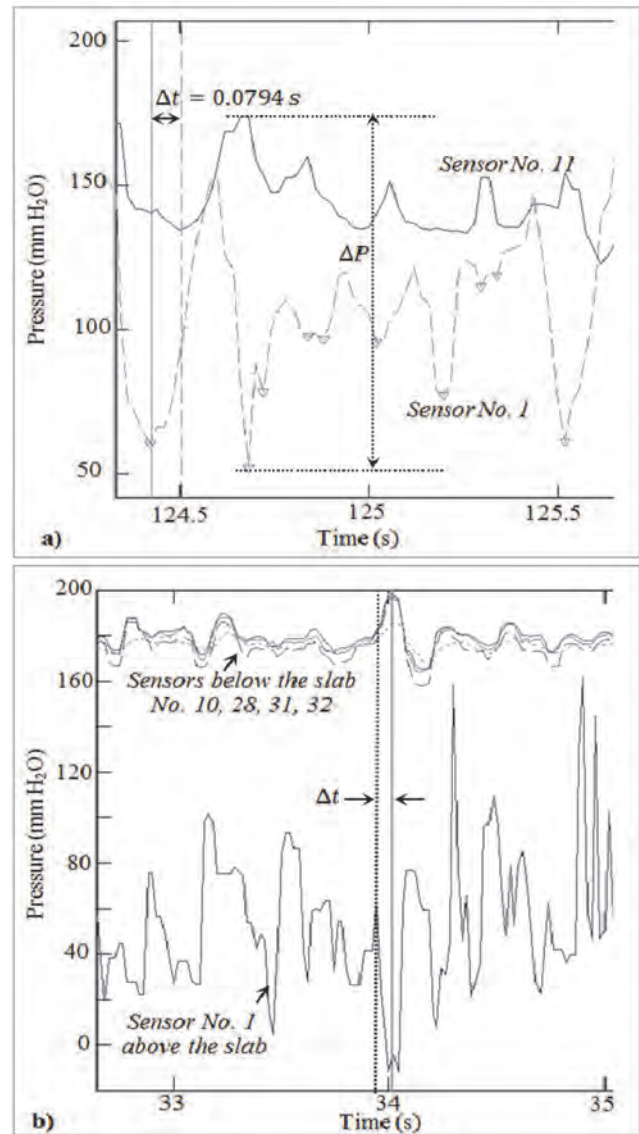


Figure 6 Pressure signal in the time of sensors closest to the joints above and below slab. Test with a) Open TJ, Fr_1 of 8.49, S1B and SHJ. b) Open FTJ, Fr_1 of 6.9, S1B and SHJ.

Source: The authors

this moment, internal erosion is a potential failure mode that cannot be completely analyzed using numerical formulae or models [35]. However, valuable information on dam and soil behavior, and on experimental hydraulic gradients can be obtained to help assess internal erosion risks.

Figs. 8, 9, and 10 present the critical hydraulic gradients below the horizontal slab in multiple tests. These include SF (Fig. 8) and HJ with undeveloped (Fig. 9) and developed (Fig. 10) inflow conditions. The hydraulic gradient (i) was the slope of the piezometric line between the pressure below the slab in proximity to FTJ and RTJ (the ratio between the pressure gradient below the slab and its length). "S1A" collected tests with four variations of δ , since its influence was not easy to perceive in the hydraulic gradient. As reference, envelope curves on the maximum hydraulic gradients are plotted in the Figs. 8, 9, and 10.

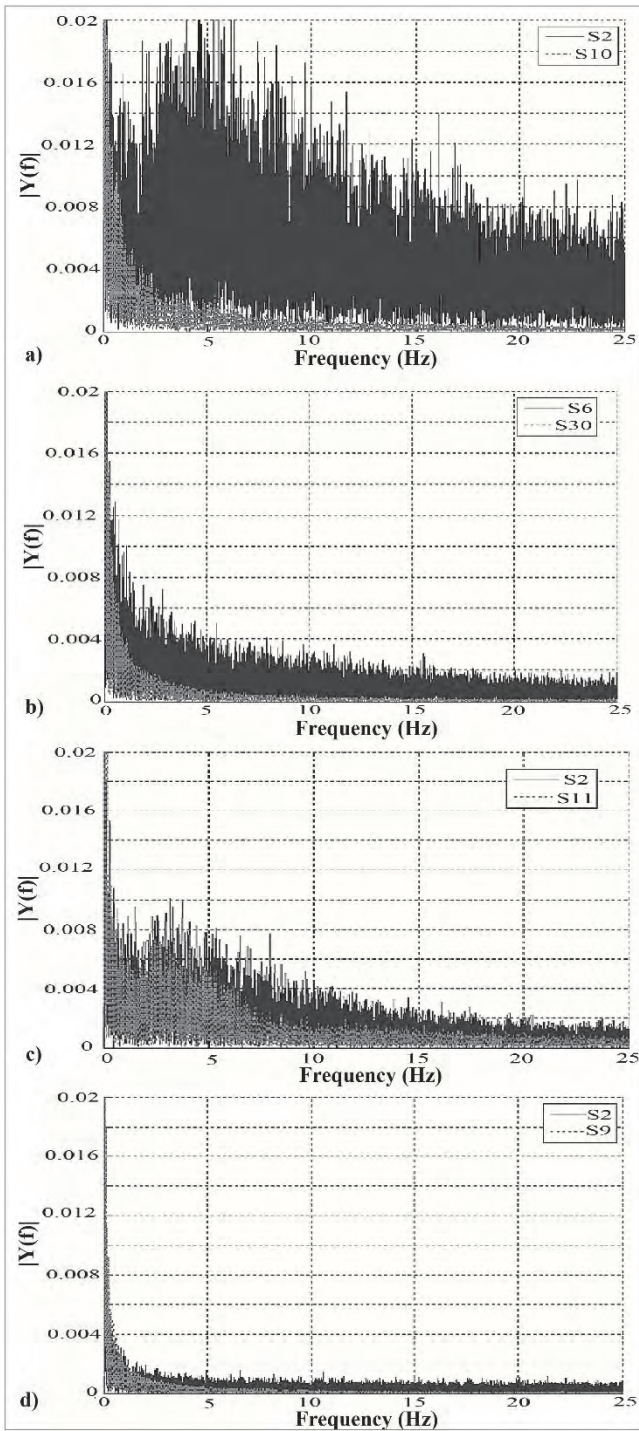


Figure 7 Typical frequency spectrum of the sensors above and below the slab. Test with: a) Open TJ, Fr_1 of 8.48, S1B and SHJ. b) FTJ, HJ, Fr_1 of 5.98, S2. c) Open AJ, Fr_1 of 6.16, S2 and HJ. d) Open TJ, Fr_1 of 11.2, S1B and SF.

Source: The authors

Critical envelope curves in the evaluation of multiple tests were found with the condition of undeveloped inflow to SF (Fig. 8) and HJ (Fig. 9). The most critical envelope curve was found with undeveloped inflow and HJ (Fig. 9). To establish the design of the hydraulic gradient calculated with envelope curves to SF, the position of HJ should be ensured by controls

at the floor of the flume. Slabs installed in sloping flumes should be considered as additional hydraulic gradients established by slab sloping.

The results of the study showed uniform pressurization below the slab with only one open transversal joint. Similar results were found by Melo, Pinheiro and Ramos [38] with slabs exposed to the impact of a jet in the plunge pool. However before considering slab design with open joints as a strategy to reduce uplift pressure with large slabs, the effect of the eccentricity of the uplift force and the likelihood of internal erosion risk should be evaluated. Large slabs are typically used to protect easily erodible materials in the foundation [39]. Therefore, waterstops are required in joints, and these must be maintained if piping, stagnation points, and future rearrangement of the slab are to be avoided.

The results of this study should be incorporated into future design criteria in an effort to optimize costs.

4. Conclusions

In this study it was observed that joints and their waterstops with turbulent flow on a horizontal apron influence the spread of pressure over a slab because:

- Joints and waterstops act as filters of pressure fluctuations generated in the flume. Pressure only propagates below the slab at points located on the joints without waterstops.
- Joints generate a time delay between the entry of the pressure wave at the joint and its arrival below the slab. This leads to a pressure differential between the top and bottom of the slab resulting in the appearance of uplift force.
- The interactions between the joints and the main stream alter the amplitude of the pressure wave below the slab.
- With only one open transversal joint pressure is uniform below the slab.
- With two or more open joints pressure gradients is generated below the slab.

The lack of uniformity in uplift pressure below the slab leads to consider the failure mechanism induced by moment. Therefore, future research is necessary in order to predict net uplift force and its point of application and, as a result, verify instability of the slab by moment of the uplift force.

The hydraulic gradients shown here constitute valuable experimental information that helps to assess internal erosion risks in hydraulic structures.

Acknowledgments

This research project was partially funded by COLCIENCIAS. Special thanks are due to Efraín del Risco for his valuable suggestions and insight in this field. The experiments necessary for the development of this research were carried out at the Hydraulic Laboratory at the University of Valle in Cali, Colombia. The support of the laboratory is sincerely appreciated.

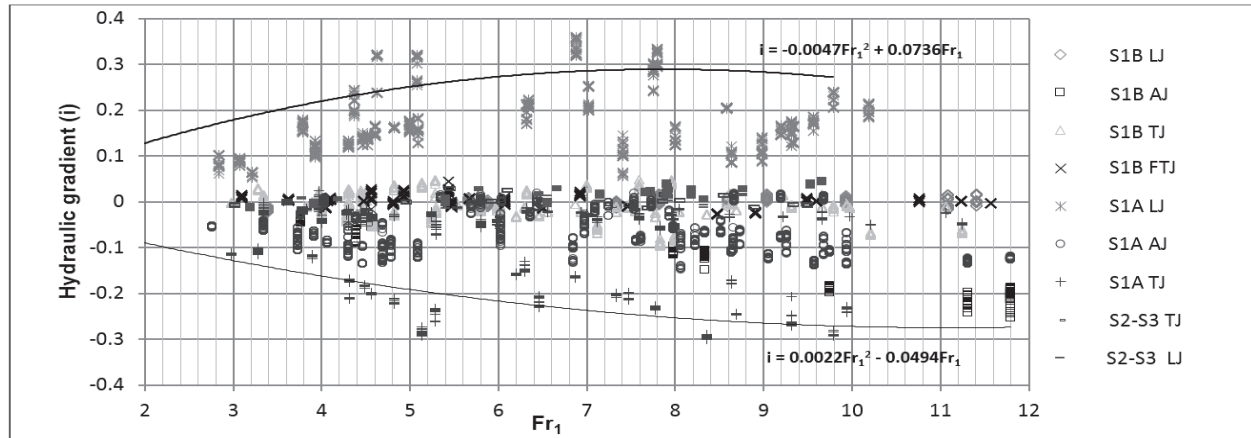


Figure 8 Hydraulic gradients below the slab as a function of Fr_1 with SF.
Source: The authors

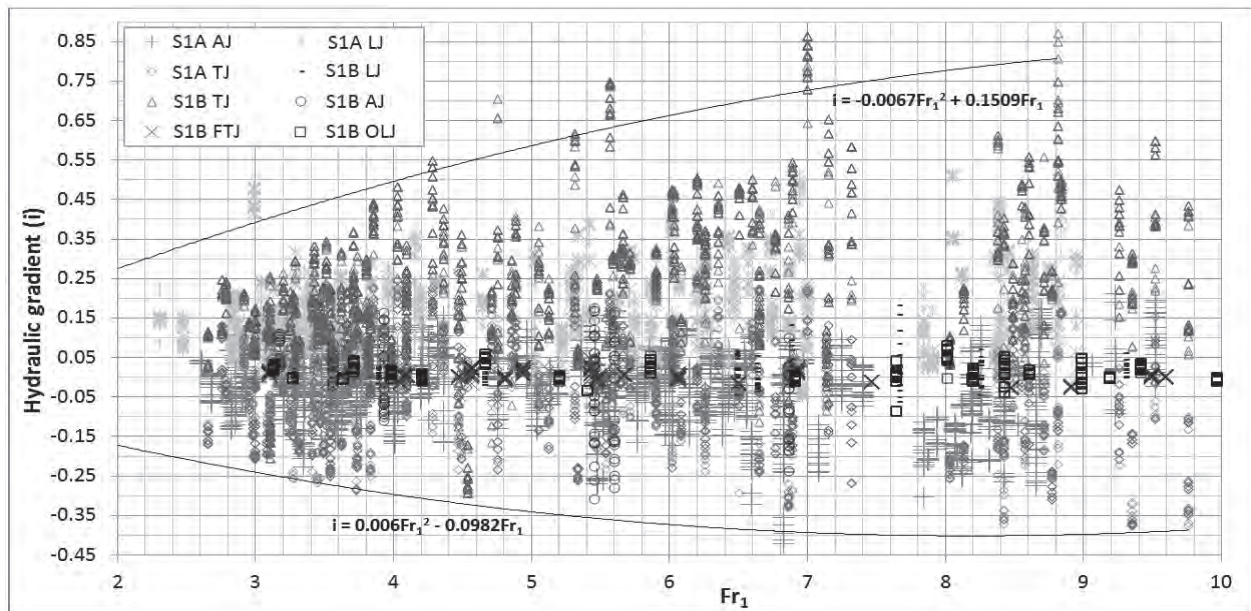


Figure 9 Hydraulic gradients below the slab 1 as a function of Fr_1 with HJ.
Source: The authors

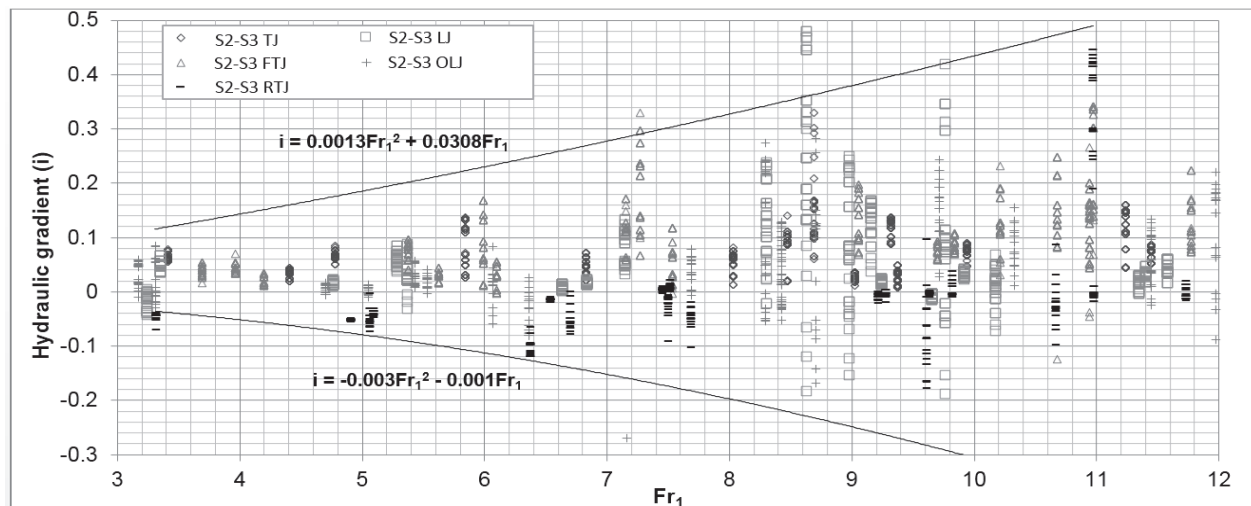


Figure 10 Hydraulic gradients below the slab 2 and 3 as a function of Fr_1 with HJ.
Source: The authors

Nomenclature

FTJ	Front transverse joint
RTJ	Rear transverse joint
OLJ	One longitudinal joint
LJ	Longitudinal joints
TJ	Transverse joints
AJ	All joints
Δ	The slab-soil separation

References

- [1] Khatsuria, R.M., Hydraulics of spillways and energy dissipators. New York: Marcel Dekker, 2005, pp 411-424.
- [2] Frizell, W.K., Uplift and crack flow resulting from high velocity discharges over. Report DSO-07-07 U.S. Department of the interior, Bureau of Reclamation. Denver, Colorado, 2007, pp 2-5.
- [3] Hepler, T.E. and Johnson, P.L., Analysis of spillway failure by uplift pressure. ASCE National Conference, Colorado Springs, Colorado, august 8–12, pp 857-862, 1988.
- [4] VSL International. Soil and rock anchors, Examples from practice. Berne, Switzerland. [Online]. pp 19-20, 1992 [Consulted: 15th of august of 2009]. Available at: www.vsl.net/Portals/0/vsl_techreports/PT_Ground_Anchors.pdf.
- [5] Bowers, C.E. and Tsai, F.Y., Fluctuating pressures in spillway stilling basins. J. Hydraulic Division, ASCE, 95(6), pp. 2071-2079, 1969.
- [6] Bribiescas, S. and Capella, V., Turbulence effects on the lining of stilling basins. Proceedings ICOLD. Congr s des Grands Barrages. Madrid. Q.41. R.83. pp. 1575-1592, 1973.
- [7] Frizell, K.H., Spillway, outlet works, and water conveyance needs—survey results. Hydraulic laboratory report HL-2005-03. U.S. Department of the Interior, Bureau of Reclamation. Denver, Colorado. pp. 10-30. 2005.
- [8] Ivanisovich, M.L., Discussion: Fluctuating uplift and lining design in spillway stilling basins, by Fiorotto, V. and Rinaldo, A., 118(4), April, 1992. Journal of Hydraulic Engineering. 119(4). pp. 526-527, 1993. DOI: 10.1061/(ASCE)0733-9429(1993)119:4(526)
- [9] Ramos, C.M., Models for study of the hydrodynamic actions on hydraulic structures. Recent advances in Hydraulic Modelling. Kluwer Academic Publishers, Netherlands. Series E: Applied Science, 165, pp. 143-189, 1989. DOI: 10.1007/978-94-009-2344-7_4
- [10] Liu, P-Q. and Li, A-H., Model discussion of pressure fluctuations propagation within lining slab joints in stilling basins, Journal of Hydraulic Engineering, 133(6), pp. 618-624, 2007. DOI: 10.1061/(ASCE)0733-9429(2007)133:6(618)
- [11] Bribiescas, S. y Fuentes, O., Tanques amortiguadores. Reporte T cnico, Proyecto 7045. UNAM, M xico D.F. pp. 1-50, 1978.
- [12] Hajdin, G., Contribution to the evaluation of fluctuation pressure on fluid currents limit areas- based on the pressures recorded at several points of the area. VIII Conference of Yugoslav Hydraulics Association, Portoroz. 1982.
- [13] Toso, J. and Bowers, C., Extreme pressures in hydraulic-jump stilling basins, J. of Hydraulic Engineering, 114(8), pp. 829-843, 1988. DOI: 10.1061/(ASCE)0733-9429(1988)114:8(829)
- [14] Farhoudi, J. and Narayanan, R., Force on slab beneath hydraulic jump. J. of Hydraulic Engineering, 117(1), pp. 64-82, 1991. DOI: 10.1061/(ASCE)0733-9429(1991)117:1(64)
- [15] Mees, A., Estudo dos esfor os hidrodin micos em bacias de dissipac o por ressalto hidr ulico com baixo n mero de Froude. MSc. Thesis. Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil. 2008.
- [16] Bellin, A. and Fiorotto, V., Direct dynamic force measurements on slab in spillway stilling basin. J. of Hydraulic Engineering. 121(10), pp. 686-693, 1995. DOI: 10.1061/(ASCE)0733-9429(1995)121:10(686)
- [17] del Risco, E., Inestabilidad del revestimiento del fondo de un tanque amortiguador. PhD dissertation, Faculty of Engineering. UNAM, M xico D.F., M xico, 1989.
- [18] del Risco, E., Estudio del Levantamiento de una losa de revestimiento. Technical Report, Faculty of Engineering, Universidad del Valle, Cali, Colombia. pp. 2-4, 2006.
- [19] del Risco, E., Hurtado, E. y Gonz lez, M., Estudio experimental de las presiones de levantamiento bajo una losa con juntas transversales al flujo. Tecnolog a y Ciencias del Agua. 1(1), pp. 31-42, 2010.
- [20] Hurtado, E., del Risco, E. y Gonz lez, M., Presiones medidas en la base de una losa de fondo con juntas paralelas al flujo en un canal. Revista de la Facultad de Ingenier a de la Universidad de Antioquia, 47, pp. 39-52, 2009.
- [21] Vasiliev, O.F. and Bukreyev, V.I., Statistical characteristics of pressure fluctuations in the region of hydraulic jump. Proceedings 12th IAHR Congress, Fort Collins, USA. 1967.
- [22] Pinheiro, A., Accoes hidrodin micas em bacias de dissipacao de energia por ressalto. PhD. dissertation, Universidad T cnica de Lisboa, Lisboa, Portugal, 353 P., 1995.
- [23] Khader, M. and Elango, K., Turbulent pressure field beneath a hydraulic jump. J. of Hydraulic Research, 12(4), pp. 469-489, 1974. DOI: 10.1080/00221687409499725
- [24] Fiorotto, V. and Rinaldo, A., Fluctuating uplift and linings design in spillway stilling basins. J. Hydraulic Engineering ASCE. 118(4), pp. 578-596. 1992a. DOI: 10.1061/(ASCE)0733-9429(1992)118:4(578)
- [25] Fiorotto, V. and Rinaldo, A., Turbulent pressure fluctuations under hydraulic jumps. J. of Hydraulic Research, 130, pp. 499-520, 1992b. DOI: 10.1080/00221689209498897
- [26] Lopardo, R., de Lio, J. and Lopardo, M., Physical modelling and design estimation of instantaneous pressures in stilling basins. Proceedings of the XXVIII IAHR Congress. Graz. 132 P., 1999.
- [27] Hassonizadeh, H. and Shafai-Bajestan, M., Experimental measurement of dynamic force on slab in stilling basins. Wetlands Engineering & River Restoration. Conference Proceeding Paper. Section: 47. Chapter: 4. 2001. DOI: 10.1061/40581(2001)125
- [28] Levi, E. y Masa, J., Estudio sobre la flotaci n de losas en tanques amortiguadores. Memorias 2do. Congreso Nacional de Hidr ulica. Culiac n. M xico. pp. 147-165, 1972.
- [29] Levi, E., and del Risco, E., Search for the cause of high-speed-flume revetment failures, ASCE, J. of Performance of Constructed Facilities, 3(2), 1989. DOI: 10.1061/(ASCE)0887-3828(1989)3:2(125)
- [30] del Risco, E., Investigaci n experimental de la falla de losas de revestimiento de tanques amortiguadores. MSc. Thesis, Faculty of Engineering. UNAM, M xico D.F., M xico, 1983.
- [31] Gonz lez, M. y Giraldo, S., Caracterizaci n din mica de sensores de presi n utilizando el principio de la botella de mariotte. Rev. Fac. Ing. Univ. Antioquia, 71, pp. 146-156, 2014.
- [32] Tatone, G., Selecci n de transductores para la medici n de presiones fluctuantes. XXIV Congreso Latinoamericano de Hidr ulica, Punta del Este, Uruguay. pp. 5-9, 2010.
- [33] Lopardo, R.A. and Henning, R.E., Experimental advances on pressure fluctuation beneath hydraulic jumps. Proc., 21st IAHR, Melbourne Australia, Vol. 3, 1985.
- [34] Gonz lez, M., Levantamiento de una losa de piso sometida a un flujo turbulento. PhD Dissertation, Facultad de Minas, Universidad Nacional de Colombia, Medell n, Colombia. 2013.
- [35] Bureau of Reclamation and USACE. Dam safety risk analysis best practices training manual. U.S. Department of the Interior Bureau of Reclamation, Colorado. Version 2.2, cap. 24, pp. 1-47, 2011.
- [36] Hansen, E., Unsteady pressure-driven viscous flows beneath concrete pavement slabs. J. of Hydraulic Engineering, 117(6), pp. 713-724, 1991. DOI: 10.1061/(ASCE)0733-9429(1991)117:6(713)
- [37] Trojanowski, J., Water operation maintenance bulletin. Proceedings United States Society on Dams, 26th annual conference in San Antonio, Texas. May. 1-5, No. 218, pp. 2-19, 2006.
- [38] Melo, J., Pinheiro, A. y Ramos, C., Forces on plunge pool slabs: Influence of joints location and width. J. of Hydraulic Engineering, 132(1), pp. 49-60. 2006. DOI: 10.1061/(ASCE)0733-9429(2006)132:1(49)
- [39] Blair, H.K. and Rhone, T.J., Design of small dams. Bureau of Reclamation, 3a ed., Washington, DC, cap. 9, Spillways structural design details. 1987, pp. 429-434.

M. Gonzalez- Betancourt, was awarded his BSc. Eng in Agricultural Engineering by the Universidad Nacional de Colombia, Palmira, Colombia, and Universidad del Valle, Cali, Colombia, in 2008. He received his PhD. in Engineering with emphasis on Hydraulic Resources in 2014 from the Universidad Nacional de Colombia, Medellín, Colombia. Between 2008 and 2010, he worked at the Hydraulic Laboratory at the Universidad del Valle. Since 2009 he has been associated with the research group “Posgrado en Aprovechamiento de Recursos Hidráulicos” (Postgraduate program on the use of Hydraulic Resources) at the Universidad Nacional de Colombia. From 2014 to date, he has been a researcher in Colombia under the terms of the joint agreement between “Organización de Estados Iberoamericanos” [Organization of Iberoamerican States], “Servicio Nacional de Aprendizaje” [National Apprenticeship Service] and “COLCIENCIAS” [Administrative Department of Science, Technology and Innovation].
ORCID: 0000-0001-5485-8043

L. Posada-García, was awarded his BSc. Eng in Civil Engineering in 1976, by the Universidad Nacional de Colombia, Medellín, Colombia. She was awarded her MSc. in Hydraulic Resources in 1982 and her PhD degree in Civil Engineering in 1995 from Colorado State University, USA. Since 1976, she has been a Professor in the Facultad de Minas, Universidad Nacional de Colombia, Medellín campus, Colombia.
ORCID: 0000-0002-5504-2203



UNIVERSIDAD NACIONAL DE COLOMBIA

SEDE MEDELLÍN
FACULTAD DE MINAS

Área Curricular de Ingeniería Civil

Oferta de Posgrados

Especialización en Vías y Transportes
Especialización en Estructuras
Maestría en Ingeniería - Infraestructura y Sistemas
de Transporte
Maestría en Ingeniería – Geotecnia
Doctorado en Ingeniería - Ingeniería Civil

Mayor información:

E-mail: asisacic_med@unal.edu.co
Teléfono: (57-4) 425 5172

Assessing the loss-of-insulation life of power transformers by estimating their historical loads and ambient temperature profiles using ANNs and Monte Carlo simulations

Andrés Arturo Romero-Quete, Enrique Esteban Mombello & Giuseppe Rattá

*Instituto de Energía Eléctrica, Universidad Nacional de San Juan – Consejo Nacional de Investigaciones Científicas y Técnicas. San Juan, Argentina.
aromero@iee.unsj.edu.ar, mombello@iee.unsj.edu.ar, ratta@iee.unsj.edu.ar*

Received: January 5th, 2015. Received in revised form: October 20th, 2015. Accepted: March 30th, 2016.

Abstract

A non-invasive method useful for asset management is to estimate the functional age of the insulating paper of the transformer that is caused by thermal aging. For this purpose, the hot-spot temperature profile must be assessed by means of some transformer characteristics, the historical load, ambient temperature profiles and a set of equations. In many in-service unit cases, the available data is incomplete. This paper proposes a method to deal with the lack of data. The method is based on the estimation of the historical load and ambient temperature profiles by using an artificial neural network and Monte Carlo simulations. The probable loss of total life percentage of a 30 MVA power transformer is obtained through the proposed method. Finally, the corresponding results for the assessed transformer, a model validation section and conclusions are presented.

Keywords: aging; artificial neural network; asset management; Monte Carlo methods; load profile forecasting.

Evaluación de la pérdida de vida del aislamiento sólido en transformadores de potencia, estimando la historia de carga y los perfiles de temperatura ambiente por medio de redes neuronales artificiales y simulaciones de Monte Carlo

Resumen

La estimación de la pérdida de vida es útil para la gestión de transformadores de potencia. Un método, no invasivo, es estimar la edad funcional del papel aislante del transformador, mediante las guías de carga. Para esto, el perfil de temperatura del punto caliente es calculado a partir de características técnicas del transformador, los perfiles de carga y temperatura ambiente y un conjunto de ecuaciones diferenciales. En la práctica, la información disponible para este análisis es incompleta. En este artículo se presenta un método para estimar la carga histórica y los perfiles de temperatura ambiente experimentados por el transformador, cuando existe falta de datos. Para este fin, el método emplea una red neuronal artificial y simulaciones de Monte Carlo. El método es aplicado a un transformador de potencia de 30 MVA. Los resultados obtenidos son analizados en una sección de validación para finalmente dar las conclusiones del trabajo.

Palabras clave: envejecimiento; red neuronal artificial; gestión de activos, simulación de Monte Carlo; pronóstico del perfil de carga.

1. Introduction

Power transformers play a fundamental, strategic role in power systems. In power transformer management, it is useful

to estimate, as closely as possible, the current condition of the unit. There are several monitoring and diagnostic techniques that aimed to estimate the transformer condition. One of these techniques seeks to assess the loss of total life percentage of the

How to cite: Romero-Quete, A.A., Mombello, E.E. and Rattá, G., Assessing the loss-of-insulation life of power transformers by estimating their historical loads and ambient temperature profiles using ANNs and Monte Carlo simulations. DYNA 83(197), pp. 104-113, 2016.

unit's insulating paper, which can be expressed in terms of the functional age resulting from thermal degradation. It is to be noted that the functional age is different from the calendar age of the transformer [1,2].

A transformer aging failure is an irreparable failure that is more probable when the loss of life percentage is approaching 100%. In this regard, this measure is useful to make appropriate decisions, e.g., the design of a Suitable scheme of maintenance to prolong unit life.

In a transformer in service, internal temperatures increase as load and ambient temperature rise, and vice versa. High internal temperatures reduce the mechanical strength of the cellulose insulation as time passes [3]. The Hot Spot Temperature (HST) is a useful measure to estimate the thermal effect of paper-aging. In fact, after developing a transformer post mortem assessment, it was found that HST is a critical factor [4]. HST is a function of load, ambient temperature, and features of the unit.

Consequently, the functional age of the transformer can be roughly assessed through HST estimation. If the historical HST profile, i.e., the data of HST values against time, is available for the whole period of the unit operation, then the thermal aging of the insulating paper can be estimated as described in the IEEE and the IEC loading guides, [5,6]. The formula to assess transformer aging is based on the condition of new paper, and can only be used to calculate the relative paper deterioration. Since the aging equation is based on the HST, and the HST calculation on load data, it is then necessary to have access to the complete loading history of the unit.

However, transformer loading history is not usually well-known in practice. To overcome this drawback, this paper proposes a method to deal with the lack of data. The methodology is founded on the fact that, nowadays, electric networks have Supervisory Control and Data Acquisition (SCADA) systems that allow for the online storage of measured values such as load, date and ambient temperature. Even if a part of this data has not been recorded, e.g., in the period previous to the SCADA implementation, the available data can be used to estimate the unrecorded historical load data.

The estimation of lacking load data relies on the statistical analysis of ambient temperatures and on the identification of historical load features. An Artificial Neural Network (ANN) is then trained in order to estimate the complete loading probable profile. Because the uncertainty in the HST estimation must be modeled, the next step requires Monte Carlo Simulations (MCS) to be performed by using both ambient temperature and loading probable profiles as inputs. Subsequently, the estimated probable HST profiles are used to assess the frequency of occurrence of the functional age of the insulating paper and its percentage loss of life.

For the HST calculation, a recent model that considers oil viscosity in temperature dependence has been used [7-9]. Finally, the probable functional age of the insulating paper and therefore, the loss-of-insulation life percentage of a real power transformer are assessed, the results are discussed, and conclusions are presented from the case study.

2. Estimation of functional age and loss-of-life percentage of the insulation paper

2.1. Ageing models

In oil-cooled power transformers, paper is used as one of

the insulating materials due to its outstanding characteristics, e.g., excellent dielectric and mechanical properties. Paper is comprised of lignin, hemicellulose and cellulose. The cellulose molecules are built by long chains of glucose rings. The average length of these chains, termed Degree of Polymerization (DP), determines the functional characteristics of the paper, [3].

When paper is exposed to water, oxygen, heat and acids, all of which are aging agents, the long molecular chains are cleaved by chemical reactions. As a consequence, the DP and the expected lifespan of the insulating paper decrease.

The aging agents act simultaneously; consequently, the mathematical modeling of the aging process is very complex, resulting in a non-linear Arrhenius plot. However, for practical purposes most of the researchers and organizations assume independent processes, and as such, the total degradation becomes the sum of degradation from each process that can be described by an individual equation, [3,5]. Annex I in [5], introduces an interesting discussion of the history of loss of insulation life. It also states that each aging agent will have an effect on degradation rate, so they must be individually controlled. Moreover, it is observed that whereas the water and oxygen content of the insulation can be controlled by the transformer oil preservation system, the transformer operating personnel is responsible for the control of heat.

This research is mainly focused on the influence of heat in the aging process. However, it must be recognized that time variations of the other agents could strongly affect the aging speed. Note that as concluded in [10,11], "water is more important than oxygen for transformer aging; and acids are of central significance for understanding the aging of paper and for evaluating the effects of oil maintenance."

This assumption could relate to the transformer in the case study reported here, since in the San Juan Province, Argentina, where the unit operates, the average ambient relative humidity is 48%. Also, as can be observed in references [12,13], there is a direct relation between moisture in air, oil, and insulating paper. The higher the relative air humidity, the higher the moisture uptake rate of the insulating liquids will be, and consequently, that of the paper. Consequently, the influence of moisture for this specific case is lower than that for transformers located in areas with a higher average relative humidity. Nevertheless, to achieve more general results, additional research must be conducted in order to model the moisture influence.

Eq. (1), as written in [5], defines the aging acceleration factor, F_{AA} , as a function of the HST in the windings, θ_{HS} in °C. The empirical constant 15000 is the aging rate constant that is better described in annex I, contained in [5]. The reference temperature, $\theta_{HS,R}$, coincides with the rated HST, which for thermally upgraded paper is $\theta_{HS,R}=110$ °C.

$$F_{AA} = e^{\frac{15000}{\theta_{HS,R}+273} - \frac{15000}{\theta_{HS}+273}} \quad (1)$$

Eq. (2) allows the functional age of insulating paper to be assessed. This expression is the summation in a determined period of time of the products obtained by multiplying intervals of time, Δt_i , (with $i=1, 2, \dots, n$, and n as the number of time intervals for the whole period of time) by its respective computed aging acceleration factor, $F_{AA,i}$, for which the insulating paper has been exposed to a given $\theta_{HS,i}$.

$$A_F = \sum_{i=1}^n F_{AA,i} \cdot \Delta t_i \quad (2)$$

In (2), both Δt_i and n are specified by the load profiles that will later be described in section 4. The functional age, A_F , is then used to calculate the loss-of-insulation life percentage as:

$$\% \text{ loss of life} = \frac{A_F}{A_{life}} \times 100 \quad (3)$$

In (3), A_{life} is the normal insulation life at the reference temperature, $\theta_{HS,R}$, in hours. The benchmark of normal insulation life for well-dried, oxygen-free 65 °C average winding temperature rise system is reported in [4,14]. A 200 retained degree of polymerization in insulation has been adopted in this article as the normal insulation life criterion, and it corresponds to 150000 hours for $\theta_{HS,R}=110$ °C.

2.2. Hot spot temperature estimation

The functional age, due to thermal aging, depends on the HST. This is because the winding insulating paper ages faster in the region in which the hottest temperature is present. When no fiber-optic sensors are installed inside the transformer, as is the common case in older units, the HST must be estimated through models based on measured oil temperatures [15].

Several HST models, ranging from simple to very complex, have been proposed in the bibliography, [5-9]. The qualitative criteria for selecting a HST model for this study were: *i*) availability of the technical transformer data and parameters the HST model is based on, *ii*) performance simulating transient or varying loads with altering ambient temperature, and, *iii*) top-oil temperature (TOT) based model, which exhibits a suitable performance [16].

A thermal model satisfying most of the above defined criteria is the HST model presented by Susa in [7-9]. In this model, HST and TOT differential equations account for oil viscosity changes with temperature. One advantage of this model is that it is only based on nameplate data and few heat run test protocol parameters. The differential equations for the TOT, θ_{TO} , and the HST, θ_{HS} , are respectively:

$$\left(\left((1 + R \cdot K^2) / (1 + R) \right) \cdot \mu_{pu}^n \cdot \Delta \theta_{TO,R} \right) = \left(\mu_{pu}^n \cdot \tau_{TO,R} \cdot \frac{d \theta_{oil}}{dt} \right) + \frac{(\theta_{TO} - \theta_{amb})^{n+1}}{\Delta \theta_{TO,R}^n} \quad (4)$$

$$\left(\left(K^2 \cdot P_{W,pu}(\theta_{HS}) \right) \cdot \mu_{pu}^n \cdot \Delta \theta_{HS,R} \right) = \left(\mu_{pu}^n \cdot \tau_{W,R} \cdot \frac{d \theta_{HS}}{dt} \right) + \frac{(\theta_{HS} - \theta_{TO})^{m+1}}{\Delta \theta_{HS,R}^m} \quad (5)$$

where, R is the ratio of load losses at rated current to no-load losses, K is the load factor and quotient of load over

rated load, μ_{pu} is the per unit oil viscosity given by the ratio of the current oil viscosity over rated oil viscosity μ/μ_R , n is an empirical constant dependent on whether the oil circulation is laminar ($n=0.25$) or turbulent ($n=0.33$), m is an empirical constant related with the shape of the thermal curve, $\Delta \theta_{TO,R}$ is the rated TOT rise over ambient (measured in °C), $\Delta \theta_{HS,R}$ is the rated HST rise over TOT (measured in °C), $\tau_{TO,R}$ is the rated top-oil time constant, $\tau_{W,R}$ is the rated winding time constant in minutes, and θ_{amb} is the ambient temperature.

Equations to model oil viscosity variations with temperature are shown in [7-9]. For Oil Natural - Air Natural (ONAN) cooling mode, the value of constants n and m is 0.25. For the Oil Natural - Air Force (ONAF) and Oil Force - Air Force OFAF cooling modes, in power transformers with external cooling these values are $n=0.5$ and $m=0.1$.

The winding losses corrected for the HST, $P_{W,pu}(\theta_{HS})$, are calculated, as per the following:

$$P_{W,pu}(\theta_{HS}) = P_{DC,pu} \frac{\theta_{HS} + \theta_K}{\theta_{HS,R} + \theta_K} + P_{E,W,pu} \frac{\theta_{HS,R} + \theta_K}{\theta_{HS} + \theta_K} \quad (6)$$

where $P_{DC,pu}$ is the ratio of DC losses to total winding losses, $P_{E,W,pu}$ is the ratio of winding eddy losses to total winding losses, θ_K is the temperature correction factor, 225 for Al and 235 for Cu, and the total winding losses are $P_W = P_{DC} + P_{E,W}$.

It should be noted that DC losses in (6) are measured by applying a direct current to the windings, whereas eddy losses in the windings are not separately measurable. Furthermore, DC losses are not necessarily given on the nameplates. Nevertheless, Jauregui states in [16] that for load levels near the transformer rated load, (i.e., in a no-overload transformer condition) the winding losses correction is unnecessary. Therefore, if necessary, $P_{W,pu}(\theta_{HS})$ can be omitted in eq. (5), which will lead to:

$$K^2 \mu_{pu}^m \Delta \theta_{HS,R} = \mu_{pu}^m \tau_{W,R} \frac{d \theta_{HS}}{dt} + \frac{(\theta_{HS} - \theta_{TO})^{m+1}}{\Delta \theta_{HS,R}^m} \quad (7)$$

Thus, it can be concluded that the Susa HST model can be simplified for applications where only nameplate data and no heat-run test protocol are available.

2.3. Necessary data to estimate the hot spot temperature

A listing of the transformer data necessary to calculate the HST at time i for the mentioned model includes the following: winding eddy losses, winding resistive losses, DC or I²R losses, load losses, no-load losses, winding-time constant (estimated by means of weight of windings), oil-time constant (estimated by means of the mass of oil), rated TOT rise over ambient temperature, rated HST rise over TOT, rated TOT, rated HST, and both ambient temperature and load at time i . Moreover, selecting appropriate transformer constants requires knowledge of characteristics such as the cooling stages and the winding conductor material. Once the HST is obtained at time i , the procedure

must be repeated for the whole operational period of the unit in order to apply (1) and (2) to estimate A_F .

3. Problem description and proposed methodology

3.1. Problem description

The aim of this paper is to present a methodology to estimate the functional age, A_F , of insulating paper of aging power transformers when part of data regarding load and ambient temperature profiles for the complete operational period of the assessed transformer is missing.

3.2. Scheme of the proposed methodology

To assess the A_F of the insulating paper, missing historical temperature and load data must be estimated to be able to then apply the models presented in section 3. Taking this into consideration, the following general scheme is proposed as a suitable methodology to address the problem.

In the following section each one of the steps in Fig. 1 will be explained in more detail.

Step 1: Acquire the following input data:

- Data as described in section 2.3, to thermally model the transformer under study.

- All available information regarding power delivered hourly through the transformer, and the ambient temperature at the power station; e.g., information acquired by the SCADA.

- All historical information about ambient temperature and monthly or yearly electric energy consumption in the district in which the power transformer operates.

Step 2: Classify, organize and filter inadequate or atypical data from the hourly load and ambient temperature profiles, acquired in step 1. This step must provide typical values to

describe the behavior of both profiles. This preprocessed data is necessary for two purposes: to construct ambient temperature profiles for normalized days and to train the load forecasting model.

Step 3: Analyze the historical ambient temperature statistically. Once the available ambient temperature profile has been preprocessed, normalized days are defined.

For this purpose, each month of available ambient temperature data can be divided into three packages of approximately ten days, i.e., $D \approx 10$. For each package the hourly mean temperature and the standard deviation, σ_h , is calculated as:

$$\bar{\theta}_h = \frac{1}{D} \sum_{d=1}^D \theta_{h,d} \quad \text{and} \quad \sigma_h = \sqrt{\frac{1}{D-1} \sum_{d=1}^D (\theta_{h,d} - \bar{\theta}_h)^2} \quad (8)$$

where $d=1,2,\dots,D$, and θ_h is the observed temperature at hour h for each day d . Moreover, there is a correlation between adjacent hourly ambient temperatures. To model this dependence, a 24×24 correlation matrix, \mathbf{R}_d , is computed for each normalized day. Based on this routine, 24 pairs of $\bar{\theta}_h$ and σ_h , with $h=1,\dots,24$, and one correlation matrix \mathbf{R}_d are obtained for each normalized day. It is noted that the covariance matrix of the normalized day \mathbf{S}_d can also be computed since it is a function of the matrix \mathbf{R}_d and the standard deviation vector of 24 elements $\sigma_{h,d}$, [16].

Step 4: Assign probability distributions of ambient temperature to the missing period of time. Completing the ambient temperature curve procedure includes assigning one of the obtained normalized days in the previous step to each day in the missing period of time. This action can be performed by comparing features of the normalized days with those known for the days in the missing period, e.g. the Euclidian distance, d , which is given by:

$$d = \sqrt{(\bar{x} - \bar{y})^2} \quad (9)$$

where \bar{x} is a vector containing a pair of values of the historical maximum and minimum daily temperatures for unknown days (i.e., whose hourly temperature profile is unknown) and \bar{y} is a vector containing the same values of a normalized day. These can be used to select the normalized day that exhibits the minimal distance in order to then assign this information to the missing day.

Step 5: Define the input data for the estimation of the historical load and implement a suitable ANN for load forecasting.

References to estimate the historical load were not found; however, load forecasting is a well-known field of work in electrical engineering. In effect, there are various techniques, which have been extensively studied and used in time series forecasting, e.g. Autoregressive Integrated Moving Average (ARIMA) and its variants, Artificial Neural Networks (ANNs), and hybrid methods such as ANNs/ARIMA [18]. However, in this paper ANNs have been selected to deal with the described problem since some researchers have compared and tested this technique, concluding that ANNs could successfully capture the nonlinear relationship between the load and the weather [19]. In effect, ANNs are better suited

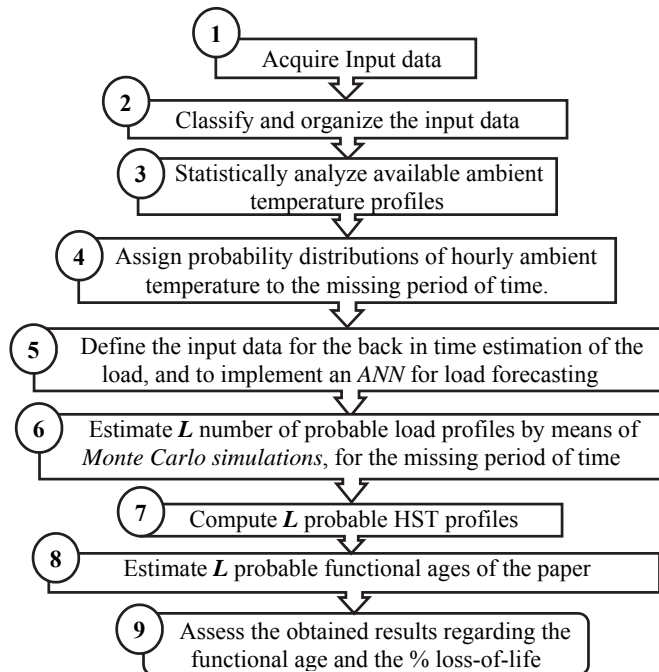


Figure 1. Proposed methodology scheme.

Source: The authors

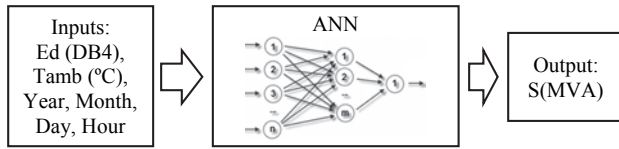


Figure 2. Conceptual scheme of the implemented ANN.
Source: The authors

in terms of ARIMA since the first can capture nonlinear patterns between input and output parameters while the second is limited by the pre-assumed linear form of the model. It must be noted that hybrid methods which combine both ARIMA and ANN are being investigated. Future research could be conducted on this subject to improve the forecasting accuracy; however, this research lies outside the scope of this paper.

In general, an ANN does non-linear curve fitting and is able to find a function to subsequently calculate load values dependent on various input parameters [20]. A two-layer feed-forward network with sigmoid hidden neurons and linear output neurons can fit multi-dimensional mapping problems arbitrarily well if provided with consistent data and enough neurons in its hidden layer [21]. In this network, the information only moves in one direction, forwards, i.e., data goes from the input nodes through the hidden nodes and then to the output node.

Therefore, we selected a feed-forward ANN structure. It basically compares the calculated output value with a desirable result, the target value. The target values are the load values that are stored by the SCADA: in effect, the available load profile. Whereas there is one single output, the load, there are several input parameters. In this approach, the input parameters are the hourly ambient temperatures (T_{amb}), year, month, weekday, hour and the monthly energy demand (E_d). Fig. 2 illustrates inputs and outputs defined for the ANN.

Step 6: Estimate a quantity of L probable ambient temperature and load profiles by means of Monte Carlo simulations, where L is the number of simulations, for the missing period of time. For this purpose, the probability distributions of ambient temperature defined in step 4, and the input data for the ANN implemented in step 5 are used.

Step 7: Compute L probable HST profiles by using the formulation presented in section 2.2 from the L probable profiles that was computed in steps 4 and 6.

Step 8: Estimate L , probable functional ages of the insulating paper by using formulation presented in section 2.1 from the L HST profiles computed in step 7.

Step 9: Assess the results obtained regarding the A_F from the L probable values computed in step 8, and compute a histogram for the % loss-of-insulation life.

4. Case study

In this section, probable HST profiles and a probability density function for the loss-of-insulation life of a 30 MVA and 132 +6/-12 x 1 %/ 34.5/13.8 kV rated step-down transformer will be estimated through the general scheme, shown in Fig. 1. The unit is connected in YN/yno/d11. The rated cooling is ONAN/ONAF at 70/100 % of load. It should

be observed that the order to start running the assessed unit fans, i.e. to start the ONAF mode, is given by a commercial temperature monitor.

It is remarkable that there have been no changes in the configuration of the power station since 1994, the year in which the unit under study was installed.

4.1. Case study description

In the case of the transformer that was being studied, the unit operator began to record the load and ambient temperature data on 20/01/08. Within this study, the period that will be covered extends until 31/07/10. Therefore, twenty-four values of both load and ambient temperature for each of the 924 days listed in the database are available. However, since the unit was installed in 1994, there is a lack of load and ambient temperature data for the period between 15/02/94 and 19/01/08. Therefore, to assess the functional age of the insulating paper, missing temperature and load data must be historically estimated for the particular period of time.

4.2. Applying the proposed methodology to case study

- 1) *Step 1.* Available data is compiled from several databases that have been classified as follows:
 - i) Database 1: 924 x 24 pairs of values of load and ambient temperature data for the period 20/01/08 - 31/07/10.
 - ii) Database 2: archives published by the local meteorological service, which includes daily minimum and maximum temperatures covering the period 01/01/01 - 31/07/10.
 - iii) Database 3: archives published by the local meteorological service, which are comprised of monthly average temperatures for the period from February 1994 to August 2010.
 - iv) Database 4: the district's monthly energy demand -where the unit is located- from February 1994 to August 2010.
 - v) Database 5: nameplate data of the assessed transformer.
- 2) *Step 2.* In *database 1*, missing or erroneous values had to be detected and the associated data (e.g., date) removed.
- 3) *Step 3.* To complete the ambient temperature curve, first, normalized days, based on the filtered measured temperature data reported in *database 1* must be established.

A total of 91 normalized daily ambient temperature profiles were determined by using the routine presented in section 3.2., Step 3. For this purpose, the time period covered by *database 1*, i.e., from 20/01/08 to 31/07/10, was considered. Each one of these 91 normalized days is representative of a day in a month of the year; i.e., there are 7 normalized days representing a day in January, 9 normalized days representing a day in February, 9 for March, ..., etc. For example, a normalized ambient temperature for a day in February, obtained from the population of 10 days from 01/02/08 to 10/02/08, is shown in Fig. 3. The vertical lines indicate the probabilistic distribution of the normalized hourly temperature. Moreover, the mesh that represents the correlation matrix for that normalized day is plotted in Fig. 4. A strong correlation between the hourly ambient temperatures, especially for those that are adjacent, is observed.

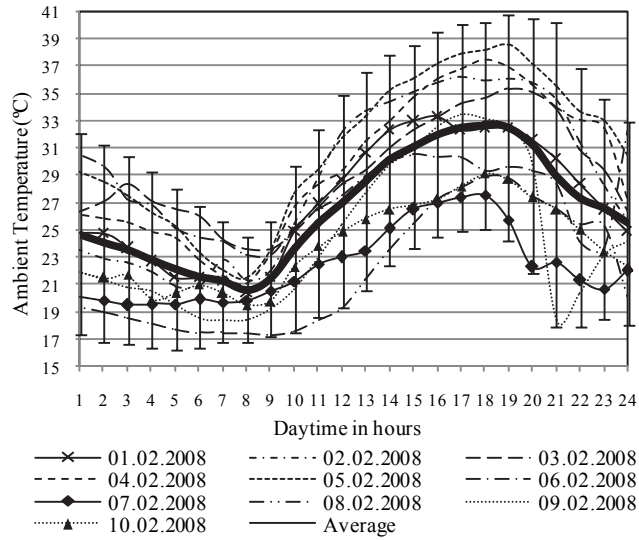


Figure 3. A normalized day in February. Vertical lines plot the 95% confidence intervals for the hourly temperature values. The measured temperature curves are shown for ten days from 01/02/08 until 10/02/08. Source: The authors

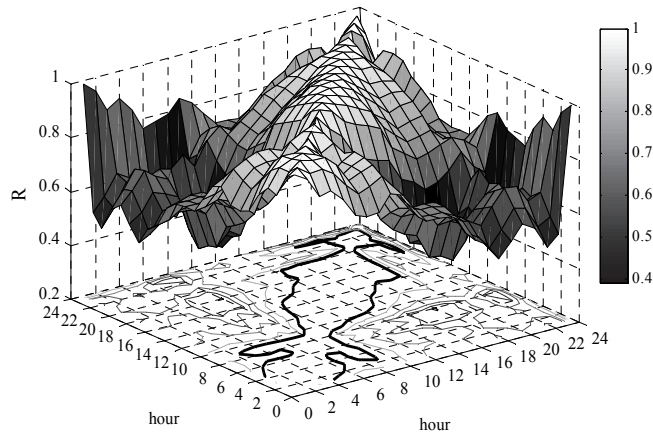


Figure 4. Correlation matrix of the hourly ambient temperature of a normalized day in February. The matrix has been computed using ten days of data from 01/02/08 until 10/02/08. Source: The authors

4) Step 4. Ambient temperatures were then determined for the 01/01/01 - 19/01/08 period. The minimum and maximum temperature values of the 91 normalized days were employed as features of typical days. From these features and from the daily minimum and maximum temperatures that are shown in database 2, the Euclidean distances were computed by using (9). In this way, one of the 91 normalized days was assigned to each day between 01/01/01 and 19/01/08. Subsequently, the corresponding monthly average temperatures were calculated for this period. The assigned normalized days were scaled by the ratio of the calculated monthly average temperatures over the real monthly average temperature given in database 3. As such, the monthly average temperatures, including assigned normalized days from the 01/01/01 - 19/01/08 period coincide with the real monthly average values that are given for the same period.

Table 1

ANN Structure Performance - MAPE in (%) and MAE in (MVA) were computed for ANN with 10, 20, 30 and 40 Hidden Neurons

Hidden Neurons	10	20	30	40
MAPE	0.137	0.126	0.119	0.118
MAE	2.193	2.015	1.864	1.826

Source: Authors

Table 2

ANN performance assessment - MAPE in (%) and MAE in (MVA) were computed for 20 ANN

ANN	1	2	3	4	5	6	7	8	9	10
MAPE	0.117	0.117	0.117	0.116	0.117	0.116	0.117	0.118	0.119	0.116
MAE	1.827	1.823	1.827	1.804	1.837	1.821	1.824	1.849	1.854	1.814
ANN	11	12	13	14	15	16	17	18	19	20
MAPE	0.116	0.117	0.117	0.119	0.117	0.116	0.116	0.117	0.119	0.117
MAE	1.813	1.828	1.825	1.859	1.827	1.815	1.814	1.826	1.854	1.825

Source: Authors

5) Step 5. The lacking load values have to be estimated for the 15/02/94 - 19/01/08 period. ANN structures, such as those shown in Fig. 1, with 10, 20, 30 and 40 hidden neurons were tested. The performance of these structures was verified by two measures: i) the mean absolute percentage error (MAPE), and ii) the mean absolute error (MAE).

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{LM_{,i} - LF_{,i}}{LM_{,i}} \right| \quad (\%) \quad \text{and}, \quad (10)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |LM_{,i} - LF_{,i}| \quad (\text{MVA})$$

where n is the total number of load values between 20/01/08 - 31/07/10, $LM_{,i}$ and $LF_{,i}$ are the measured and forecasted hourly load values, in MVA, respectively. The performance values for each ANN are given in Table 1.

By increasing the number of hidden neurons, the performance of the ANN improves. However, there is only a small improvement for more than 30 hidden neurons. Therefore, a structure of 40 hidden neurons was implemented. Moreover, when training an ANN, the initial weights and biases of the network are usually randomly generated taking values between -1 and 1 . This initialization has the ability to produce an improved network. Therefore, twenty two-layer feed-forward networks with 40 hidden neurons were created and trained for the 20/01/08 - 31/07/10 period. With the real load values and those estimated through the ANN, the performance of these 20 ANNs was verified again by means of MAPE and MAE. The performance values for each one of these 20 ANNs are given in Table 2.

Network 4, which showed the best performance, was chosen for its back-in-time estimation of load. Diagrams in Fig. 5 show the forecast loads for two different fourteen-day periods. It can be observed from the graphical results that the forecast load follows the real measured values closely.

6) Step 6. Probable ambient temperature profiles were created by using MCS. A thousand ambient temperature profiles, $L=1000$, each containing 123168 temperature values included in the period

15/02/94 - 19/01/08, were generated by multivariate normal random numbers. These numbers were computed from the hourly normal distributions characterized by the mean hourly ambient temperature, θ_h , the corresponding standard deviation, σ_h , and the daily covariance matrix, S_d that were obtained in steps 3 and 4.

Once the probable ambient temperature profiles were obtained, they were constrained to fit the reported information given in *database 3*, e.g., 10 randomly generated ambient temperature profiles for the first 15 days of January of 2001 are shown in Fig. 6.

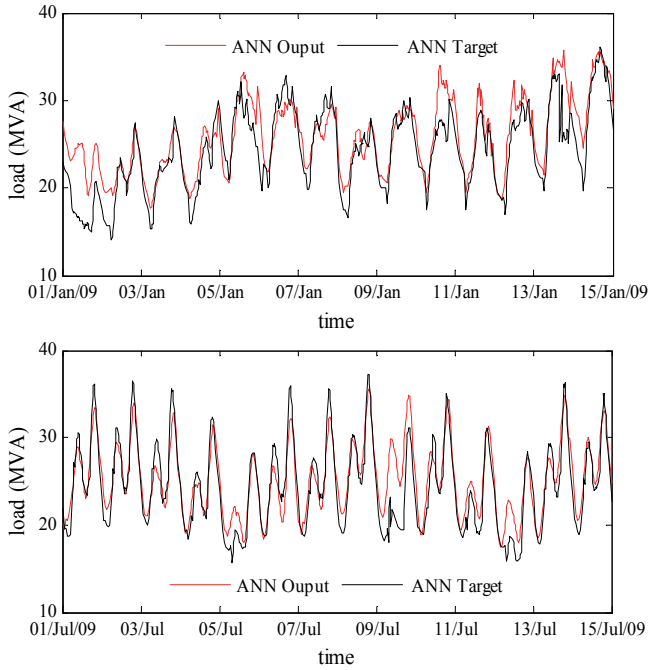


Figure 5. Forecast load profiles for two fourteen-day periods. Period one: summer season, corresponding to 01.01.2009 at 01:00 until 15.01.2009 at 01:00. Period two: winter season, corresponding to 01.07.2009 at 01:00 until 15.07.2009 at 01:00.

Source: Authors

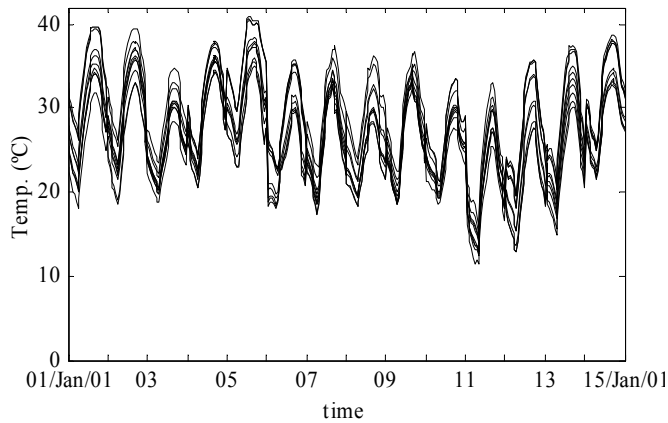


Figure 6. Ambient temperature profiles generated randomly for the period 01/01/01 - 15/01/01.

Source: Authors

The selected ANN network was then used to estimate 1000 of the transformer's probable load profiles. Each hourly load profile was estimated from a corresponding probable temperature curve and the input data from *database 4*.

7) *Step 7*. The HST model presented in section 2.2 was adjusted for the unit characteristics of the case study. Moreover, to deal with the changes in the cooling modes, ONAN/ONAF, it was necessary to model the behavior of the temperature monitor that commands the fans. This type of device uses the following approach to estimate the control HST:

$$\theta_{HS,c} = \theta_{TO} + H \cdot \Delta\theta_{HS,R} \cdot K^y \quad (11)$$

where, H and y are set values which depend on the transformer characteristics, in fact, in this case study, $H=1.3$ and $y=1.6$. It must be noted that the control HST ($\theta_{HS,c}$) computed from (11) is an approximation, the only purpose of which is to command the fans. In fact, fans are turned-on when $\theta_{HS,c}$ is higher than 85°C, and turned-off when $\theta_{HS,c}$ becomes lower than 75°C so as to avoid repetitive on-off fan operations.

In the algorithm that was implemented, when both θ_{TO} , computed from (4), and K (which depends on the load) yield $\theta_{HS,c}$ values higher than 85°C; when the temperature is increasing, or lower than 75°C; or when the temperature is decreasing, then the empiric constants n and m in eq. (4)-(5) change and assume their respective values in agreement with the activated cooling mode.

The HST calculation was divided into two parts. The first part is the deterministic HST calculation based on the load profile data covering the period from 20/01/08 to 31/07/10.

The second part is the probabilistic HST calculation that was made for all 1000 estimated load and ambient temperature profiles determined for the 02/15/94 to 19/01/08 period. Therefore, 1000 HST curves were computed. One of these HST-curves is shown in Fig. 7.

Moreover, Fig. 8 presents a diagram showing the average annual loading. This data shows the loading trend of the 30 MVA transformer. A fact it is worthwhile pointing out is the sudden decrease in the demand that timely coincided with the 2001 political and financial crisis in Argentina; this is also reflected in the forecasted load.

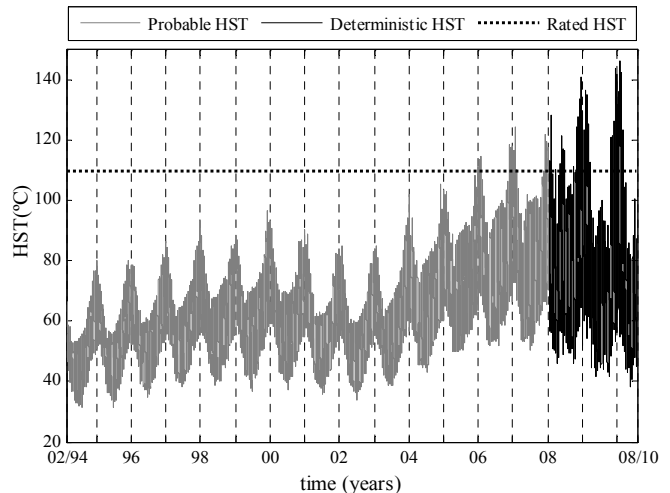


Figure 7. A randomly estimated HST profile.

Source: Authors

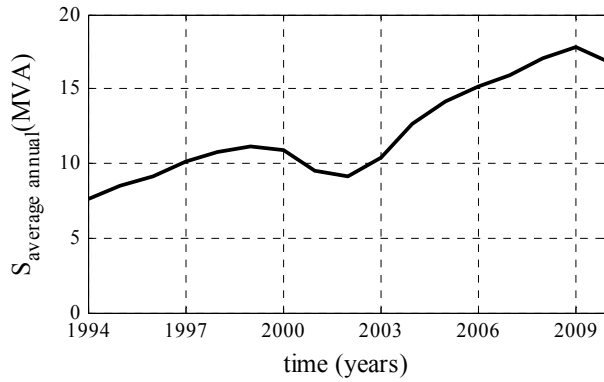


Figure 8. Forecasted average annual loading of the 30 MVA transformer. Note that computed average annual loading for the 2010 year is only comprised of the period from January to July, which is the reason for a value lower than that computed with 2009.

Source: Authors

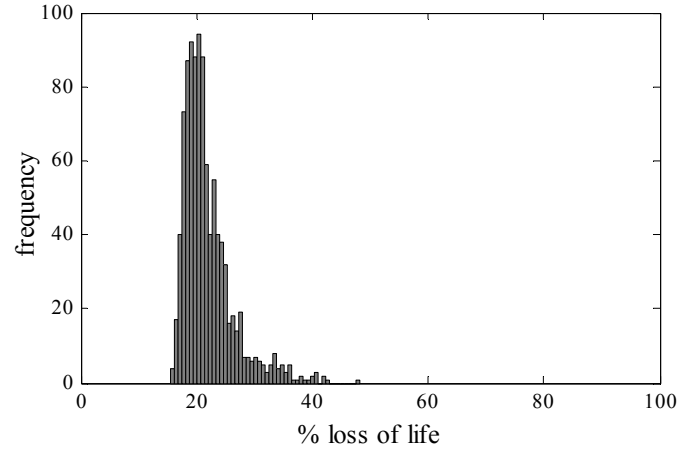


Figure 10. Loss of life percentage of the insulating paper.

Source: Authors.

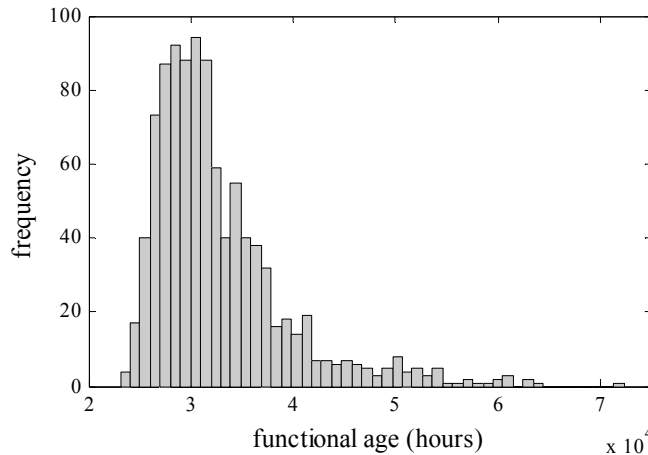


Figure 9. Histogram for functional age of the insulating paper, A_F .

Source: Authors.

- 8) Step 8. After pre-processing the load profile data, estimating the ambient temperature and the load, and subsequently determining the 1000 probable HST profiles for the 15/02/94 – 31/07/10 period, the functional age of paper was determined. The first component functional age is deterministic, and corresponds with the accumulated age during the period 20/01/08 to 31/07/10. This was computed by using, (1), the accelerated aging factor for thermally-upgraded paper, and, (2), the functional age, A_F , of insulating paper. The resulting functional age for this period was 8415.5 hours.

The second component of A_F is probabilistic, and it was estimated for all the 1000 probable HST profiles, also by means of (1) and (2). Therefore, $L=1000$ probable values for the A_F of the insulating paper were estimated for the 15/02/94 to 19/01/08 period. The total A_F of the insulating paper was computed by adding both the deterministic and the probabilistic components. Fig. 9 shows the distribution of the total A_F in a histogram of 50 bins.

- 9) Step 9. It was observed from the results that 90% of computed A_F are less than 41376 hours, and 95% (950) of the simulations yield to A_F values lower than 47739 hours.

Finally, the los-of-insulation life percentage, due to thermal aging, was computed by using (3), and the $DP=200$ end of life criterion. As such we obtained the histogram shown in Fig. 10, in which it can be observed that the studied power transformer's insulating paper is still in good condition, from a thermal aging point of view. In fact, from the histogram, it can be seen that the bin centered at 21.4% loss of life value reports the highest frequency of occurrence, i.e. it represents the mode.

5. Validation

In this section the performance of the proposed methodology is assessed.

5.1. Probabilistic method assessment

The result obtained in section 4.2, step 8, ($A_F = 8415.5$ hours for the 20/01/08 to 31/07/10 period), i.e. the period for which hourly ambient temperature and load profiles are available, is compared with those obtained by assuming that data was not recorded. For this purpose, historical daily minimum and maximum temperatures, obtained from 91 normalized daily ambient temperature profiles and trained ANN were used to assess the probable A_F by means of MSC. The most frequent bin (in effect, that containing 91 simulations) is the one centered in 8680 h, with edge values of 8580 h and 8790 h. Additionally, 950 simulations (95%) yield lower values than 11119 h.

5.2. Assessment of ambient temperature reconstruction methods, founded on sinusoidal functions of time

References [22] and [23] present deterministic methods to produce hourly temperature data by assuming that yearly temperature amplitudes and hourly temperatures follow sinusoidal functions of time. To assess the performance of such proposals, again, a comparison of the results obtained from the period 20/01/08 to 31/07/10 was undertaken.

In Fig. 11, actual and estimated by the sinusoidal functions of time ambient temperature profiles are plotted. From the estimated profile, the A_F calculation is performed, obtaining a result of 6091 h.

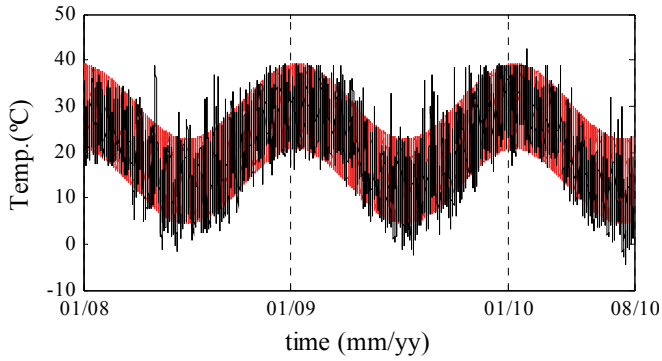


Figure 11. Actual (black) and estimated by sinusoidal functions (red) ambient temperature profiles.
Source: Authors.

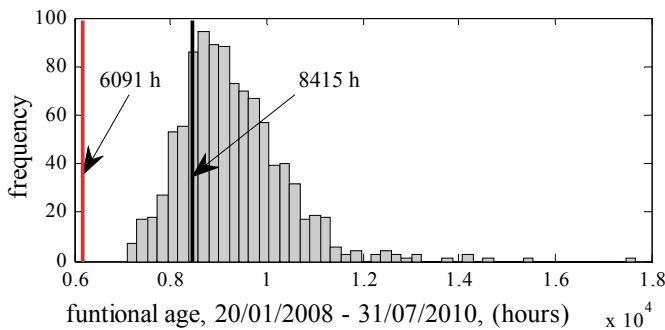


Figure 12. Histogram for functional age of the insulating paper assessed for the period 20.01.2008 - 31.07.2010. The black line indicates the value obtained from available data. The red line indicates the value obtained from the sinusoidal method to reconstruct the ambient temperature curve.
Source: Authors.

Table 3

Comparison of the Proposed Method vs. the Method based on Sinusoidal Functions of time

Actual A_F	Proposed method	Difference	Sinusoidal method	Difference
8415.5 h	8680 h	3.1%	6091 h	27.6%

Source: Authors

This result shows that A_F is underestimated when the sinusoidal functions of time are employed to reconstruct the ambient temperature. A possible explanation for this result is that hot days are omitted for periods falling out of the hottest month; this can be observed in Fig. 11.

5.3. Discussion

The results obtained from test in sections 5.1 and 5.2 are shown in Fig. 12. The deterministic result obtained in 4.2, and the central value of the most frequent bin obtained in 4.1 are compared with the reference of 8415.5 h in Table 3.

These results show that the probabilistic method proposed for the reconstruction of ambient temperature curve could lead to results that do not underestimate the functional age of the insulating paper. Moreover, the results also highlight that the proposed method is able to model the uncertainty involved in the problem that is the motivation for this paper.

To conclude this discussion, there is a lot of research that needs to be done in order to achieve a higher degree of accuracy in the predictions about the transformer condition. It is widely recognized that in this problem there are several sources of uncertainty. For instance, there is uncertainty in the measurement and posterior diagnostic model of frequency response analysis when looking for faults in core and coils [24-26]. The situation is the same for DGA when oil is analyzed (in fact, there are several methods to infer incipient faults in the transformer through DGA, but none of these can be considered as the best one). Furan analysis to infer the DP, is affected by changes in oil, uncertainty about the mass of insulating paper, the type of paper, etc.

In this context, it is expected that models to assess the loading history of a transformer may also be affected by uncertainty. There is uncertainty affecting (1); i.e., about the constant $B=15000$. There have been different values reported in [5]. Moreover, as stated in section 2.1, the reference value for end-of-insulation life and the Arrhenius plot are also affected by the three insulating paper aging mechanisms, hydrolysis, pyrolysis and oxidation.

Finally, there is also uncertainty when one attempts to reconstruct historical load and ambient temperature profiles (especially in the adopted forecasted models, which undoubtedly can be improved, e.g., by testing other ANN structures, or employing hybrid methods as ARIMA-ANN). However, it is noted that one of the major contributions reported in this paper is related to the capability of the proposed method in modeling uncertainty. Most of the standard methods are deterministic. Therefore, these do not recognize the uncertain nature of the analyzed problem.

6. Conclusions

This paper introduces a methodology to estimate missing load data to be used in order to determine the functional age of power transformers' insulating paper. The methodology is founded on the statistical analysis of local ambient temperatures, and on ANNs and MCS to estimate probable historical HST profiles. Subsequently, through the thermal aging principles of insulating paper, it is possible to assess the functional age and loss-of-insulation life percentage due to thermal effects.

Later on in the research, a case study for a 30 MVA rated transformer with its corresponding database was presented. Based on this database and on long term temperature measurements from an external weather station, through MCS, 1000 probable ambient temperature curves were generated for the whole operational period of the unit. Then, using an ANN 1000, probable load profiles were obtained. With these results, a corresponding number of HST profiles were then gathered.

Test results were processed to estimate the frequency of the occurrence of the studied transformer's functional age. The results obtained show that the insulating paper from the unit under investigation is still in good condition, in terms of a thermal aging point of view and from its loading history analysis. Moreover, it was shown that calculation of functional age is also useful to estimate the loss of life percentage due to thermal aging. The presented methodology will contribute to the transformer condition monitoring as a

cheap, comparably fast and easily accessible tool, even for load profiles that do not contain the entire load and ambient temperature data for the whole period of unit operation.

Further research on this issue must be conducted in order for several other important aspects to be considered. These include the following in particular: to include in the model influence of moisture on the aging speed, to improve the forecasting accuracy by testing hybrid methods as ARIMA-ANN, and to assess both the functional age and the loss-of-insulation life percentage in forecasted future load scenarios. Finally, the proposed method is an applicable tool for a major project research that is searching for the determination of a transformer risk index.

References

- [1] Li, W., Risk assessment of power systems, models, methods and applications. IEEE Press Series on Power Engineering, 2005.
- [2] Romero, A.A., Mombello, E.E. and Rattá, G., An overview on power transformer management: Individual assets and fleets, Transmission and Distribution Latin America Conference and Exposition (T&D-LA), Sixth IEEE/PES, pp. 1-7, 2012. DOI: 10.1109/TDC-LA.2012.6319081
- [3] CIGRE Brochure 323. Aging of cellulose in mineral-oil insulated transformers. Taskforce D1.01.10, Ed. 2007.
- [4] Lelekakis, N., Wenyu, G., Martin, D., Wijaya, J. and Susa, D., A field study of aging in paper-oil insulation systems, IEEE Electrical Insulation Magazine, 28(1), pp. 12-19, 2012. DOI: 10.1109/MEI.2012.6130527
- [5] IEEE Guide for loading mineral-oil-immersed transformers and step-voltage regulators. IEEE Standard C57.91, 2011.
- [6] IEC Power Transformers, Part 7: Loading guide for oil-immersed power transformers. IEC 60076-7, 2005.
- [7] Susa, D. and Nordman, H., A simple model for calculating transformer hot-spot temperature, IEEE Transactions on Power Delivery, 24(3), pp. 1257-1265, 2009. DOI: 10.1109/TPWRD.2009.2022670
- [8] Susa, D. and Lehtonen, M., Dynamic thermal modeling of power transformers: further Development-part I, IEEE Transactions on Power Delivery, 21(4), pp. 1961-1970, 2006. DOI: 10.1109/TPWRD.2005.864069
- [9] Susa, D. and Lehtonen, M., Dynamic thermal modeling of power transformers: further Development-part II, IEEE Transactions on Power Delivery, 21(4), pp. 1971-1980, 2006. DOI: 10.1109/TPWRD.2005.864068
- [10] Lundgaard, L.E., Hansen, W., Linhjell, D. and Painter, T.J., Aging of oil-impregnated paper in power transformers, IEEE Transactions on Power Delivery, 19(1), pp. 230-239, 2004. DOI: 10.1109/TPWRD.2003.820175
- [11] CIGRE, Moisture equilibrium and moisture migration within transformer insulation systems, Vol. A2.30, Ed. 2008
- [12] Du, Y., Zahn, M., Lesieutre, B.C., Mamishev, A.V. and Lindgren, S.R., Moisture equilibrium in transformer paper-oil systems, IEEE Electrical Insulation Magazine, 15(1), pp. 11-20, 1999. DOI: 10.1109/57.744585
- [13] Fofana, I., Borsi, H. and Gockenbach, E., Fundamental investigations on some transformer liquids under various outdoor conditions, IEEE Transactions on Dielectrics and Electrical Insulation, 8(6), pp. 1040-1047, 2001. DOI: 10.1109/94.971463
- [14] Oommen, T. and Prevost, T., Cellulose insulation in oil-filled power transformers: Part II – Maintaining insulation integrity and life, IEEE Electrical Insulation Magazine, 22(2), pp. 5-14, 2006. <http://dx.doi.org/10.1109/MEI.2006.1618996>
- [15] Lobo-Ribeiro, A.B., Eira, N.F., Sousa, J.M., Guerreiro, P.T. and Salcedo, J.R., Multipoint fiber-optic hot-spot sensing network integrated into high power transformer for continuous monitoring, IEEE Sensors Journal, 8(7), pp. 1264-1267, 2008. DOI: 10.1109/JSEN.2008.926926
- [16] Jauregui-Rivera, L. and Tylavsky, D., Acceptability of four transformer top-oil thermal models – Part II: Comparing metrics, IEEE Transactions on Power Delivery, 23(2), pp. 868-872, 2008. DOI: 10.1109/TPWRD.2007.905576
- [17] Peña, D., Análisis de datos multivariantes, Mc Graw Hill, 2002.
- [18] Zhang, G.P., Time series forecasting using a hybrid ARIMA and neural network model, Neurocomputing, 50, pp.159-175, 2003
- [19] El-Desouky, A.A. and Elkateb, M.M., Hybrid adaptive techniques for electric-load forecast using ANN and ARIMA, IEE Proceedings Generation, Transmission and Distribution, 147(4), pp.213-217, 2000. DOI: 10.1049/ip-gtd:20000521
- [20] Feinberg, E. and Genethliou, D., Applied mathematics for restructured electric power systems – Chapter 12: Load Forecasting. Springer, Hamburg, 2005, pp. 269-285.
- [21] Hornik, K., Stinchcombe, M. and White, H., Multilayer feedforward networks are universal approximators, Neural Networks, 2(5), pp. 359-366, 1989.
- [22] IEC, Loading guide for oil-immersed power transformers, IEC 60354, Second edition, 1991.
- [23] Gelezenis, J.J., Estimation of hourly temperature data from their monthly average values: case study of Greece, Renewable Energy, 18(1), pp. 49-60, 1999.
- [24] Gómez-Luna, E., Fernando-Navas, D., Aponte-Mayor, G. and Betancourt-Buitrago, L.A., Literature review methodology for scientific and information management, through its structuring and systematization, DYNA, 81(184), pp. 158-163., 2014 DOI: 10.15446/dyna.v81n184.37066
- [25] Gomez-Luna, E., Duarte, C., Aponte, G., Guerra, J.P. and Goossen, K.W., Experiences with non-intrusive monitoring of distribution transformers based on the on-line frequency response, Ingeniería e Investigación, 35(1), pp. 55-59, 2015. DOI: 10.15446/ing.investig.v35n1.47363
- [26] Gomez-Luna, E., Aponte, G., Herrera, W. and Pleite, J., Experimentally obtaining on-line FRA in transformers by injecting controlled pulses, Ingeniería e Investigación, 33(1), pp. 43-45, 2013.

A.A. Romero-Quete, was born in Colombia in 1978. He received a degree in Electrical Engineering in 2002, from the Universidad Nacional de Colombia, and his Ph.D. from the Universidad Nacional de San Juan (UNSJ) in 2009. He is currently a researcher at the Consejo Nacional de Investigaciones Científicas y Técnicas, CONICET, at the IEE UNSJ-CONICET. His research interests are: asset management, power quality, and high voltage test techniques.
ORCID: 0000-0002-6530-852X

E.E. Mombello, was born in Buenos Aires, Argentina, in 1957. He received a B.S. in electrical engineering and a Ph.D. in electrical engineering in 1982 and 1998, respectively. He has worked for CONICET, at the Instituto de Energia Eléctrica of the University of San Juan (IEE-UNSJ), Argentina, as a lecturer and researcher since 1982. His main fields of interest are design, modeling and diagnostics of power transformers, transformer life management, electromagnetic transients in electric machines and networks, modeling of equipment, low frequency electromagnetic fields.
ORCID: 0000-0003-0425-596X

G. Rattá, was born in Italy in 1950. He received his degree in Electromechanical Engineering from the Universidad Nacional de Cuyo-Argentina in 1974. Since 1997 he has been director of the IEE-UNSJ, Argentina. Prof. Rattá is currently an Assistant Professor in the UNSJ. His research interests include transient behavior of power system components and power quality. Email: ratta@iee.unsj.edu.ar.
ORCID: 0000-0003-3663-6310

Thermal assessment of ecological tiles in physical models of poultry houses

Flavio Alves Damasceno ^a, Leonardo Schiassi ^a, Tadayuki Yanagi Junior ^a, Jairo Alexander Osorio-Saraz ^b
& Jofran Luiz-de Oliveira ^c

^a Departamento de Ingeniería, Universidad Federal de Lavras, Lavras, Brasil. flavio.damasceno@deg.ufla.br, leoschiassi@yahoo.com.br, yanagi@deg.ufla.br

^b Facultad de Ciencias Agrarias, Universidad Nacional de Colombia, Medellín, Colombia. aosorio@unal.edu.co

^c Departamento de Ingeniería Agrícola, Universidad Federal de Mato Grosso, Cuiabá, Brasil. joframluiz@gmail.com

Received: January 27th, 2015 Received in revised form: August 11th, 2015. Accepted: February 25th, 2016.

Abstract

In countries with tropical climates, such as Brazil, the high summer temperatures associated with high relative humidities are a stress factor in animal production. Excessive heat within a poultry facility causes a reduction in feed intake and production, and increased bird mortality. With the knowledge that about 75% of the radiant heat load within a facility comes from the roof, it is necessary to study alternatives that can minimize this radiation. Thus, the objective of the present work is to analyze the thermal environment inside physical broiler housing models constructed on a reduced scale (1:10), where the thermal comfort was evaluated by the black globe humidity index (BGHI) and radiant heat load (RHL). Five models built with different roofing materials were evaluated. Based on the results, it can be concluded that roofs constructed with the channel clay roofing tile (T_{B30}), natural fiber tiles with their external side painted white (T_{FVP15}) and tiles made of recycled long-life packaging (T_{LV15}), provide better thermal environments within the models.

Keywords: reduced models, thermal environment, alternative materials, thermal capacity.

Evaluación térmica de tejas ecológicas en modelos físicos de galpones avícolas

Resumen

En países con climas tropicales, como Brasil, las altas temperaturas asociadas con altas humedades relativas en el verano contribuyen como un factor de estrés en la producción animal. El calor excesivo dentro de una instalación avícola causa una reducción en el consumo de alimento y en la producción, y aumento en la mortalidad de las aves. Conociendo que el 75% de la carga de calor radiante dentro de una instalación viene desde el techo, es necesario estudiar alternativas que pueden minimizar esta radiación. El objetivo del presente trabajo fue analizar el ambiente térmico dentro de los modelos físicos de instalaciones de pollos de engorde construidos a escala reducida (1:10), donde el confort térmico se evaluó por el índice de humedad y de globo negro (BGHI) y la carga de calor radiante (RHL). Cinco modelos construidos fueron evaluados con diferentes materiales para techos. Basados en los resultados, se puede concluir que los techos construidos con tejas de cerámica (T_{B30}), placas de fibra natural pintadas de blanco con en el lado externo (T_{FVP15}) y placas de fabricadas con envases larga vida reciclado (T_{LV15}), proporcionan los mejores ambientes térmicos.

Palabras clave: modelos a escala, ambiente térmico, materiales alternativos, capacidad térmica.

1. Introduction

The use of ecological tiles is an uncommon practice despite the potential to use industrial and urban waste when manufacturing them.

For many years, in the construction and agricultural

industries, roofs have been built with traditional materials such as fiber-cement, ceramics and aluminum. Current environmental concerns have forced architects and engineers to review the materials used in civil construction, and they now favor the use of renewable and less polluting energy sources [1].

How to cite: Damasceno, F.A., Schiassi, L., Yanagi, T., Osorio-Saraz, J.A. and Luiz-de Oliveirac, J., Thermal assessment of ecological tiles in physical models of poultry houses. DYNA 83(197), PP. 114-119, 2016.

The roof most commonly used in Brazilian poultry facilities are fiber-cement tiles, which are simple to install, maintain and clean (in comparison to traditional ceramic tiles) in addition to costing less. This is mainly because the support structure is lighter and requires less labor [2,3]. Another advantage is that these tiles may also be combined with other methods, such as a reflective paint and water sprinkling on the external surface in order to decrease the temperature inside the facility [4].

In Brazil, a country with a tropical climate and high summer temperatures that receives intense solar radiation, the materials to construct roofs should provide good thermal insulation for the internal environment of the installations so they are less influenced by the outside climate [5].

According to some researchers [6], of the thermal radiation received from various parts of the installation surrounding an animal in the shade, 28% of the radiant heat load comes from the sky, 21% from the roofing, 18% from the unshaded area and 33% from the shaded area.

The current trend of studies on agricultural projects in South American countries that occupy a prominent place on the world stage, such as Brazil, Colombia and Argentina, is focused on the technical and economic evaluation of technologies seeking to increase efficiency since the global economy requires the poultry industry to be more productive and at the same time more profitable.

In tropical and subtropical countries, the importance of studying new roofing materials and especially ecological materials lies in the wealth of natural resources that these countries possess and the high incidence of direct solar radiation, especially in the summer. According to [7], incident solar radiation generates thermal discomfort, which is mostly related to the roofing material and the type of thermal cooling used, in humans and in animals housed inside production installations.

This discomfort increases the risk of work-related accidents in addition to exposing the body to various conditions such as hyperthermia, dizziness and others. Animal discomfort causes a reduction in growth performance and increased mortality rate [8].

In this context, the construction of ecological roofs has begun to use plant cellulose fiber, paper and recycled long-life packaging. These types of roofs have some advantages including: being eco-friendly, having good thermal and acoustic insulation, good mechanical strength, impermeability, and are socially inclusive and generate employment and income.

As these new eco-friendly materials are being used in the construction of roofing tiles, the agro-industrial sector requires technical information on the damping capacity of the radiant heat load provided by these materials.

Due to limited economic resources and to save time performing studies seeking to evaluate roofing materials, it is common to use physical broiler house models built on a reduced scale [9-14]. According to [12], the use of these physical models is based on the theory of similarity, which intends to establish relationships that allow for real predictions to be made based on observations of smaller models.

Therefore, the objective of this study was to evaluate the efficiency of ecological tiles to reduce the radiant heat load.

It seeks to look for improvements in the thermal environment of facilities for broilers.

2. Material and methods

2.1. Construction and field deployment of the reduced models

In order to undertake this experiment, five models at a reduced scale of 1:10 were constructed, based on the methodology used by [9,11-14] for east-west oriented poultry houses. They had different types of roofing as indicated in Table 1, were constructed according to the cross-section schematic of the models (Fig. 1) and arranged according to the overview of the experimental area (Fig. 2).

The physical models were deployed in the experimental area at the Construction and Ambience Sector of the Engineering Department, Federal University of Lavras, located in southern Minas Gerais - Brazil (21°14' S latitude and 45°00' W longitude, located 918 m above sea level). The local climate, according to the Köppen climate classification, is Cwa: a humid temperate with a dry winter.

In order to construct small-scale models, real dimensions of a commercial poultry house for production of broiler chickens were used as the basis, with a standard that is adopted worldwide (12.0 m wide, 120.0 m long, ceiling height of 3.5 m and distance between roof trusses of 5.0 m).

In the construction, 1 cm thick plywood panels were used that were reinforced by a 3 x 3 cm wood structure. After construction, the models were 1.2 m wide, 1.5 m long and had a distance between roof trusses of 0.5 m. There were three sections of 5 m that represented the distance between pillars. All constructed roofs had 2.0 meter overhangs. The central sector of physical models was used to set the measuring sensors of environmental variables.

The east and west sides of the buildings were completely closed with plywood. The side had a height of 0.5 cm, corresponding to the real sheds' existing 5 cm low walls. The buildings' floors were elevated from the soil by 0.4 m, making the air flow at floor height approximately the same, with respect to a real shed [11,12]. All reduced physical models had the ridges of the roofs oriented in the east-west direction since they are located in the southern hemisphere.

Table 1.
Treatments tested in the experiment.

Description of the Treatments	
T _{AL15}	am roof, 0.5 mm thick, 15° inclination, with no louver and a ceiling height of 3.5 m.
T _{B30}	clay roofing tile (Roman), 30° inclination, with no louver and a ceiling height of 3.5 m.
T _{VN15}	tile of cellulose fiber (recycled paper) and bitumen, 0.5 mm thick, 15° inclination, with no louver and a ceiling height of 3.5 m.
T _{FVP15}	Roofing tile of cellulose fiber (recycled paper) and bitumen, externally painted with two coats of hydrated lime, 0.5 mm thick, 15° inclination, with no louver and a ceiling height of 3.5 m.
T _{LV15}	tiles made of recycled long-life packaging, with a thickness of 6 mm, ceiling height of 3.5 m and 15° inclination, with no louver.

Source: The authors.

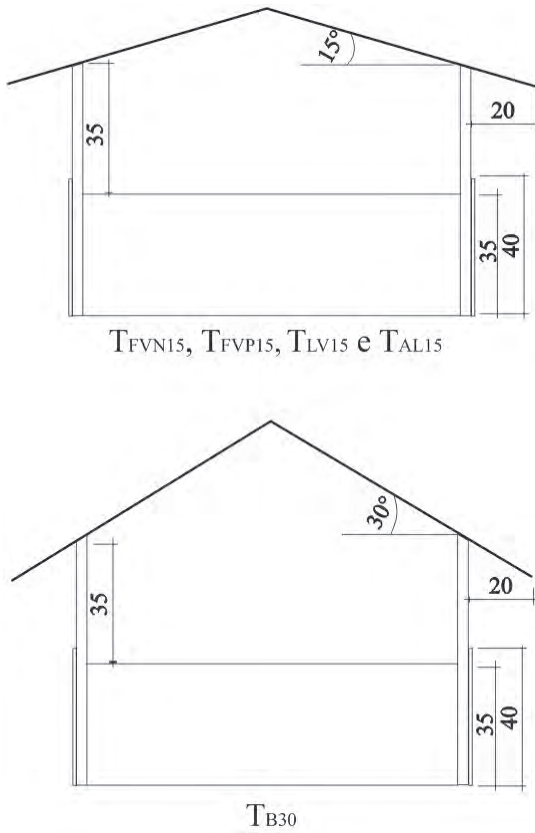


Figure 1. Schematic illustration of the cross sections of the physical models built on a reduced scale (unit: cm).

Source: The authors.

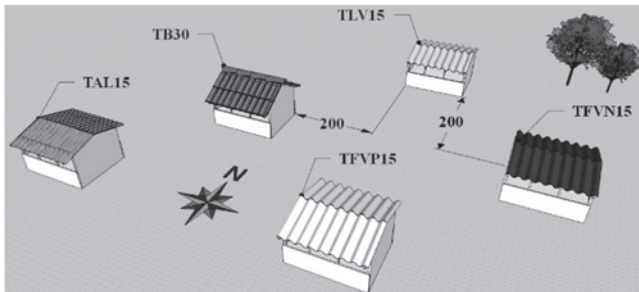


Figure 2. Overview of the experimental area and distribution of the physical models (unit: cm).

Source: The authors.

2.2. Instrumentation and measurements

The dry bulb temperature (T_{db}), black globe temperature (T_{bg}), relative humidity (RH) and air velocity (V) variables were measured in the constructed physical reduced scale models and in the external environment over a period of seven non-consecutive typical summer days, with a clear sky, between 12 and 16 h. These conditions present severe thermal discomfort conditions, at 15-minute intervals. After this process, the respective values of BGHI and RHL were calculated for the measurement times.

Inside the reduced scale physical models, sensors were installed at the height corresponding to the geometric center of the birds, i.e., 3 cm from the floor (30 cm in the actual

buildings). In the external environment, climate data was collected in a weather shelter that was installed in the experimental area, except for T_{bg} and V, whose sensors were allocated at the height corresponding to the geometric center of the birds.

2.2.1. Dry bulb, wet bulb and black globe temperatures

The thermal environment was assessed by the T_{db} , T_{bg} and RH, in which these three variables were measured by means of sensors / recorders (Hobo®, mod. U12-013, accurate to $\pm 3\%$). In each reduced model, one sensor / T_{db} , T_{bg} and RH recorder were installed. These sensors were housed inside a perforated protective container in order to prevent damage to equipment. Readings were compared to a sensor outside of the protection in order to verify any interference that the protective container had on equipment measurements. In the case of T_{bg} , only the recorder was housed in the protective container, since the temperature sensor was inserted into a plastic sphere measuring 3.6 cm in diameter and painted matte black. The recorder-sensor assembly was calibrated against a standard made of copper, with a 15 cm diameter and 0.5 mm thickness, painted black.

2.2.2. Air velocity

The air velocity was measured using a digital propeller anemometer (Kestrel® mod. 4000, accurate to $\pm 3\%$). Measurements were obtained in the vicinity of each black globe, on the same days and at the same times.

2.2.3. Climatic variables of the external environment

External climate data was collected using a weather shelter installed in the experimental area. In this shelter, sensors were installed to measure air temperature and humidity. Thus, T_{db} and RH data could be developed using measurements at 15-minute intervals, indicating the average, maximum and minimum. A black globe thermometer was installed in the external environment to quantify the BGHI and RHL of the external environment.

2.3. Determination of the thermal indices

From the T_{db} data, dew point temperature (T_{dp}), T_{bg} and V measured at the predetermined times, the respective BGHI and RHL values for each measurement time were calculated according to eq. 1 and 2, respectively.

$$BGHI = T_{bg} + 0.36 \cdot T_{dp} - 330.08 \quad (1)$$

where:

T_{bg} : black globe temperature, K;

T_{dp} : dew point temperature, K.

$$RHL = \sigma \cdot (T_m)^4 \quad (2)$$

where:

σ : Stefan-Boltzmann constant ($\sigma = 5.67 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$);

T_m : mean radiant temperature, calculated by eq. 3.

$$T_m = 100 \cdot \left[2.51 \cdot V^{0.5} \cdot (T_{bg} - T_{db}) + \left(\frac{T_{bg}}{100} \right)^4 \right]^{\frac{1}{4}} \quad (3)$$

Classification of materials used in the roofs or their designs was performed by effectiveness (ϵ), which may be defined in relation to the BGHI [18, 3, 28, 11] by eq. 4.

$$\epsilon = \frac{BGHI_{sun} - BGHI_{tested\ roof}}{BGHI_{sun} - BGHI_{aluminium\ roof}} \quad (4)$$

2.4. Experimental design and statistical analysis

We used a randomized block design (RBD), in which the treatments were arranged in a split-plot design, with seven repetitions (measurement days) The plots were attributed to the constructed reduced scale models with the different roofs, and subplots with the measurement times. The mean BGHI and RHL for the treatments were compared with the Skott-Knott test at a 5% probability. The Sisvar 4.6 software [14] was used for statistical analysis.

3. Results and discussion

Results from the analysis of variance for the response variables BGHI and RHL ($W \cdot m^{-2}$) are presented in Table 2. It was verified that for both the BGHI and RHL there was a significant difference ($p < 0.05$) for the treatments and measurement times. However, no significant difference was observed for the interaction of treatments x measurement time ($p > 0.05$).

Regarding the BGHI, the coefficients of variation for the plots and subplots were 2.64% and 1.33%, respectively; and for the RHL they were 3.37% and 2.65%. This demonstrated little variability of the observed models in relation to the means.

For both the RHL and BGHI (Table 3), the treatments that resulted in the best thermal comfort conditions inside the reduced physical models were T_{B30} , T_{FVP15} and T_{LV15} , which are statistically equal (Skott-Knott test, $p < 0.05$). The other treatments presented poorer results (Skott-Knott test, $p > 0.05$).

The results obtained for the BGHI and RHL are in accordance with those observed by [16] and [17], where T_{B30} presented the best efficiency when compared with the other materials due to its reflective capacity and porosity. It is therefore statistically equal to T_{FVP15} and T_{LV15} and different in relation to T_{FVN15} and T_{AL15} .

When considering tiles constructed of natural plant fiber sealed with bitumen, a better efficiency was observed when they were painted with two coats of hydrated lime on the external surface. This effect was similar to that encountered when using white paint on fiber-cement tiles, as observed by [16,18,19], where it was shown that the use of reflective paint on the tiles resulted in improved thermal performance.

Mean BGHI values observed in treatments T_{B30} and T_{FVP15} were in agreement with those found by [20], from 69.0 to 77.0, where adult birds have the highest productivity and best performance parameters.

Table 2.

Summary of the analysis of variance for the temperature index response variables: black globe humidity index (BGHI) and radiant heat load (RHL).

FV	DF	Mean square	
		BGHI	RHL
Blocks	6	55.34*	3330.45*
Treatments	4	22.29*	1110*
Error (a)	24	4.17	256.86
Times	16	2.36*	143.81*
Treatment x Time	64	0.33 ns	18.41ns
Error (b)	240	1.06	61.81
C.V. Error (a) %		2.64	3.37
C.V. Error (b) % =		1.33	2.65

*: significant at 5% probability (F test); ns: non-significant (F test).

Source: The authors.

Table 3.

Mean values of the radiant heat load (RHL), reduction of the RHL and black globe humidity index (BGHI) and reduction of the BGHI in relation to the external environment for the various treatments tested.

Trea t.	BGH I	Reduction of BGHI (%)	RHL (W.m ⁻²)	Reduction of RHL (%)
	76.61			
T_{B30}	a	7.09	471.23 a	20.24
T_{FVP15}	a	7.01	472.75 a	19.98
5	77.07			
T_{LV15}	a	6.54	475.49 a	19.52
T_{FVN15}	77.72			
5	b	5.75	480.45 b	18.68
	77.84			
T_{AL15}	b	5.6	479.47 b	18.84
T_{EXT}	82.46		590.82	

Source: The authors.

The RHL values for all treatments were lower than those found in [19], 498.3 ($W \cdot m^{-2}$), and [21], 502.7 $W \cdot m^{-2}$ (average of the times) for sheds with clay tiles. Reduction of the BGHI by the roofs varied from 7.09 to 5.60% for the best (T_{B30}) and worst (T_{AL15}) treatment, respectively. For the RHL, the change was 20.24% to 18.68% for the best (T_{B30}) and worst (T_{FVN15}) treatment. The RHL reduction results are within the range suggested by several authors, which is from 20 to 40% [22-25].

Fig. 3 shows the average internal and external behaviors of the BGHI and RHL, depending on the time of day for the tested treatments, and the lower and upper limits of thermal comfort for the BGHI. The lower and upper limits of BGHI that prevented the broilers from suffering stress had values of 69.1 and 77.5, respectively, and were in accordance with the recommendations reported by other studies [21,27,28].

The classification of effectivities (ϵ) of the roofs is presented in Table 4. Behavior was observed similar to that found for [11,28]. These results complement the information in the literature related to the classification of roofs made of different types of materials, especially those obtained by [3,11].

Through analysis of the BGHI, RHL and ϵ , it was verified that the roofs T_{B30} , T_{FVP15} and T_{LV15} provided the best thermal comfort conditions inside the physical reduced-scale models. This is due to the increased thermal resistance of the roof, an effect of the thickness of the material and the light color, as shown by [11].

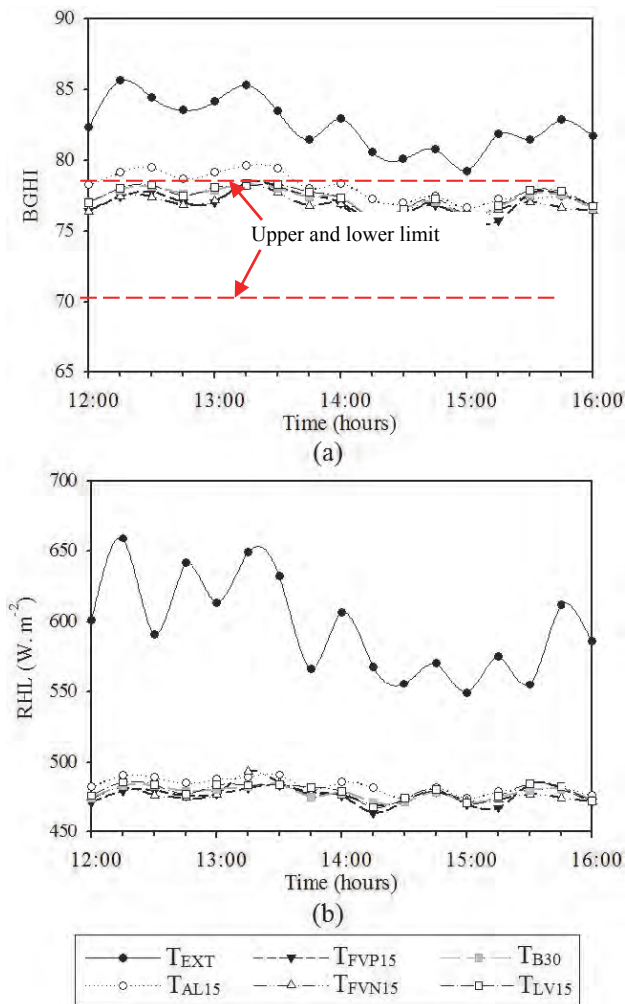


Figure 3. Mean values of the black globe humidity index - BGHI (a) and radiation heat load - RHL (b), corresponding to the height of the birds. Source: The authors.

Table 4.

Mean values and standard deviations of the effectiveness (ϵ) of the roofs in reducing the black globe humidity index (BGHI) in relation to the roof constructed with aluminum sheets.

Treatments	Effectiveness (ϵ)
T _{FVN15}	1.28 ± 0.13
T _{FVP15}	1.26 ± 0.18
T _{B30}	1.18 ± 0.12
T _{LV15}	1.16 ± 0.14
T _{AL15}	1.00 ±

Source: The authors.

4. Conclusions

The roofs constructed with tiles made of ceramic, plant fiber, painted bitumen and recycled long-life packaging provided the best thermal comfort conditions inside the physical broiler housing models built on a reduced scale (1:10). Statistically equal results were obtained. Average reduction of the black globe humidity index (BGHI) and radiant heat load (RHL) were 6.88% and 19.91%, respectively.

Acknowledgements

The authors thank the CAPES, CNPq and FAPEMIG for their financial support of the project.

References

- [1] Silva, R.B., Nääs, I.A., Moura, D.J. e Silveira, N.A., Insalubridade do trabalhador na produção animal: Uma questão de educação e informação. Segurança e Trabalho, Brasil, 2006.
- [2] Conceição, M.N., Alves, S.P., Telatin Júnior, A., Silva, I.J.O., Piedade, S.M.S., Savastano-Júnior, H. e Tonoli, G., Desempenho de telhas de escória de alto forno e fibras vegetais em protótipos de galpões. Revista Brasileira de Engenharia Agrícola e Ambiental, 12(5), pp. 536-539, 2008. DOI: 10.1590/S1415-43662008000500015
- [3] Moraes, S.R.P., Tinôco, I.F.F., Cecon, P.R. e Baêta, F.C., Conforto térmico em galpões avícolas, sob coberturas de cimento-amianto e suas diferentes associações. Revista Brasileira de Engenharia Agrícola e Ambiental, 3(1), pp. 89-92, 1999. DOI: 10.1590/1807-1929/agriambi.v3n1p89-92
- [4] Abreu, P.G. e Abreu, V.M.N., Análise termográfica da temperatura superficial de telhas. Revista Brasileira de Engenharia Agrícola e Ambiental, 15(11), pp. 1193-1198, 2011. DOI: 10.1590/S1415-43662011001100013
- [5] Abreu, P.G., Abreu, V.M.N. e Dalla-Costa, O.A., Avaliação de coberturas de cabanas de maternidade em sistema intensivo de suínos criados ao ar livre (Siscal), no verão. Revista Brasileira de Zootecnia, 30(1), pp. 1728-1734, 2001. DOI: 10.1590/S1516-35982001000700010
- [6] Bond, T.E., Morrison, S.R. and Givens, R.L., Influence of surrounding on radiant heat load of animals. Transactions of the ASAE, 12(2), pp. 246-248, 1969. DOI: 10.13031/2013.38809
- [7] Oliveira, L.M.F., Efeitos climáticos no conforto térmico animal e humano. Tese de MeSc, Departamento de Engenharia Agrícola, Universidade Federal de Lavras, Lavras, Brasil, 2004.
- [8] Damasceno, F.A., Yanagi-Junior, T., Lima, R.R., Gomes, R.C.C. e Moraes, S.R.P., Avaliação do bem-estar de frangos de corte em dois galpões comerciais climatizados. Ciência Agrotecnica, 34(2), pp. 1031-1038, 2010. DOI: 10.1590/S1413-70542010000400033.
- [9] Murphy, G.C.E., Similitude in engineering. New York: Ronal Press, 1950.
- [10] Pattie, D.R. and Milne, W.R., Ventilation air-flow patterns by use of models. Transactions of the ASAE, 9(5), pp. 646-9, 1966. DOI: 10.13031/2013.40062
- [11] Santos, P.A., Yanagi-Junior, T., Teixeira, V.H. e Ferreira, L., Ambiente térmico no interior de modelos de galpões avícolas em escala reduzida com ventilação natural e artificial dos telhados. Engenharia Agrícola, 25(3), pp. 575-584, 2005. DOI: 10.1590/S0100-69162005000300002
- [12] Jentzsch, R., Baêta, F.C., Tinôco, I.F.F., Damasceno, F.A., Cecon, P.R. e Tibiriçá, A.C.G., Predição de parâmetros térmicos ambientais no interior de modelos físicos de galpões avícolas construídos em escalas reduzidas. Interciencia, 36(1), pp. 1-5, 2011.
- [13] Jentzsch, R., Baêta, F.C., Tinôco, I.F.F., Damasceno, F.A. e Osório, J.A., Parâmetros arquitetônico-ambientais para construção e testes em modelos reduzidos, representativos de galpões avícolas, com base em similitude. Engenharia na Agricultura, 21 (1), pp. 19-30, 2013. DOI: 10.13083/1414-3984.v21n01a02
- [14] Carvalho, H.G., Materiais de cobertura e suas associações a forros e materiais isolantes no ambiente térmico de protótipos abertos e fechados com vistas a produção de frangos de corte em clima quente. Dissertação de Dr., Departamento de Engenharia Agrícola, Universidade Federal de Viçosa, Viçosa, Brasil, 2013.
- [15] Ferreira, D.F., Sistema de análise estatística para dados balanceados - SISVAR. Lavras: UFLA/DEX, 2000.
- [16] Sevegnani, K.B., Ghelfi-Filho, H. e Silva, I.J.O., Comparação de vários materiais de cobertura através de índices de conforto térmico. Scientia Agrícola, 51(1), pp.1-7, 1994. DOI: 10.1590/S0103-90161994000100001
- [17] Jácome, I.M.T.D., Furtado, D.A., Leal, A.F., Silva, J.H.V. e Moura, J.F.P., Avaliação de índices de conforto térmico de instalações para poedeiras no nordeste do Brasil. Revista Brasileira de Engenharia

- Agrícola e Ambiental, 11(5), pp. 527-531, 2007. DOI: 10.1590/S1415-43662007000500013
- [18] Kelly, C.F. and Bond, T.E., Effectiveness of artificial shade materials. *Agriculture Engineering*, St. Joseph, 39(12), pp.758-764, 1958.
- [19] Rosa, Y.B.C.J., Influência de três materiais de cobertura no índice de conforto térmico em condições de verão, para Viçosa. Tese de MSc., Departamento de Engenharia Agrícola, Universidade Federal de Viçosa, Viçosa, Brasil, 1984.
- [20] Medeiros, C.M., Baêta, F.C., Oliveira, R.F.M., Tinôco, I.F.F., Albino, L.F.T. e Cecon, P.R., Efeitos da temperatura, umidade relativa e velocidade do ar em frangos de corte. *Engenharia na Agricultura*, 13(4), pp. 277-286, 2005.
- [21] Furtado, D.A., Azevedo, P.V. e Tinoco, I.F.F., Análise do conforto térmico em galpões avícolas com diferentes sistemas de acondicionamento. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 7(3), pp. 559-564, 2003. DOI: 10.1590/S1415-43662003000300025
- [22] Santos, A.C., Baêta, F.C., Cecon, P.R. e Cardoso, R.M. Análise de diferentes bezerreiros individuais móveis, para região de Viçosa. *Engenharia na Agricultura*, 2(7), pp. 1-8, 1993.
- [23] Turco, S.H.N., Baêta, F.C., Costa, P.M.A., Cardoso, R.M. e Cecon, P.R., Modificações das condições ambientais de verão em maternidades de suínos. *Engenharia na Agricultura*, 3(11), pp. 1-12, 1994.
- [24] Abreu, P.G., Baêta, F.C., Soares, A.R., Abreu, U.M.N. e Maciel, N.F., Utilização de piso aquecido eletricamente na criação de aves. *Engenharia na Agricultura*, 4(12), pp. 1-19, 1995.
- [25] Baêta, F.C., Souza, C.F., *Ambiência em Edificações Rurais: Conforto Animal*. 2ª Ed. Viçosa-MG: Editora UFV, 2010.
- [26] Teixeira, V.H., Estudo dos índices de conforto em duas instalações de frango de corte para as regiões de Viçosa e Visconde do Rio Branco, MG. Tese de MSc. Departamento de Engenharia Agrícola, Universidade Federal de Viçosa, Viçosa, Brasil, 1983.
- [27] Tinôco, I.F.F., Resfriamento adiabático (evaporativo) na produção de frangos de corte. Tese de MSc, Departamento de Engenharia Agrícola, Universidade Federal de Viçosa, Viçosa, Brasil, 1988.
- [28] Ferreira-Júnior, L.G., Yanagi-Junior, T., Damasceno, F.A., Silva, E. e Silva, G.C.A., Ambiente térmico no interior de modelos físicos de galpões avícolas equipados com câmaras de ventilação natural e artificial. *Engenharia na Agricultura*, 17(1), pp. 166-178, 2009. DOI: 10.13083/1414-3984.v17n03a01

F.A. Damasceno, is adjunct professor at the Universidade Federal de Lavras (UFLA), Brazil. PhD in Environment Control from the Federal University of Viçosa (UFV), Brazil with training in Biosystems and Agricultural Engineering from the University of Kentucky, as a visiting researcher. MSc. of Environment Control from the UFLA in 2008, Sp. Management Specialist and Environmental Management in Agribusiness from UFLA in 2008, and BSc. in Agricultural Engineering from UFLA in 2006. He has publication scientific papers in conferences and national and international journals. ORCID: 0000-0002-4181-0611

L. Schiassi, is professor at the Federal University of Lavras (UFLA), Brazil, in the Engineering Department. Researcher in COAMBI - Research Group Constructions and Ambience in Biosystems - DEG / UFLA. PhD in Rural and Ambience Constructions, Master of Systems Engineering. He has experience in Agricultural Engineering with an emphasis on Rural Construction Engineering, working mainly in the areas of construction and ambience, Building Materials, Bioclimatology, Noise, Animal Welfare and Human, Evaluation and Building Systems Modeling and Ambience, Analysis Fuzzy Systems and Neural Networks. ORCID: 0000-0002-8284-7496

T. Yanagi-Junior, has a BSc degree in Agricultural Engineering in 1992 and is a Professor at the Federal University of Lavras, Brazil. His MSc. degree is in Agricultural Engineering from Federal University of Lavras in 1995 and PhD in Agricultural Engineering from the Federal University of Viçosa, Brazil, in 2002. He has experience in Agricultural Engineering, and focuses on Agricultural Constructions and Environment. ORCID: 0000-0001-9653-205X

J.A. Osorio-Saraz, has a BSc. degree in Agricultural Engineering in 1998, a Post Graduate Degree in Environmental Legislation in 2001 and a MSc. in Materials Engineering in 2006, all from the Universidad Nacional de Colombia, Medellín, Colombia). In 2011 he gained a Dr. degree in Rural Constructions from the Federal University of Viçosa, Minas Gerais state, Brazil. Since 2003 he has worked as a professor at the Universidad Nacional de Colombia, teaching and advising in the following areas: Design of Livestock Housing, Materials Technology for Livestock Housing, air quality and animal welfare, thermal comfort for animals, use of the CFD tool to predict air motion patterns in agro industry facilities. He is member of AMBIAGRO (Center for Research in Agro-Industrial Ambience and Engineering Systems) at the Department of Agricultural Engineering, Federal University of Viçosa and of the research group in agricultural engineering (AI in Colciencias Colombia) at the UNAL. He is currently the Dean of the Faculty of Agricultural Sciences at UNAL Medellín, and has been since 2012. ORCID: 0000-0002-4358-3600

J. Luiz-de Oliveira, has a BSc. degree in Agricultural Engineering in 2010 from the Federal University of Viçosa, Minas Gerais state, Brazil. Since 2010 he has worked as a Professor at the Universidade Federal de Mato Grosso (UFMT), Brazil, teaching and advising in the following areas: Thermal Comfort in Livestock Facilities and Materials technologies for Thermal Control Improvement. He is member of AMBIAGRO (Center for Research in Agro-Industrial Ambience and Engineering Systems) at the Department of Agricultural Engineering, Federal University of Viçosa and ECOAGRI (Engineering and Sustainability in Agriculture) at UFMT. ORCID: 0000-0002-9093-3773



UNIVERSIDAD NACIONAL DE COLOMBIA

SEDE MEDELLÍN
FACULTAD DE MINAS

Área Curricular de Ingeniería
Geológica e Ingeniería de Minas y Metalurgia

Oferta de Posgrados

Especialización en Materiales y Procesos
Maestría en Ingeniería - Materiales y Procesos
Maestría en Ingeniería - Recursos Minerales
Doctorado en Ingeniería - Ciencia y Tecnología de
Materiales

Mayor información:

E-mail: acgeomin_med@unal.edu.co
Teléfono: (57-4) 425 53 68

Statistical performance of control charts with variable parameters for autocorrelated processes

Oscar Oviedo-Trespalacios ^{a,b} & Rita Peñabaena-Niebles ^a

^a Departamento de Ingeniería Industrial, Universidad del Norte, Barranquilla, Colombia. oooviedot@gmail.com

^b Centre for Accident Research and Road Safety – Queensland (CARRS-Q), Institute of Health and Biomedical Innovation (IHBI), Queensland University of Technology (QUT), Kelvin Grove, Australia

Received: February 25th, 2015. Received in revised form: February 29th, 2016. Accepted: March 9th, 2016.

Abstract

Chemical process and automation of data collection for industrial processes are known to result in auto-correlated data. The independence of observations is a basic assumption made by traditional tools that are used for the statistical monitoring of processes. If this is not adhered to then the number of false alarms and quality costs are increased. This research considers the use of variable parameters (VP) charts in the presence of autocorrelation. The objective is to determine the impact on the detection velocity and false alarms of different degrees of autocorrelation and their interaction with process conditions—the aim being to effectively select parameters. The VP chart showed improvements in its performance when detecting large average runs as the autocorrelation coefficient increased in contrast with the traditional variable sample interval (VSI) and \bar{x} quality systems. This research demonstrates the superiority of variable parameters quality monitoring architectures over traditional statistical process monitoring tools.

Keywords: Autocorrelation, Variable parameters control charts, Statistical Quality Monitoring, Simulation, Quality

Desempeño estadístico de cartas de control con parámetros variables para procesos autocorrelacionados

Resumen

Procesos químicos y automatización en la recolección de datos en procesos productivos son reconocidos por producir datos autocorrelacionados. La independencia de las observaciones es uno de los supuestos básicos de las herramientas tradicionales para el monitoreo estadístico de procesos, omitirlo hace que se incremente el número de falsas alarmas y los costos de calidad. Esta investigación considera a través de técnicas de simulación, la utilización de cartas de control con parámetros variables (VP) en presencia de datos autocorrelacionados con el objetivo de determinar el impacto en la velocidad de detección y falsas alarmas de diferentes grados de autocorrelación y su interacción con diferentes condiciones de proceso como la varianza y longitudes del corrimiento de la media, en búsqueda de realizar una selección efectiva de parámetros. La carta VP mostro mejoras en su desempeño en la detección de grandes corrimientos de media en la medida que incrementaba el coeficiente de auto correlación y sostenidamente mejores resultados en una mayor cantidad de condiciones de simulación, en contraste al sistema VSI y \bar{x} tradicional. Esta investigación demuestra la superioridad de sistemas de calidad basados en esquemas con parámetros variables comparado con las técnicas tradicionales de control estadístico de procesos.

Palabras clave: Autocorrelación, Cartas de control con Parámetros Variables, Monitoreo Estadístico de la Calidad, Simulación, Calidad

1. Introduction

Traditional tools used in the statistical monitoring of processes assume no correlation between consecutive

observations of the quality characteristic being controlled. This assumption is violated if there is a set of data that shows a trend towards moving in moderately long “runs” to both sides of the mean [1]; this is known as memory process.

How to cite: Oviedo-Trespalacios, O. and Peñabaena-Niebles, R., Statistical performance of control charts with variable parameters for autocorrelated processes. DYNA 83(197), pp. 120-128, 2016.

Specifically, the effect of violating this assumption in traditional control charts such as the \bar{x} chart will produce an increase in out of control false alarms [2]. Similarly, if the correlation is negative, it is possible that the control chart is not detecting disturbances or assignable causes different to the natural variation of the process [3].

Barr [4] demonstrated that correlation between data means that Shewhart control charts are ambiguous tools to identify assignable causes. This means that it is necessary to invest efforts into their adjustment or to develop new monitoring alternatives. The strategies that were designed to handle autocorrelation in productive processes include five procedures: (i) Modifying the traditional control chart parameters, (ii) Adjusting time series models to use normal residuals in typical control charts, (iii) Applying transformations that correct the correlation in the data, (iv) Forming an observation vector that facilitates the application of multivariable techniques, and (v) Applying artificial intelligence techniques to discover patterns in the data.

The first study on the procedure of varying parameters in control charts was undertaken by Reynolds, et al. [5], who suggested modifying the sampling interval (h) according to the performance observed for the process. The second parameter considered in the literature was the size of the samples (n), which was investigated by Prabhu, et al. [6], and the factor that determines the width of the control limits (k) that was studied by Costa [7]. Similarly, some models, such as the one proposed by Mahadik [8], include the warning limit coefficient (w). Initially, each of these variables was considered independently, and subsequently, there have been schemes in which the sampling interval and size of the sample change simultaneously. The concept of varying all the parameters simultaneously can be attributed to Costa [7], who used possible maximum and minimum values to adjust all the parameters based on information of the previous sample; this type of chart is currently known as an adaptive chart or a variable parameter (VP) chart.

The present research work aims to use simulation to evaluate the robustness of a VP chart process-control system in scenarios in which observations are autocorrelated. In order to do so, different design parameters in the control chart will be considered, including the following: control coefficients; sample size and interval; and process conditions such as the variance, autocorrelation coefficient, and run length.

VP control charts require the use of new performance indicators, given that the average run length (ARL), which is the most frequently used indicator, is not adequate for adaptive charts because it has no constant sampling intervals and sample sizes. Accordingly, the following indicators were used to evaluate the statistical performance of the VP chart: the average number of observations until a signal is emitted (ANOS), the average time between the point in which an average run occurs and the emission of a signal (AATS), and the average number of false alarms per cycle (ANFA). Including a metric for false alarms improves the performance evaluation proposals of VP charts in environments with autocorrelated observations because the literature that has been revised only lists comparisons based on the detection velocity of an average run.

This paper follows a structure that primarily includes a

presentation of the variable parameter chart model. This is followed by a historical review of the studies that are dedicated exclusively to analyzing the statistical performance of control chart models with variable parameters in environments with univariate data with correlations between observations. Next, the characteristics of the autocorrelated quality variables are presented, and we describe the methodology used to evaluate the performance of VP charts with different process conditions and design parameters for the control chart in the presence of autocorrelation. Finally, we present the results of the simulation, making a comparison between the \bar{x} chart and the VSI chart, and then state our conclusions.

2. Control Charts with Variable Parameters (VP)

When designing a VP chart, as well as a traditional \bar{x} chart, parameters such as the sample size (n), sampling interval (h), and control limit and warning limit coefficients (k , w) must be taken into account. These parameters vary between two values, minimum and maximum, the selection of which reduces or increases the strictness of the control. Consequently, the selection of the parameters must fulfill the following conditions: $n_1 < n_2$, $h_1 > h_2$, $w_1 > w_2$ y $k_1 > k_2$, in such a manner that the scenario with values (n_1 , h_1 , w_1 , k_1) corresponds to the relaxed control and the scenario with values (n_2 , h_2 , w_2 , k_2) corresponds to the strict control.

The VP chart is divided into three different regions: the central region that is bounded by the lower warning limit and the upper warning limit (LAI, LAS, for their initials in Spanish); the warning region that is bounded by the areas between the lower control limit up to the lower warning limit (LCI, LAI, for their initials in Spanish) and the upper warning limit up to the upper control limit (LAS, LCS, for their initials in Spanish); and the action region that is bounded by the values that surpass the upper control limits (LCS, for its initials in Spanish) and the values that are lower than lower control limit (LCI) (see Fig. 1).

The decision policy states that the position in which each sample falls determines which set of control parameters (relaxed or strict) will be used in the following sample. Equation (1) follows:

$$(n_i, h_i, k_i, w_i) = \begin{cases} (n_2, h_2, k_2, w_2) & \text{if } w < Z_{i-1} < LCS \\ (n_1, h_1, k_1, w_1) & \text{if } -w < Z_{i-1} < w \\ (n_2, h_2, k_2, w_2) & \text{if } LCI < Z_{i-1} < -w \end{cases} \quad (1)$$

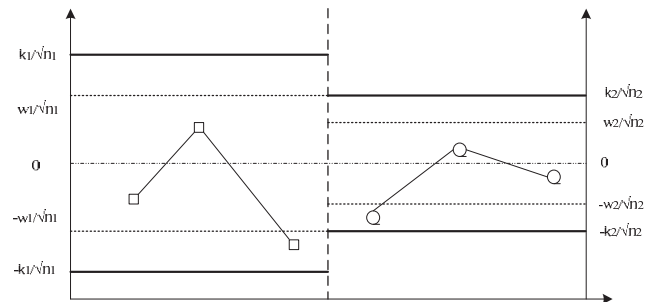


Figure 1. Control chart model with variable parameters (VP)

Source: The authors

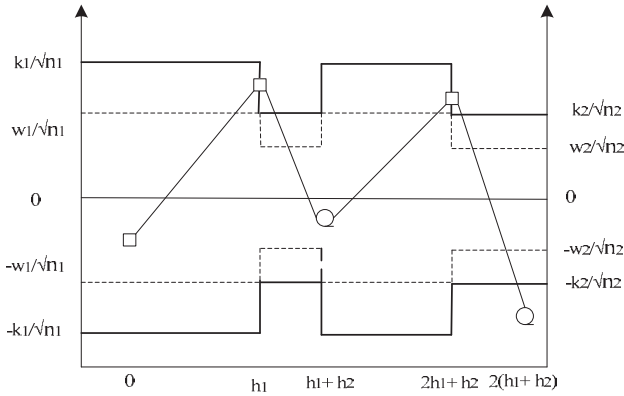


Figure 2. Control chart model with variable parameters (VP)
Source: The authors

On the one hand, if the point falls in the central region, control is reduced, the sample size must be small n_1 , and the sampling interval and control and warning coefficients must be large: h_1 , k_1 , and w_1 . On the other hand, if the point falls in the warning region, control is increased, resulting in a large sample size n_2 , and the sampling interval and control and warning coefficients must be small: h_2 , k_2 , and w_2 (see Fig. 2). Finally, if the point falls in the action region, the possible occurrence of an assignable cause must be investigated, and if pertinent, an intervention must be initiated.

To implement the VP chart, a statistical design must be used by selecting parameters that optimize the statistical properties of interest (see equation 1). However, when it is preferred that the costs associated with the control process are low, without taking into account the loss of statistical characteristics of the process, we refer to this as the economical design of a control chart. When it is desirable to select parameters that balance the statistical behavior and costs of the process, we are referring to an economical-statistical design of a control chart. Consequently, the type of design is based on what is being sought: cost reduction and high performance in detecting imbalances in the process or equilibrium between these two variables [9, 10]. Identically to the traditional \bar{x} chart, the VP chart emits a signal to stop the process when an observation is found to be outside of the control limits [7].

3. Literature Review

A limited number of published studies address the problem of autocorrelation in the performance of charts with variable parameters. Among the first studies is the one conducted by Reynolds, et al. [11], which uses the efficiency shown by control chart schemes with a variable sampling interval (VSI) as a reference to detect changes in the mean more quickly than a fixed model does, thereby exploring the impact of the presence of autocorrelation on the chart performance. This study was motivated by the assumption that the correlation between observations would be more evident in this type of chart when having shorter sampling intervals than in those with the traditional scheme. Experimentally, a first order, autoregressive, temporal, series model AR (1) was selected to model observations;

subsequently, with the help of a Markovian process, the properties of the VSI scheme were contrasted with those of the traditional scheme. As a result, as the autocorrelation levels increased, no significant differences were found between the performance of the variable scheme and those of the traditional scheme.

Prybutok, et al. [12], compared the response time of the fixed scheme of traditional \bar{x} charts with that of the proposal of control with VSI proposed by Reynolds, et al. [5] in the presence of autocorrelation. In their research, they explored the impact of utilizing a policy of pre-establishing control limits based on theoretical parameters that are typically determined as the three-sigma limits (σ_x) for data that follow a standard normal distribution. This was contrasted with a policy based on calculating control limits that understood the first 25 samples according to Montgomery [13]. The limits calculated are based on the theoretical relationship between the autocorrelation process and its (σ_x) according to the model proposed by Wardell, et al. [14]. These authors explain how σ_x depends, for certain types of autocorrelated processes, on the parameter ϕ and on the standard deviation (σ_a) [see Equation (2)]. The use of values different from the σ_a values in the control charts was also considered a method of adjusting the chart to the number of false alarms detriment.

$$\sigma_x = \sqrt{\frac{1}{1-\phi^2}} \sigma_a \quad (2)$$

The results of the present study reveal that adjusting the control limits in a fixed sampling scheme for large values of ϕ helps to reduce the rate of false alarms. With moderate autocorrelation levels, the pre-established process limits demonstrate a better performance, identifying when the process is out of control in contrast with the calculated limits; however, when the process is under control, the result that comes from using a predetermined scheme shows a considerable increase in the number of false alarms. In processes with high levels of correlation between observations, the pre-established limits are more effective in maintaining a low rate of false alarms but are inefficient in detecting changes in the mean. In general, a variable sampling scheme is beneficial for statistical monitoring because it improves the average time before emitting a stoppage signal when the process is under control.

Zou, et al. [15] indicated years later that the literature using charts with variable parameters is scarce; specifically, they mention that there is still work to be undertaken on the dependence of the correlation between observations on their distance at the time of sampling. This is because, in a VSI scheme there is uncertainty regarding the interval length and also because, in many cases, it does not match the natural periods of the process. In addition, these authors state that very little is known about the level of advantage offered by such a scheme, and there is still no correct way of estimating the control parameters despite the impact they have in determining the power of the charts. This study establishes the VSI chart parameters to take samples in fixed time periods, limiting the study to cases in which the same power of the original scheme was registered with independent data. It is, therefore, possible to compare their performance with the process modeled through an AR(1) process. The results

show that VSI charts with a fixed time and sampling rate, despite providing implementation benefits, do not show improvements in the monitoring of autocorrelated processes within the levels evaluated $0.4 < \phi < 0.8$.

Lin [16] evaluated the performance of the VP control charts in the presence of autocorrelated observations, through a comparative study between the traditional \bar{x} chart, VSI, VP, and the cumulative sum chart (CUSUM) in an AR(1) process. The results indicate that, in principle, high levels of autocorrelation require more time and samples to detect changes in the process mean. Of the different \bar{x} chart schemes studied (traditional, VSI, and VP), the VP chart has a substantially lower AATS to detect small changes in the mean when the correlation is not very high. This is because it appears to become ineffective as the variable parameters increase. The CUSUM scheme, in contrast with the VP scheme, has a better detection capacity and also exhibits the lowest sampling costs for the detection of small changes. When the autocorrelation increases, this advantage becomes more important. However, to detect large changes in the mean, the VP chart is generally faster than the CUSUM. The correlation levels considered in this study were $0.4 < \phi < 0.8$, and the running magnitude was $0.5 < \delta < 2$.

Recently, Costa and Machado [17] developed comparisons between the detection velocities of an out-of-control, first-order, autoregressive process AR(1) whilst following a Markovian approach of a VP chart and a double sampling chart scheme. Changes in the mean between $0.5 < \delta < 2.0$ and correlation levels of $\phi=0.4$ and $\phi=0.8$ were considered. The results obtained reveal that: i. With high ϕ and variance proportion levels due to the random average ψ that is defined in Equation (3), the use of VP schemes is not justified because the change in the temporal detection efficiency of an assignable cause in the process is marginal; ii. The control chart with variable parameters has a better performance when the values of the small samples are used to choose the size of the next sample in the process instead of the state of the process; and iii. The double sampling scheme works better when the process never sends signals before proceeding to the second stage of sampling.

$$\psi = \frac{\sigma_\mu^2}{\sigma_x^2} = \frac{\sigma_\mu^2}{(\sigma_\mu^2 + \sigma_\varepsilon^2)} \quad (3)$$

4. Autocorrelated quality variables

The usual assumption that generally justifies the use of control chart systems is that the data generated by the process of the quality characteristic of interest X are normally distributed and independent, with a mean μ and known variance σ^2 [11]. For this study, we will operate under the assumption that there is a correlation between the observations. In order to model it, a temporarily discretized random variable has to be considered to describe t -th element is contained in the following Equation (4).

$$X_t = \mu + \varepsilon_t \quad (4)$$

where $X(t)$ is a stochastic process $I = \{X(t), t \in I\}$ is a family of indexed random variables, where I is the set of discrete indices, (e.g., $I = \{\dots, -2, -1, 0, 1, 2, \dots\}$). In the

model of control charts, this implies that $\{X_t\}_{t \in I}$ contains all the discrete points in time t [18].

We refer to autocorrelation when the linear association between observations of the process is significant. For processes in which the mean and covariance is finite and independent of time and X_t , observations are generated at a time t ($t = 1, 2, 3, \dots$), the autocorrelation in the delay k is given by Equation (5) [19].

$$\rho_k = \frac{E[(X_t - \mu)(X_{t+k} - \mu)]}{\sigma_x^2} \quad (5)$$

Autoregressive models are frequently used in the literature to represent the correlation of productive processes. Gilbert, et al. [19] demonstrated that there are no systematic works cited in the literature that directly address the monitoring of different autoregressive models. This approach is characterized by the fact that the behavior of a variable in a specific moment in time depends on the past behavior of the variable. Therefore, the value taken by the variable at time t can be written linearly with the values taken by the variable at times $t-1, t-2, t-3, \dots, t-p$. In this scheme, p in the term AR(p) indicates the delays taken into account to describe or determine the value of x_t . If the dependency relationship is established with the p previous delays, the process will be first order autoregressive AR(1), according to Equation (6).

$$AR(1) x_t = \phi_1 x_{t-1} + \varepsilon_t \quad (6)$$

where ε_t is a white noise process, with a mean of 0, a constant variance of σ_ε^2 , and zero covariance, and ϕ corresponds to the autoregressive parameter.

5. Methodology

This study was conducted by simulation to compare the performance of VP charts in monitoring autocorrelated processes under different process conditions and design control chart parameters. The following performance indicators were used: the AATS, ANOS out of control, and the ANFA. To develop the simulations, the methodology was organized as follows:

5.1. Modeling of the AR(1) series

To simulate the autocorrelation of the process through an autoregressive model, we take into account the considerations made by Reynolds, et al. [11] and Lin [20], in which observation X_i can be written as $X_i = \mu_i + \varepsilon_i$, $i = 1, 2, 3, \dots$, where μ_i is the main random process of the sample taken in time t_i and ε_i is the random error. To include an AR(1) process, the expression developed in (6) is included in Equation (4), resulting in:

$$\mu_i = (1 - \phi)\xi + \phi\mu_{i-1} + \alpha_i, \quad i = 1, 2, 3, \dots, n \quad (7)$$

In Equation (7), ϕ represents the correlation level between μ_i and μ_{i-1} . The parameter ξ corresponds to the global mean of the process, specifically, $\xi = E(\mu_i)$. The variable α_i corresponds to a random shock that is normally

distributed with a mean of 0 and a variance of σ_α^2 and which is independent from other sources of random error. The initial value of the mean μ_0 is assumed to be normally distributed with a mean ξ and variance σ_μ^2 . Reynolds, et al. [11] demonstrated that μ_i follows a normal distribution with mean ξ and variance $\sigma_\mu^2 = \sigma_\alpha^2/(1 - \phi^2)$. As a result, the distribution of X_i has mean ξ and variance $\sigma_X^2 = \sigma_\mu^2 + \sigma_\varepsilon^2$. Similarly, Lu and Reynolds [21] considered that σ_μ^2 can be seen as the long-term variability, while σ_ε^2 is a combination between the short-term variance and measurement errors. Subsequently, Lin [20] simplifies the model by assuming a ratio ψ as the proportion of the process variance due to the random mean μ_i , where $0 \leq \psi \leq 1$. With this step, we find that the proportion of the variance due to the random error ε_i is $1 - \psi$.

5.2. Definition of the process parameters

For this study, the process parameters considered were the autocorrelation coefficient, the proportion of the variance of the process due to random mean μ_i , the occurrence of the assignable cause, and the average run length. The autocorrelation coefficient with which we will work is positive. This follows the recommendation by Reynolds, et al. [11] regarding the scarce presence of processes with a negative correlation between observations in the manufacturing industry. Because the number of simulation runs must remain computationally manageable, the values of the autocorrelation coefficient were limited to $\rho = 0.2, 0.4, 0.6, 0.8, 0.99$, ensuring that a wide range of values were included. The proportions of the process variance due to random mean μ_i used follow the Lin's [20] recommendation, who defined three levels: $\psi = 0.2, 0.5, 0.9$.

The occurrence of the assignable cause or the point at which the mean of the process suffers a run of known magnitude δ , is determined randomly through an exponential distribution with mean $1/\lambda$. It is noteworthy that the model of a single assignable cause with a known effect does not respond to real process conditions; however, it provides us with an acceptable and quasi-optimal approximation to carry out the economic and statistical designs in a more realistic process that is subject to multiple assignable causes [22].

The running length uses changes in the absolute value δ , in an attempt to detect increments and decreases in the value of the process mean. This is allowed because the formulation of the \bar{x} control charts for both sides of the process mean is exactly the same [22]. Increments of $\delta = 0.5\sigma_x, 1\sigma_x, 2\sigma_x, 3\sigma_x$ are appropriate to establish patterns of change in the behavior; in addition, they take advantage of the values proposed by Prybutok, et al. [12] by allowing the evaluation of small changes in the mean ($0.5\sigma_x$) and widening the spectrum of values considered by Lin [20], who changed the values without considering the natural variance of the process. Particularly, in this study, we will consider run lengths of $0.5\sigma_x$ as low, $1\sigma_x$ medium, $2\sigma_x$ as high, and $3\sigma_x$ as very high.

5.3. Defining the chart parameters

The interest of this study is to evaluate the impact of the autocorrelation coefficient on the performance of the VP

chart. The objective is to offer a guide to its performance under certain combinations of chart parameters. To select the levels over which the simulation will be developed, we took into account the study performed by Lin [20]. The values and combinations of h_1, h_2, n_1, n_2, k_1 , and k_2 are listed in Table 1.

Mathematically, we can express the control chart according to equations (8-9), where ξ_0 is the mean of the process when it is under control, $\sigma_{\bar{x}_l}$ is the standard deviation of \bar{X} , k_1^* is the control limit parameter, and w_1^* the warning limit parameter with sample size n_1^* , where $k_1^* \geq w_1^*$ for the case $l = 1, 2$.

$$UCL_l = \xi_0 + k_1^* \sigma_{\bar{x}_l} \text{ y } LCL_l = \xi_0 - k_1^* \sigma_{\bar{x}_l} \quad (8)$$

$$UWL_l = \xi_0 + w_1^* \sigma_{\bar{x}_l} \text{ y } LWL_l = \xi_0 - w_1^* \sigma_{\bar{x}_l} \quad (9)$$

Selecting the initial policy to be strict or relaxed is conducted randomly to avoid the effect it might have on the detection capacity of the chart in the presence of false alarms.

5.4. Performance indicators

We take the model proposed by Lin [16] and Lin [14], which uses the average number of observations until a signal is emitted (ANOS) and the average adjusted time calculated from the point in which an average run occurs and a signal is emitted (AATS) when the process is out of control as performance metrics for variable parameter schemes. An important contribution of this research, in contrast with the work undertaken by Lin [14], is that the values of false alarms are not maintained constant and that they are considered a response variable of the simulations. To that end, we take the model by Franco, Costa, and Machado (2012), which records the average number of false alarms per cycle (ANFA). In general, it is desirable for ANFA values to be small in order to reduce the frequency of false alarms [20].

The simulation was executed by programming an application in MATLAB 8.1 and programming 32,400 runs, for each combination of the autocorrelation coefficient/proportion of the process variance due to random mean μ_i /sampling interval/sample size/control coefficient/run length, following the logic described in Fig. 3. This means a new data series is simulated five times in each of the experimental conditions and stopped until the chart emits a signal to warn that the process is out of control and recording false alarms. Finally, the values of AATS, ANOS out of control, and ANFA are calculated as criteria to measure the performance of the control chart.

Table 1.
Parameters of the VP chart to be evaluated

Sampling interval (h1, h2)	Sample size (n1, n2)	Control limit coefficient (k1, k2)
(0.75, 0.1)	(1, 5)	(3.5, 2.8)
(1.25, 0.1)	(1, 10)	(3.5, 2.9)
(1.9, 0.1)	(1, 20)	(3.5, 3)
	(3, 5)	(3.5, 3.1)
	(3, 10)	(3.5, 3.2)
	(3, 20)	

Source: The authors

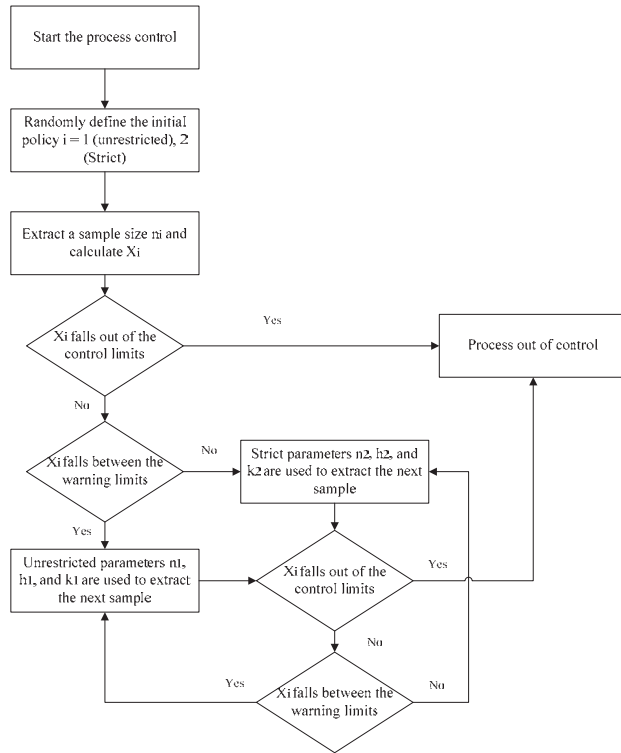


Figure 3. Process of out-of-control detection for X_i processes
Source: The authors

5.5. Simulation conditions

6. Results and discussion

The results for AATS, ANOS out of control, and ANFA for each autocorrelation coefficient/proportion of the process variance due to random mean μ_i /sampling interval/sample size/control coefficient/run length were calculated. The chart performance was evaluated by distinguishing between the different run lengths, considering the impact of increments in the coefficient of correlation and the different levels of variance. The appendix shows the results for different simulation conditions.

6.1. Short run length $\delta=0.5\sigma_x$

The chart's detection capacity improves as the autocorrelation coefficient increases because the AATS and ANOS values decrease. However, the rate of false alarms increases with the coefficient of autocorrelation, which is consistent with the results of tests conducted by Harris and Ross [23], Mastrangelo [2], and Barr [4]. As the proportion of the process variance due to random mean μ_i increases, the AATS and ANOS results decrease while the ANFA increases. Regarding the VP chart parameters, the results with larger n_2 values (e.g., $n_2 = 20$) and lower k_2 values (e.g., $k_2 = 2.8$ and 2.9) have the lowest AATS and ANOS. The highest ANFA values were found with lower k_2 values (e.g., $k_2 = 2.8, 2.9$, and 3.0).

6.2. Medium run length $\delta=1\sigma_x$

The detection capacity of the chart is not affected by increments in the autocorrelation coefficient. However, the

rate of false alarms was increased as the autocorrelation coefficient increased. The highest value for the process variances due to the random mean μ_i (e.g., $\psi = 0.9$) is associated with a higher AATS, ANOS, and ANFA. There appears to be no difference between $\psi = 0.1$ and $\psi = 0.5$. The prevalence of the VP chart's better performance was found for (h_1, h_2) values = $(0.75, 0.1)$, higher n_2 values (e.g., $n_2 = 20$), and lower k_2 values (e.g., $k_2 = 2.8, 2.9, 3.0$). The highest ANFA values were found for lower k_2 values (e.g., $k_2 = 2.9$ and 3.1) and $(n_1, n_2) = (1, 20)$.

6.3. High run length $\delta=2\sigma_x$

The detection capacity of average runs of the VP chart and the presence of false alarms exhibits a growing trend because the autocorrelation coefficient and the process variance increase due to the random mean μ_i . The design parameters of the VP chart with the best performance in AATS and ANOS were $(h_1, h_2) = (0.75, 0.1)$ and $n_1 = 3$. Low ANOS values were also found $[(h_1, h_2) = (1.25, 0.1)$ and $n_1 = 3]$. The highest ANFA values were found for $(h_1, h_2) = (1.9, 0.1)$, $n_1 = 1$ and the lowest k_2 values (e.g., $k_2 = 2.8, 2.9$).

6.4. Very high run length $\delta=3\sigma_x$

The detection capacity of the chart was not affected by autocorrelation coefficient increments and the proportion of the variance due to random mean μ_i ; however, the ANFA did exhibit increments. The prevalence of better performance of the VP chart was found for (h_1, h_2) values = $(0.75, 0.1)$ and $n_1 = 3$, and the same result was found for the high run length case. The number of false alarms increased with h_1 values = 0.75 and 1.25 , $(n_1, n_2) = (3, 20)$ and $h_2 = 2.8$ y 3 .

6.5. Impact of the Autocorrelation in the VSI and \bar{x} Charts

For short run lengths (e.g., $\delta = 0.5\sigma_x$) the VSI control chart consistently exhibited better performance levels to detect autocorrelation coefficients close to 0.99 ; however, the highest ANFA values were also found. In the \bar{x} chart, the increments in autocorrelation do not improve performance in terms of the AATS and ANOS out of control; however, there is a tendency to increase false alarms. Better \bar{x} chart performances the are exhibited in low ψ levels.

With medium, high, and very high run lengths (e.g., $\delta = 1\sigma_x, 2\sigma_x, 3\sigma_x$), the performance of the VSI chart appears not to be affected by increments in the autocorrelation coefficient for any of the metrics being studied (AATS, ANOS, and ANFA). This phenomenon also occurs in the \bar{x} chart for high and very high run lengths (e.g., $\delta = 2\sigma_x, 3\sigma_x$); however, for a value of $\delta = 1\sigma_x$, the \bar{x} chart improves its performance (AATS and ANOS) and increases the ANFA as the autocorrelation coefficient increases.

In the \bar{x} chart for the cases with $\delta = 0.5\sigma_x$ and $3\sigma_x$, the proportion of the variance causes a reduction in performance in terms of the ANFA, while it has no effect on the AATS and ANOS. In contrast, the VSI chart for the intermediate level of ψ exhibits the worst performance. The best results for all the conditions simulated for the VSI chart were obtained with the combination of sampling intervals $(h_1, h_2) = (0.75, 0.1)$.

Table 2

AATS, ANOS, and ANFA values for the VP, VSI, and X charts for $\psi = 0.5$

Adaptive Parameters	$\delta = 0.5$			$\delta = 1$			$\delta = 2$		
	VP	VSI	X	VP	VSI	X	VP	VSI	X
(h_1, h_2)	(0.75, 0.1)	(1.25, 0.1)	(1, 1)	(1.25, 0.1)	(1.25, 0.1)	(1, 1)	(0.75, 0.1)	(1.25, 0.1)	(1, 1)
(n_1, n_2)	(3, 20)	(5, 5)	(5, 5)	(3, 20)	(5, 5)	(5, 5)	(3, 10)	(5, 5)	(5, 5)
(k_1, k_2)	(3.5, 2.8)	(3, 3)	(3, 3)	(3.5, 3.1)	(3, 3)	(3, 3)	(3.5, 2.8)	(3, 3)	(3, 3)
ϕ	AATS	AATS	AATS	AATS	AATS	AATS	AATS	AATS	AATS
0.2	20.32	44.28	26.45	1.23	0.91	2.90	0.44	0.68	0.35
0.4	25.57	57.07	29.05	3.71	2.90	5.00	0.54	1.31	0.54
0.6	30.55	56.69	54.57	6.47	2.86	3.32	0.81	0.62	1.00
0.8	34.55	37.98	8.19	2.95	5.72	4.38	0.82	0.82	0.69
0.99	4.36	8.25	14.19	3.84	3.25	2.15	0.91	1.37	1.76
ϕ	ANOS	ANOS	ANOS	ANOS	ANOS	ANOS	ANOS	ANOS	ANOS
0.2	28.60	41.40	26.20	1.00	0.80	2.40	0.60	0.80	0.00
0.4	35.80	51.60	28.40	2.80	2.80	4.40	0.60	0.80	0.20
0.6	41.80	54.80	54.00	5.80	3.00	3.00	1.00	0.20	0.60
0.8	48.40	35.40	7.80	2.00	6.40	4.00	1.00	0.40	0.00
0.99	6.20	7.60	13.60	3.00	3.40	1.40	1.00	1.20	1.20
ϕ	ANFA	ANFA	ANFA	ANFA	ANFA	ANFA	ANFA	ANFA	ANFA
0.2	0.00	0.00	0.00	0.00	0.20	0.00	0.00	0.00	0.40
0.4	2.40	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.8	0.00	0.00	0.00	0.40	0.00	0.00	0.00	0.00	0.00
0.99	0.00	0.00	1.00	0.00	0.00	0.40	0.00	0.00	0.00

Source: The Authors

6.6. Performance comparison between the VP, VSI and \bar{x} charts

To evaluate the statistical performance of the VP chart relative to other traditionally proposed schemes in the literature, we contrast its performance in terms of the AATS, ANOS out of control, and ANFA under different levels of autocorrelation and run length (see Table 2). The results show how the increase in the autocorrelation coefficient of the process negatively affects the chart performance. However, with an elevated autocorrelation level of 0.99, there appears to be no negative impact in the performance of the chart, and in the case of detection of short runs $\delta = 0.5\sigma_x$, it appears to improve its power.

The VP chart consistently exhibits better performance compared with traditional VSI and \bar{x} schemes. The VP chart registers its best performance for large runs using $\delta = 2\sigma_x$. In general, it is possible to observe the superiority of the adaptive schemes over fixed monitoring schemes.

7. Results and discussion

The VP chart is presented in this article as an alternative to monitor processes with autocorrelated data. For that purpose, we developed an experiment based on simulation, considering different process conditions and control chart parameters in order to evaluate their performance. During the data monitoring, it was observed that the chart works better with larger runs using $\delta = 1\sigma_x$. In addition, it was demonstrated that the VP chart exhibits better performance indicators when the data present low autocorrelation (e.g., $\rho = 0.2$) and a higher proportion of the process variance due to the random mean μ_i (e.g., $\rho = 0.9$).

The VP chart designed for autocorrelated data proved to be a statistically good alternative to monitor processes showing autocorrelation. The authors suggest new studies

should be conducted that focus on the design of detection policies that enable this type of chart to detect changes in the mean more effectively.

In contrast with the VP and \bar{x} charts, the VSI chart exhibited the best performance in terms of the absolute value of AATS and ANOS in conditions with elevated correlation coefficients for observations with a larger number of simulation conditions. Similarly, the VP chart appears to be a better monitoring alternative compared with the traditional \bar{x} chart in environments with high autocorrelation levels.

8. Conclusions

This paper has shown that high levels of autocorrelation can have a significant effect on the performance of control charts. When autocorrelation is present, traditional control chart methodology should not be applied without modification. The VP and VSI Charts performed better than traditional \bar{x} charts. The VSI is among these two the recommended charts for high correlation.

For future research, we suggest the use of techniques to evaluate the individual impact of each of the factors used to determine the best combinations that help optimize statistical performance and minimize false alarms. Recent studies have showed the effectiveness of combining optimization techniques and simulated systems [24]. It is also relevant that the study of other data models is still scarce in the literature and may provide important research opportunities.

Acknowledgement

The authors are grateful for the financial support that the Colombian Research Council COLCIENCIAS granted to this work, project 121552128846 under contract 651 2011.

References

- [1] Montgomery-Douglas, C., Engineering Statistics., 2011.
- [2] Mastrangelo, C.M., Statistical Process Monitoring of Autocorrelated Data. Arizona State University, 1993.
- [3] Wieringa, J.E., Statistical process control for serially correlated data, 1999. DOI: 10.1002/cjce.5450690106
- [4] Barr, T.J., Performance of quality control procedures when monitoring correlated processes. Thesis. School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, USA, 1993.
- [5] Reynolds, M.R., Amin, R.W., Arnold, J.C. and Nachlas, J.A., Charts with variable sampling intervals, Technometrics, 30, pp. 181-192, 1988. DOI: 10.1080/0040 1706.1988.10488366
- [6] Prabhu, S., Runger, G. and Keats, J., X chart with adaptive sample sizes, The International Journal of Production Research, 31, pp. 2895-2909, 1993. DOI: 10.1080/00207549308956906
- [7] Costa, A.F., X charts with variable parameters, Journal of Quality Technology, 31, pp. 408-416, 1999.
- [8] Mahadik, S.B., X charts with variable sample size, sampling interval, and warning limits, Quality and Reliability Engineering International, 29, pp. 535-544, 2013. DOI: 10.1002/qre.1403.
- [9] Penabaena-Niebles, R., Oviedo-Trespalcacios, O., Ramirez, K. and Moron, M., Statistical-economic design for variable parameters control chart in the presence of autocorrelation, Revista Iberoamericana de Automatica e Informática Industrial, 11, pp. 247-255, 2014. DOI: 10.1016/ j.riai.2014.02.007
- [10] Peñabaena-Niebles, R., Oviedo-Trespalcacios, O., Cuentas-Hernandez, S. y García-Solano, E., Metodología para la implementación del diseño económico y/o estadístico de cartas de control X-barra con parámetros variables (VP), DYNA, 81(184), pp. 150-157, 2014.
- [11] Reynolds, M.R., Arnold, J.C. and Baik, J.W., Variable sampling interval X charts in the presence of correlation, Journal of Quality Technology, 28, pp. 12-30, 1996.
- [12] Prybutok, V.R., Clayton, H.R. and Harvey, M.M., Comparison of fixed versus variable sampling interval shewhart control charts in the presence of positively autocorrelated data, Communications in Statistics - Simulation and Computation, 26, pp. 83-106, 1997. DOI: 10.1080/036109 19708813369.
- [13] Montgomery, D.C., Introduction to statistical quality control: John Wiley & Sons, USA, 2013.
- [14] Wardell, D.G., Moskowitz, H. and Plante, R.D., Control charts in the presence of data correlation, Management Science, 38, pp. 1084-1105, 1992. DOI: 10.1287/MNSC.38.8.1084.
- [15] Zou, C., Wang, Z. and Tsung, F., Monitoring autocorrelated processes using variable sampling schemes at fixed-times, Quality and Reliability Engineering International, 24, pp. 55-69, 2008. DOI: 10.1002/qre.86.
- [16] Lin, Y.-C., The variable parameters control charts for monitoring autocorrelated processes, Communications in Statistics-Simulation and Computation, 38, pp. 729-749, 2009. DOI: 10.1080/03610910802645339.
- [17] Costa, A.F.B. and Machado, M.A.G., Variable parameter and double sampling charts in the presence of correlation: The Markov chain approach, International Journal of Production Economics, 130, pp. 224-229, 2011. DOI: 10.1016/J.IJPE.2010.12.021.
- [18] Del Castillo, E., Statistical process adjustment for quality control. Wiley, New York, USA, 2002.
- [19] Gilbert, K.C., Kirby, K. and Hild, C.R., Charting autocorrelated data: Guidelines for practitioners, Quality Engineering, 9, pp. 367-382, 1997. DOI: 10.1080/08982119708919057.
- [20] Lin, Y.-C., The variable parameters control charts for monitoring autocorrelated processes, Communications in Statistics - Simulation and Computation, 38, pp. 729-749, 2009. DOI: 10.1080/03610910802645339.
- [21] Lu, C.W. and Reynolds, M. R., EWMA control charts for monitoring the mean of autocorrelated processes, Journal of Quality Technology, 31, pp. 166-188, 1999.
- [22] Tagaras, G. and Nikolaidis, Y., Comparing the effectiveness of various Bayesian X control charts, Operations Research, 50, pp. 878-888, 2002. DOI: 10.1287/OPRE.50.5.878.361.
- [23] Harris, T.J. and Ross, W.H., Statistical process control procedures for correlated observations, The Canadian Journal of Chemical Engineering, 69, pp. 48-57, 1991. DOI: 10.1002/cjce.5450690106.
- [24] Oviedo-Trespalcacios, O. y Peñabaena, R.P., Optimización de sistemas simulados a través de técnicas de superficie de respuesta, Ingeniare. Revista Chilena de Ingeniería, 23, pp. 421-428, 2015.

O. Oviedo-Trespalcacios, received his BSc. in Industrial Engineering in 2006, and his MSc. in Industrial Engineering in 2013, both from the Universidad del Norte, Barranquilla, Colombia. From 2012 to 2014, he worked as full professor in the Industrial Engineering Department, Division of Engineering, Universidad del Norte, Barranquilla, Colombia. He is currently a Researcher Officer in human factors engineering applied to road safety at the Centre for Accident Research and Road Safety – Queensland (CARRS-Q), the Institute of Health and Biomedical Innovation (IHBI), and Queensland University of Technology (QUT), Australia. His research interests include: Traffic safety, statistical methods, human factors, cognitive engineering and industrial safety.
ORCID: 0000-0001-5916-3996

R. Peñabaena-Niebles, received her BSc. in Industrial Engineering in 1995, and her MSc. degree in Industrial Engineering in 2004 from the Universidad del Norte, Barranquilla, Colombia. She has an MSc degree in Industrial and Systems Engineering in 2013, from the University of South Florida (USF), USA and a PhD in Civil Engineering in 2015, from the Universidad de Cantabria, Santander, Spain. She is currently a full professor in the Industrial Engineering Department, Division of Engineering, Universidad del Norte, Barranquilla, Colombia. Her research interests include: Traffic engineering, industrial statistics, data mining and quality control.
ORCID: 0000-0003-4227-3798.



UNIVERSIDAD NACIONAL DE COLOMBIA

SEDE MEDELLÍN
FACULTAD DE MINAS

Área Curricular de Ingeniería Administrativa e
Ingeniería Industrial

Oferta de Posgrados

Especialización en Gestión Empresarial
Especialización en Ingeniería Financiera
Maestría en Ingeniería Administrativa
Maestría en Ingeniería Industrial
Doctorado en Ingeniería - Industria y Organizaciones

Mayor información:

E-mail: acia_med@unal.edu.co
Teléfono: (57-4) 425 52 02

Symbolic modeling of the Pareto-Optimal sets of two unity gain cells

Said Polanco-Martagón ^a, José Ruiz-Ascencio ^b & Miguel Aurelio Duarte-Villaseñor ^c

^a Departamento en Tecnologías de la Información, Universidad Politécnica de Victoria, Cd. Victoria, México, spolancom@upv.edu.mx

^b Laboratorio de Inteligencia Artificial, Centro Nacional de Investigación y Desarrollo Tecnológico (CENIDET), Cuernavaca, México, josera@cenidet.edu.mx

^c Cátedra CONACYT, Instituto Tecnológico de Tijuana, Tijuana, México, miguel.duarte@tectijuana.edu.mx

Received: May 29th, de 2015. Received in revised form: November 20th, 2015. Accepted: December 16th, 2015

Abstract

A method to data-mine results of an analog circuit sizing in order to extract knowledge that can be immediately used by designers or students of analog integrated circuit design is presented. The symbolic models, which have been generated using multi-objective genetic programming, are human-interpretable mathematical models. The procedure presented involves two steps: the generation of samples of Pareto-optimal performance sizes of two unity gain cells using the NSGA-II genetic algorithm; and the generation of models of each of the objectives sized by symbolic regression via genetic programming. The functionality of the method to describe circuits is shown in three cases: two current followers and one voltage follower.

Keywords: Symbolic Models, Symbolic Regression, Genetic Programming, White-box models, Multi-objective Optimization, Analog Circuit Sizing, Analog Circuit Design.

Modelado simbólico del conjunto Óptimo de Pareto de dos celdas de ganancia unitaria

Resumen

En este trabajo es presentado un método para extraer datos acerca de los resultados de un dimensionamiento de circuitos analógicos para extraer características que puede ser inmediatamente utilizado por diseñadores o estudiantes de diseño de circuitos integrados analógicos. Los modelos simbólicos que han sido generados utilizando Programación Genética Multi-Objetivo son modelos matemáticos humano interpretables. El procedimiento presentado abarca dos etapas: la generación de muestras del frente de Pareto de las dimensiones con desempeño óptimo de dos celdas de ganancia unitaria mediante el algoritmo genético NSGA-II; y la generación de modelos de cada uno de los objetivos del dimensionamiento utilizando regresión simbólica vía programación genética. La funcionalidad del método se muestra al describir tres casos: dos seguidores de corriente y un seguidor de voltaje.

Palabras clave: Modelado Simbólico, Regresión Simbólica, Programación Genética, Modelos transparentes, Optimización Multi-Objetivo, Dimensionamiento de circuitos analógicos, Diseño de Circuitos Analógicos.

1. Introduction

Analog integrated circuit (IC) design using Metal Oxide Semiconductor Field-Effect-Transistors (MOSFETs) imposes challenges in sizing and selecting the right circuit topology, because of the huge plethora of feasible active devices [1-4]. The sizing problem is about finding the optimal width and length (W, L) values of the MOSFETs, to guarantee optimal performance of active filters, oscillators,

sensors and so on [5-8]. The problem of choosing the right circuit topology is known as circuit synthesis [2,3,9] and it is accompanied by a sizing procedure. The selected topology must accomplish a specific function where optimal behavior is achieved through the appropriate sizing (W, L) of MOSFETs. Thus, sizing is the main problem encountered when trying to meet target requirements, such as wider bandwidth, lower power consumption and gain, all at the same time. However, some of these objectives are in conflict

How to cite: Polanco-Martagón, S., Ruiz-Ascencio, J. and Duarte-Villaseñor, M.A., Symbolic modeling of the Pareto-Optimal sets of two unity gain cells. DYNA 83(197), pp. 129-138, 2016.

with each other, thus, a multi-objective optimization is required [10]. The sizing problem for analog integrated Circuit is more complex than for digital IC due to the very large quantity of performance parameters, and the complex correlation between them.

The basic cells for designing analog ICs are known as Unity Gain Cells (UGC) [9]: Voltage Follower (VF), Current Follower (CF), Voltage Mirror (VM) and Current Mirror (CM). Among these basic cells VFs and CFs have had much interest in the recent technical literature because of their use, as an alternative to other more complex building blocks in the design of analog circuits such as: universal biquad filters, sinusoidal oscillators, impedance converters, precision rectifiers, chaotic oscillators, and so on [5-8,11].

The selection of optimal sizes of MOSFETs is a key factor in achieving higher analog IC performance. Therefore, the understanding of circuits that a designer could have is the main ingredient for improving those circuits and thus the circuit design will be more effective. However, the highly nonlinear behavior and complexity of analog IC makes it difficult to obtain insight in a method.

Data-driven modeling (DDM) can help to create models that capture the behavior of analog ICs and also to increase a designer's knowledge of analog ICs. However, the models created using some DDM techniques do not show the functional relationships between input and output parameters.

Human-interpretable mathematical models can be deployed to capture the optimal performance behavior and to find the MOSFET sizes behind it. Symbolic analysis and symbolic modeling can be used to derive human-interpretable mathematical models of analog IC behavior. Symbolic analysis of analog ICs addresses the generation of symbolic expressions for the parameters that describe the performance of linear and nonlinear circuits [11]. Thus, symbolic analysis is a systematic approach to obtaining an understanding of analog blocks in an analytic form. Symbolic analysis techniques offer a complementary way to analyze analog ICs. The results of symbolic analysis are mathematical forms that explicitly describe the functionality of the circuit as a symbolic expression of the design variables. Designers use them to gain insight into the behavior of the circuit. However, their main weaknesses are: 1) limited to linear or weak nonlinear circuits; 2) large and complex models; 3) require a great deal of experience from the analog IC design expert [11].

Symbolic modeling has similar goals to symbolic analysis, but a different core approach to solving the problem. Symbolic modeling or symbolic regression is defined as the use of simulation data to generate interpretable mathematical expressions for circuit applications [12]. It extracts mathematical expressions from SPICE simulation data by means of genetic programming (GP). Since SPICE simulations readily handle nonlinear circuits [13], its simulation results may be leveraged for design purposes by means of symbolic modeling.

Other methods of DDM use a data set obtained from a statistical sampling of the global behavior of the analog IC [12,13]. These models are used instead of the circuit simulator in different stages of analog design. However, due to their global nature they do not answer questions such as:

What are the relevant factors in the optimal gain? What are the relevant factors in the optimal bandwidth? How do the trade-offs affect these factors? Answering these kinds of questions is of utmost importance so as to enrich the experience of an analog circuit designer.

A promising approach to explore these trade-offs is the use of the Pareto front of optimally sized solutions of analog IC [14]. The Pareto optimal greatly reduces the design space, keeping the best configurations of the circuit and the best trade-offs possible among the multiple objectives employed. Thus, the Pareto front provides answers to the questions posed above.

The objective of this paper is to show a two-step methodology for the extraction of human-interpretable mathematical models of the feasible solutions of two Unity Gain Cells: a Voltage Follower and a Current Follower. The first step (section II), the optimal selection of design variables to maximize design objectives is formulated as a multi-objective optimization problem. The second step, Genetic Programming (GP) is used to create human-interpretable models starting from the feasible solution sets obtained from the first step. In Section III a brief introduction to the Symbolic Regression by Genetic Programming theory, and the general scheme of the MultiGene Genetic Programming (MGGP) method used for this work is given. In Section IV the details of the methodology proposed are presented. The experimental results of the methodology proposed to obtain symbolic models of two UGCs are presented in Section V. A brief discussion of the results is provided in Section VI. Finally, in Section VII the conclusions are presented.

2. Background

Data-driven models are built using observed and captured properties of a system, exhibited under different conditions or over time, and expressed in numerical form. The task of empirical modeling, or data modeling, lies in using a limited number of observations of system variables (data) to infer relationships among these variables.

The task of modeling is to identify input variables driving the response variables and formulate the exact relationship in the form of an accurate model. Fig. 1 depicts a schematic view of the general process of data-driven modeling. A system model is the system under study; it can be a real or simulated process. In Fig. 1 (x_1, \dots, x_d) is a set of input variables, (y_1, \dots, y_r) is a set of response variables, and DDM is a collection of relationships that model the observed responses: $(y_1, \dots, y_r): \{\hat{y}_j = \hat{f}(x_1, \dots, x_d), j = 1:r\}$.

The models obtained through the use of data-driven techniques can be classified according to their transparency; i.e., their ease in transmitting relationships between input and output variables. The first category, "black box" models, corresponds to those techniques from which information about relationships between inputs and outputs cannot easily be obtained. Some of the drawbacks of black-box models are: 1) lack of significant variables determination, 2) difficult to invert, 3) fundamental structure difficult to interpret, 4) lack of adaptability to new changes and 5) difficulty of

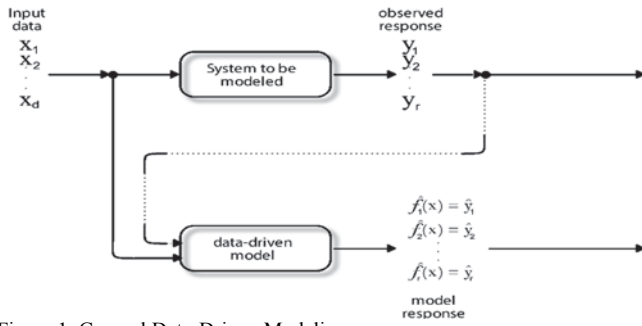


Figure 1. General Data-Driven Modeling process.
Source: Authors' own.

incorporating expert knowledge. This makes it impossible to answer simple questions about the underlying system, e.g.: What are the ranges of input variables that cause a response to take certain values, and not necessarily the optimal values? Some black-box model techniques include Neural Networks [15-18], support vector machines [19], and kriging [20]. To gain insight into an implicit black-box model, and facilitate its interpretation and manipulation, it is possible to create explicit models from the black-box inputs and outputs parameters. This process is called meta-modeling.

The second category is "gray-box" models; these correspond to those techniques where some insight into the relations between input and output variables can be extracted from the model obtained. Gray box modeling techniques include latent variable regression [21] and nonlinear sensitivity analysis [22]. The most important criteria for these modeling techniques are prediction ability, model building time, scalability, and model simulation time.

However, in contrast to the prediction error minimization method, the problem of knowledge extraction aims to obtain insight from a data set regardless of the complexity of the model obtained. Ideally, a functional relationship of some variables to other variables is sought. The third category is "white-box" model or "glass-box model", this technique can provide insight into relationships between input-output variables. This category includes Symbolic Regression via Genetic Programming (SRGP) which is a technique to automatically generate symbolic models that provide functional relations from given data [23]. As a consequence, SRGP has been successfully used for extracting white box models from samples of the full design space of an analog IC [13,24,25].

3. Symbolic regression by genetic programming

Genetic Programming is a bio-inspired method used in the field of machine learning [26]. An introduction and review of the GP literature is provided by [27]. The main difference between Genetic Algorithms (GA) and GP is the representation of individuals. In the GA individuals are expressed as linear strings of fixed length (chromosome) through all evolution steps; whereas in GP, individuals are expressed as nonlinear objects with different sizes and shapes, normally these nonlinear objects are expression trees. Thus, motivated by the evolutionary process observed in nature, computer programs are evolved to solve a given task. This process is done by randomly generating an initial

population of trees and then, by using operations such as crossover and mutation, a new population of offspring is created [26]. A new parent population is selected from the initial and offspring populations, based on how good these trees perform for the given task. The best offspring resulting from this process is the solution of the problem. GP is implemented as per the following five steps:

- 1) Generate an initial population of individuals formed as expression trees, consisting of the functions (arithmetic, basic mathematical and Boolean functions) and terminals (constants and inputs variables) of the problem.
- 2) The fitness of each individual in a population is measured through a criterion such as the mean squared error.
- 3) Likewise, the complexity of the individual is measured through a criterion such as the sum of all nodes in the tree.
- 4) Better parents are usually selected through a tournament involving comparison of two parents at a time and thereafter, short listing the winner for further competition with the hope that it has a better chance of producing better offspring. The concept of what is "best" may be multi-objective, this means that there are two or more criteria used for selecting "good individuals" for further propagation. Often these criteria are prediction error and model complexity.
- 5) New offspring are generated by changing the parents through three genetic operations: reproduction, crossover and mutation. After finishing the first cycle of the five steps listed, steps 2 to 5 are iterated. After some condition is reached, the evolutionary process is stopped. This gives rise to a set of solutions ranging from less complex individuals but with high error, to those highly complex individuals that have minimum error.

When GP is used to create empirical mathematical models of data acquired from a process or system, it is commonly referred to as symbolic regression [28]. SRGP is an iterative search technique, that looks for appropriate expressions of the response variable in a space of all valid formulae containing some of the given input variables and some predefined functional operators, like summation, multiplication, division, exponentiation, sine, and so on.

The task of symbolic regression can be stated as follows. Given a set of $\{x(t), y(t)\}, t = 1, \dots, N$; where $x(t)$ is a d -dimensional input t , $y(t)$ is a corresponding output value, P is a set of user-specified functions, F is the universe of all mathematical expressions formed from P , the task of symbolic regression is to determine a function $f \in F$, which provides a small prediction error. The resulting function f is later used to predict the output for a new given input. The quality of f is determined by how well it maps the given inputs to their corresponding outputs. An advantage of symbolic regression in contrast to classical regression analysis, is that GP automatically evolves the structure and the parameters of the mathematical model.

In contrast to the classical GP, in MultiGene Genetic Programming (MGGP) each symbolic model (expression tree) in the GP population is made from several trees instead of one [28,29]. All of the genes have specific optimal weights

and the summation of the weighted genes plus a bias term form the final formula. The mathematical form of the multigene representation is shown in (1) [28].

$$y = \sum_{i=1}^n d_i G_i + d_0 \quad (1)$$

Where y is the predicted output, G_i is the value of the i -th gene and is a function of one or more of the input variables, d_j is the i -th weighting coefficient, n is the number of genes and d_0 is a set term. For example, a multigene model that predicts an output variable (y) using input variables x_1 and x_2 is shown in Fig. 2.

The software GPTIPS [29] is used in the current study to perform MGGP, which produces the symbolic modeling of the optimal sizes W and L of a VF and a CF whose data sets are collected in Section 5. GPTIPS is a SRGP code which was written based on Multigene Symbolic Regression to be used with MATLAB®. This software has the ability of avoiding most bloat problems by imposing restrictions on the maximum value of parameters such as the maximum number of genes, the maximum depth of trees and genes, the maximum number of nodes per tree, and so on.

4. Methodology

The methodology used is comprised of two steps: Pareto Front Generation and obtaining the white box model by using MGGP.

4.1. Pareto front generation

The basic cells for designing analog integrated circuits are several, such as OPAMP, OTA, CC, and so on; but [9] highlights the UGCs: VF, CF, VM and CM. The interconnection of UGCs generates more complex ICs, such as the Current Feedback Operational Amplifier (CFOA). These kinds of ICs have been sized with different optimization methodologies such as the Evolutionary Algorithms [9,10,30,31], with the aim of generating optimal behavior through the evaluation of multiple objectives [32-34] (performance parameters). Evolutionary Algorithms generate populations of feasible solutions, which approximate the Optimal Pareto set [35]. However, in the optimization of analog ICs, much of the time the best solutions include extreme performance requirements such as ultra-low power, low voltage and high frequency. Therefore,

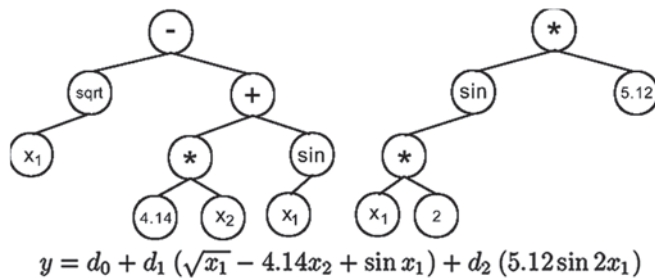


Figure 2. Example of a MultiGene Symbolic Model.
Source: Authors' own.

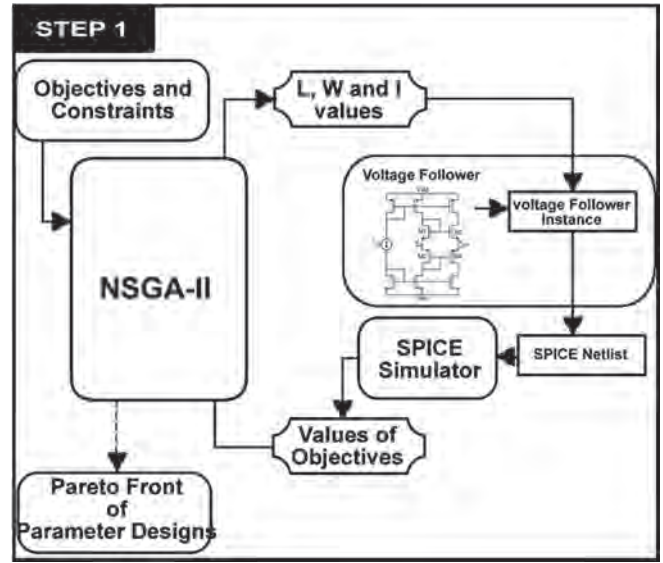


Figure 3. Step 1: UGC sizing optimization for Pareto Front Generation.
Source: Authors' own.

the set of optimal solutions is located in some region of the periphery of the feasible solutions.

Multi-objective optimization is a convenient tool to exploit the feasible and optimal solutions in situations where two or more objectives are in conflict and have to be maximized or minimized at the same time. Thus, the first step of the methodology is to identify the optimal I , W and L sizes of two UGCs, a VF and a CF, maximizing the unity gain and bandwidth of such circuits. This is carried out using the process depicted in Fig. 3 [14].

The optimization stage is performed by applying the NSGA-II algorithm [14, 32, 36, 37], to minimize a problem of the form:

$$\begin{aligned} \text{minimize } f(x) &= [f_1(x), f_2(x), \dots, f_m(x)]^T \\ \text{subject to } h_k(x) &\geq 0, k = 1 \dots p, \end{aligned} \quad (2)$$

where function $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$, $X = [x_1, \dots, x_n]^T$ is the decision vector and n is the number of variables; $x \in X$, where $X \subset \mathbb{R}^n$ is the decision space for the variables.

All objective functions $f_j(x): \mathbb{R}^n \rightarrow \mathbb{R}$, $j = 1 \dots m$ ($m \geq 2$) and $h_k(x)$, $k = 1 \dots p$ are performance constraints. Very often, since the objectives in (2) contradict each other, no point $x \in X$ minimizes the objectives simultaneously. The best trade-offs among the objectives can be defined in terms of the Pareto optimal [36,37].

The NSGA-II Algorithm is based on Pareto ranking. Firstly, two populations P_0 and Q_0 are generated, each one of size N . The NSGA-II procedure in each generation consists of rebuilding the current (t) population (R_t) from the two original population fronts [36]. In the next step, new offspring (P_{t+1}) are created from the current population R_t (previously ranked and ordered by sub-front number), in order to choose from a population of size $2N$, N solutions belonging to the first sub-front. In addition, the last sub-front could be greater than necessary, and the measure ($i_{distance}$) is used to preserve diversity by the selection of the solutions that are far from the rest [38]. To build

new generations, we use multi-point crossover and single point mutation as genetic operators. The optimization loop is iteratively applied until some stopping criterion is met, e.g., a certain number of iterations are executed or a sufficient convergence to the Pareto front with appropriate diversity is achieved [37,39].

In this case, each variable x represents the Width (W) or Length (L) of the MOSFETs, or the bias current (I). Those values are integer-multiples of a minimum value allowed by the fabrication process. The objectives (f_j) are the circuit performances which are unity gain and high bandwidth, that are extracted from the output of the SPICE simulation. The output of the first step is a set of points of the Pareto front, as is shown in Fig. 3.

4.2. White box model generation

A set of samples of the Pareto front can provide information to answer questions on feasibility and performance trends. The proposed approach includes a second step: white box model generation. For this paper the models are of gain and bandwidth performances of the sizes W and L of the UGCs. These are obtained by using symbolic modeling through MGGP. Fig. 4 depicts the process of construction of the white box models of feasible performance solutions of the UGCs.

The aim of symbolic modeling is to use simulation data to generate interpretable mathematical expressions that relate the circuit performances to the design variables [13]. Thus, those mathematical expressions replace complex systems whose performance evaluation is computationally expensive. Moreover, the symbolic models obtained can provide insight into the underlying system, unlike other modeling methods like kriging, neural networks or support vector machines.

Generating a white box model by symbolic regression usually involves three steps:

- 1) Design of experiments. It encompasses the selection of sample vectors to build the model. A wide variety of techniques are available, ranging from the classical Monte Carlo sampling, to latin hypercube sampling [40]. Basically, it tries to achieve a more uniform coverage of the search space with the minimum number of samples.

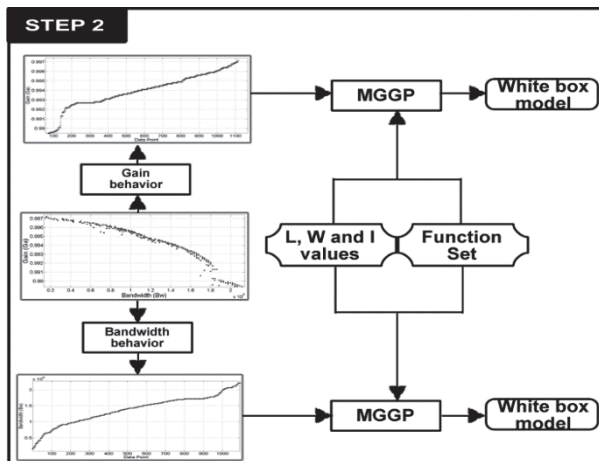


Figure 4. Step 2: White box model generation.

Source: Authors' own.

- 2) Accurate evaluation. SPICE simulations are used to evaluate the performances for the sample circuits selected in step 1. This is the most computationally expensive step.
- 3) Model construction. This part concerns building the symbolic model using the data sets obtained in step 2. For the model to be easily understood, low complexity is essential.

Experimental design is not used in step 2. The sample vectors to build the symbolic model correspond to the samples of the Pareto front obtained in step 1 (Fig 3). In this case, the Pareto optimal front has relatively few samples representing the best configurations in the design space. Multi-objective optimization algorithms use different strategies to ensure a good convergence to the ideal Pareto front and a good diversity of solutions within that front; e.g., ranking of solutions and maximization of the crowding distance of non-dominated solutions are used in NSGA-II [36]. Therefore, Multi-objective algorithms play the role of experimental design, i.e. the NSGA-II algorithm gets the sampling set to perform the white box modeling instead of using statistical sampling.

5. Experiments and Results

The optimization system has been programmed in standard C, and evolves the Widths (W), Lengths (L) and current bias (I) for each transistor of the UGC. Results are provided by TopSPICE as a listing of output data and are used for evaluating the individuals. The UGCs are coded using eight design variables for the VF and two or three for the CF, as can be seen in Table 1 and Table 2 respectively. For illustrative purposes, two conflicting objectives are shown in Fig. 4; but any number of objectives may be used. The UGCs used are shown in Fig. 5.

To perform symbolic modeling using GPTIPS, the parameters shown in Table 3 are used. Parameter settings play a major role in implementing the MGGP efficiently. The majority of parameters were selected based on the most frequently cited values in the literature [2,3,9,29,31]. The function set may include a broad selection of primitives so as

Table 1.
Coding of the VF of Figure 5(b).

Gene	Variable	Transistor
x_1	I	M_1, \dots, M_{10}
x_2	L	M_1, \dots, M_{10}
x_3	W_1	M_1, M_2
x_4	W_2	M_3, M_4
x_5	W_3	M_5, M_9
x_6	W_4	M_6, M_{10}
x_7	W_5	M_7
x_8	W_6	M_8

Source: Authors' own.

Table 2.
Coding of the CF of Figure 5(a)

Gene	Variable	CF 2var	CF 3Var
x_1	W_1	M_1, \dots, M_8	M_1, \dots, M_8
x_2	W_2	M_{b1}, \dots, M_{b8}	M_{b1}, \dots, M_{b8}
x_3	L	—	$M_1, \dots, M_8, M_{b1}, \dots, M_{b8}$

Source: Authors' own.

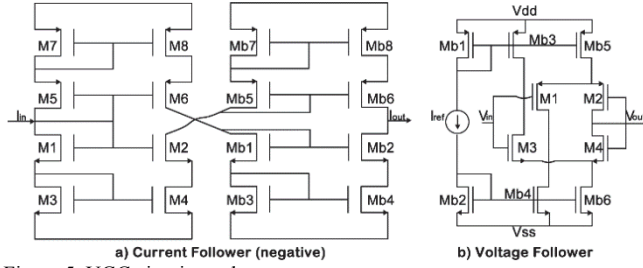


Figure 5. UGC circuits under test.
Source: Authors' own.

Table 3.
Parameter settings for GPTips.

Parameters	Values Assigned
Runs	10
Population size	500
Number of Generations	500
Tournament size	4
Max depth	10
Max Genes	8
Function set (P)	$+, -, *, t^2, t, \exp(t), \log t$
Terminal set (T)	$+, -, *, t^2, t, \exp(t), \log t, \tanh t, \sinh t, \cosh t$ $x_1, x_2, [-10, 10]$ $x_1, x_2, x_3, [-10, 10]$ $x_1, x_2, x_3, x_4, x_5, x_6, [-10, 10]$
Crossover probability	0.6, 0.75, 0.85
Mutation probability	0.01, 0.05, 0.1
Direct reproduction	0.05

Source: Authors' own.

to evolve a variety of nonlinear models. Of the three functional sets in Table 3, it chooses the smallest. Although the larger sets give slightly more accurate models, their complexity makes them too obscure for practical use. The values of population size and number of generations depend on the complexity of the data. In addition, the number of genes and the maximum depth of a gene influences the size and the number of models to be searched for in the global space.

The evaluation criteria for symbolic models are model precision and model complexity. Model precision is the adjustment for training and test data, using the Root Mean Square Error (RMSE) as an adjustment measure, as in (3), where G_i is the value of the i th data sample predicted by the symbolic model, A_i is the actual value of the i th data sample and N is the number of training samples. Model complexity was evaluated as the sum of all nodes in the genes. Using this method we selected symbolic models that were sufficiently accurate and highly interpretable.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (G_i - A_i)^2}{N}} \quad (3)$$

5.1. Current Follower

The first example of using this methodology is the modeling of the Pareto Optimal Front (POF) of a CF shown in Fig. 5(a). It has L fixed at $0.7 \mu m$, so there are only two variables to model, W_1 and W_2 ; MOSFET N and P , respectively. These feasible solutions are depicted in Fig 6.

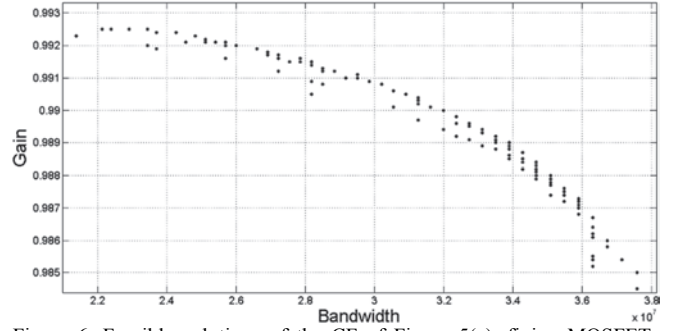


Figure 6. Feasible solutions of the CF of Figure 5(a), fixing MOSFETs Length.
Source: Authors' own.

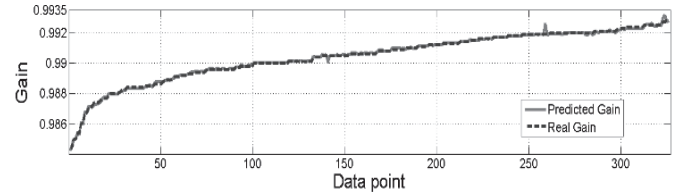


Figure 7. Predicted gain and Real gain for the CF of Figure 5(a), fixing the MOSFETs' Length.
Source: Authors' own.

$$\begin{aligned} Ga = & -(2.664 \cdot 10^{-7}) x_1^3 x_2 \\ & - (1.438 \cdot 10^{-6}) x_1^2 x_2 + 0.0001055 x_1^2 \\ & + (5.074 \cdot 10^{-5}) x_1 x_2 - 0.0021 x_1 \\ & - (4.685 \cdot 10^{-6}) x_2^2 + (7.389 \cdot 10^{-5}) x_2 \\ & + 0.9928 \end{aligned} \quad (4)$$

Once the feasible solutions are obtained, each performance parameter is modeled separately using the MGGP method, as described in Section IV. The symbolic model obtained for the unitary gain of the CF of Fig. 5(a) is shown in (4), where $x_1 = W_1$ and $x_2 = W_2$. Its approximation behavior is depicted in Fig. 7 where the predicted gain is the value obtained from the symbolic model (4) using the test data; its RMSE is $8.6274e^{-05}$.

The symbolic model obtained for the bandwidth of the feasible solutions for the CF (Fig. 5(a)) is shown in (5).

Figure 8 shows the results of the symbolic model and the SPICE simulation using test data.

$$\begin{aligned} Bw = & 3240.0 x_1^5 - 95333.0 x_1^4 + (1.061 \cdot 10^6) x_1^3 \\ & + (-5.176 x_2^3 - 697.6 x_2 \\ & - 5.774 \cdot 10^6) x_1^2 \\ & + ((1.63 \cdot 10^5) x_2 \\ & - 2.588 x_2^3 (x_2 - 13.12) \\ & - 348.8 x_2^2 + 1.558 \cdot 10^7) x_1 \\ & + (23277.0 x_2^2 \\ & - (1.494 \cdot 10^6) x_2 + 2.228 \\ & \cdot 10^7) \end{aligned} \quad (5)$$

The second example is the modeling of the POF of the same CF, but in this case, one variable was added to the optimization process. The modeling variables are $x_1 = L$, $x_2 = W_1$ and $x_3 = W_2$. L can take values of $0.4 \mu m$, $0.7 \mu m$, $1 \mu m$, $1.05 \mu m$

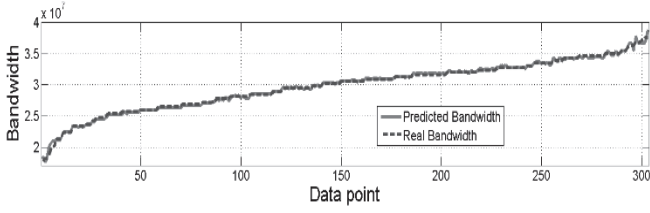


Figure 8. Predicted BW and Real BW for the CF of Figure 5(a), fixing the MOSFETs' Length.
Source: Authors' own.

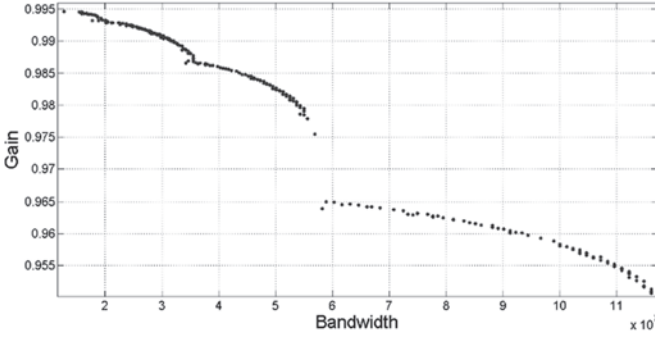


Figure 9. Pareto front of the CF of Figure 5(a).
Source: Authors' own.

and $1.4\mu m$. The feasible solutions are depicted in Fig. 9. The feasible solutions from Fig. 6 are part of this set.

Following the steps in Section IV, feasible design parameter solutions for the CF (Fig. 5(a)) are first obtained, and then each performance parameter of the optimization process is modeled with GPTIPS. These produce symbolic models of the gain and CF bandwidth. The symbolic model obtained for the current follower bandwidth is expression (6). The approximation behavior is depicted in Fig. 10; the predicted gain is the value obtained from the symbolic model (6) and the actual gain is the value obtained from the test data (SPICE). The symbolic model obtained for the bandwidth of the feasible solutions for the CF is shown in (7). Fig. 11 shows the behavior of the obtained symbolic model (7) versus the real data obtained using the test data (SPICE).

$$\begin{aligned} Ga = & -0.004957x_1^5 + 0.1235x_1^3 \\ & + (-0.001047x_2^2 - 0.3462)x_1^2 \\ & + (0.001047x_2^2 + 0.005884x_2 \\ & - 0.001047x_3 + 0.3535)x_1 \\ & + (0.00132x_3 - 0.00736x_2 \\ & + 0.866) \end{aligned} \quad (6)$$

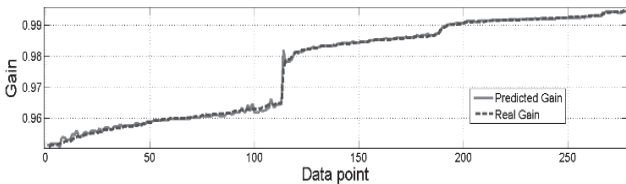


Figure 10. Predicted Gain and Real Gain for the CF of Figure 5(a).
Source: Authors' own.

$$\begin{aligned} Bw = & (4.6 \cdot 10^7)x_1 + (4.325 \cdot 10^7)x_2 \\ & - (8.329 \cdot 10^6)x_3 \\ & + (2.606 \cdot 10^5)x_1x_2 \\ & - (2.606 \cdot 10^5)x_2(x_2 - 1.828x_1^2) \\ & - (1.007 \cdot 10^7)x_1(x_1^3 + 2x_1 \\ & - x_3) - (4.45 \cdot 10^7)x_1^3 \\ & + (2.606 \cdot 10^5)x_1(6.68x_3 \\ & - 30.39) \\ & - (2.606 \cdot 10^5)x_1^2x_2, (x_1 - 9.1) \\ & + (2.797 \cdot 10^6)x_1(x_1 \\ & - 1.835)(16.63x_2 - 33.26x_1 \\ & + 134.4) \\ & + 352.6(x_1 \\ & - (x_1 - 9.1)(x_1 + x_2 \\ & - 7.265))(x_2^2 - 9.286x_1x_2 \\ & + 2x_1)(16.22x_1 + 3.963x_2 \\ & - 60.39)(x_1 + x_1x_2 + x_1^2x_3 \\ & - 9.1) + 2.672 \cdot 10^8 \end{aligned} \quad (7)$$

5.2. Voltage Follower

Finally, the proposed methodology is applied for modeling the Pareto optimal front of a VF, shown in Fig. 5(b). This VF has eight variables to encode, as seen in Table 1. The Pareto optimal front to be modeled is depicted in Fig. 12. The symbolic models obtained for gain and bandwidth are presented in (8) and (9).

The behavior of the Gain symbolic model (8) and the SPICE simulation result (Actual Gain) are shown in Fig. 13. It can be observed that MGGP is capable of approximating the behavior of actual gain solutions; its RSME is $4.915e^{-4}$. The symbolic model obtained by using the MGGP method is shown in (8). As can be seen, the model is highly interpretable given that it is formed only by additions and subtractions of polynomials. Likewise, the BW symbolic model obtained is shown in (9). Fig. 14 shows the behavior

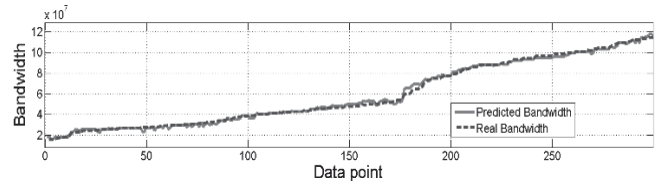


Figure 11. Predicted BW and Real (actual) BW for the CF of Figure 5(a).
Source: Authors' own.

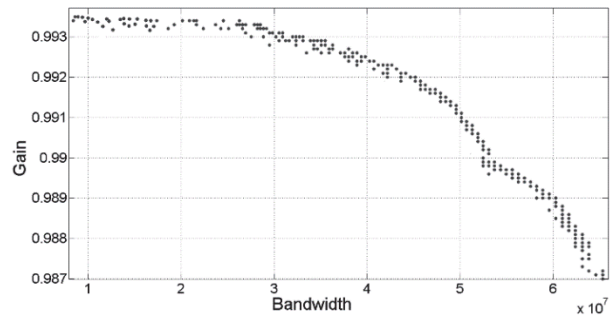


Figure 12. Pareto front of the VF of Figure 5(b).
Source: Authors' own.

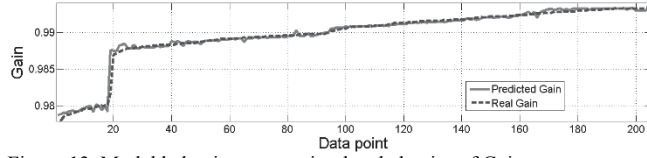


Figure 13. Model behavior versus simulate behavior of Gain.
Source: Authors' own.

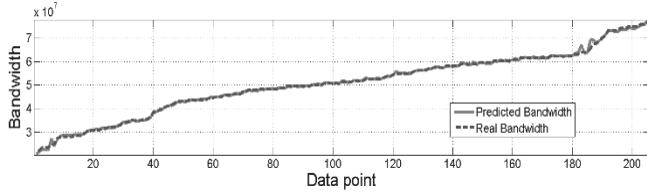


Figure 14. Model behavior versus Simulate behavior of BW.
Source: Authors' own.

of the symbolic model for the bandwidth and the behavior of the actual (SPICE simulated) bandwidth.

$$\begin{aligned}
 Ga = & (1.487 \cdot 10^{-5})x_2^{13} + (1.351 \cdot 10^{-8})x_2^4 \\
 & - (1.351 \cdot 10^{-8})x_2^2x_3 \\
 & - 0.03551x_2^2 + 0.08916x_2 \\
 & - (1.351 \cdot 10^{-8})x_3^2 \\
 & + (1.474 \cdot 10^{-5})x_3 \\
 & - (4.053 \cdot 10^{-8})x_1 \\
 & + (2.702 \cdot 10^{-8})x_4 + 0.9329
 \end{aligned} \quad (8)$$

$$\begin{aligned}
 Bw = & (3.44 \cdot 10^5)x_1 - (6.336 \cdot 10^7)x_2 \\
 & - 68155.0x_3 - 32300.0x_4 \\
 & - 21533.0x_5 - 10777.0x_6 \\
 & + (1.571 \cdot 10^7)x_2^2 \\
 & + 10.68x_2^4x_3 (-1x_1x_2^2 - x_1 \\
 & + x_3 + x_4 + x_5) + 8.878 \cdot 10^7
 \end{aligned} \quad (9)$$

6. Discussion of Results

The symbolic regression process provides an abundant population of different types of symbolic models with different values in RMSE and complexity. Some having a smaller error, have complexities exceeding the hundreds of nodes, making them uninterpretable. In addition to the above, enlarging the functional set drastically increases the complexity of the symbolic models obtained. For that reason, the functional set is limited to only three functions: addition, subtraction and multiplication. In Section 5 symbolic models and their behavior against SPICE simulated data are shown. As mentioned above, symbolic models that are highly interpretable by human beings and can be of practical use are selected.

Given the quantity of individuals that form the population of symbolic models and the complexity and adjustment of each one, this work focused on those models that dominate others for the selection of symbolic models. In other words, individuals in the knee of the Pareto Front of models are selected; this guarantees that selected models are highly interpretable and have a good adjustment.

As is well known, the Pareto-domination based selection

of the NSGA-II algorithm succeeds at driving the whole population towards the Pareto optimal set or Pareto Front. However, it has no direct control over the evolution of each individual in the population and thus it has no good mechanism to control the distribution over the PF. That is why repetition is recommended to ensure the location of the real Pareto front. Moreover, the NSGA-II groups individuals from different regions of the design space in the Pareto front. The interpolation of antecedents of two neighbors of the Pareto front may not necessarily be valid.

7. Conclusions

This paper has presented a procedure to data-mine the results of a circuit sizing to extract domain knowledge that can be immediately used by IC experts. Thus, the knowledge can be distributed to designers or students, without the need for more synthesis, since the symbolic models, which have been generated using MGPP, are human-interpretable mathematical models.

The procedure presented involves two steps: the generation of samples of the Pareto-optimal performance sizes of a UGC using the MOGA and the generation of white box models of each of the objectives of this Pareto Front.

The symbolic models are formed by additions and subtractions of polynomials; thus, their form is highly interpretable. Nevertheless, these are very hard to invert, so it is necessary to go back to the antecedent simulations.

For values between two or more simulation points, interpolation does not always make sense. This is because the antecedent points come from distant regions in the design space. It is important to highlight that in this work, less complex symbolic models were selected despite having a less precise adjustment. Thus, selecting complex models that are obscure as a neural network is avoided.

References

- [1] Rutenbar, R.A., Gielen, G. and Antao, B., Computer-aided design of analog integrated circuits and systems, Piscataway, NJ, USA: Wiley-IEEE Press, 2002. ISBN: 978-0-471-22782-3
- [2] Razavi, B., Design of analog CMOS integrated circuits: Tata McGraw-Hill Education, 2002. ISBN-10: 0072380322
- [3] Tlelo-Cuautle, E., Integrated circuits for analog signal processing. Tlelo-Cuautle E., ed., Springer New York, USA, 2013. ISBN: 9781461413820
- [4] Sánchez-López, C., Martínez-Romero, E. and Tlelo-Cuautle, E., Symbolic analysis of OTRAs-based circuits. Journal of applied research and technology [online], 9(1), pp. 69-80, 2011. [date of reference Dec. 11th, 2015]. Available at: <http://www.revistas.unam.mx/index.php/jart/article/view/25278/23750>
- [5] Khateb, F. and Biolk, D., Bulk-driven current differencing transconductance amplifier, Circuits, Systems, and Signal Processing, 30(5), pp.1071-1089, 2011. DOI: 10.1007/s00034-010-9254-9
- [6] Campos-Cantón, I., Desarrollo de celdas lógicas por medio del espacio de estados en un sistema bidimensional, Revista Mexicana de Física, [online], 57(2), pp. 106-109, 2011.[date of reference Dec. 11th, 2015] Available at: <http://revistas.unam.mx/index.php/rmf/article/view/25225/23707>
- [7] Swamy, M., Mutators, generalized impedance converters and inverters, and their realization using generalized current conveyors. Circuits, Systems, and Signal Processing, 30(1), pp.209-232, 2011. DOI: 10.1007/s00034-010-9208-2
- [8] Pathak, J., Singh, A.K. and Senani, R., Systematic realisation of quadrature oscillators using current differencing buffered amplifiers.

- IET Circuits, Devices & Systems, 5(3), pp. 203-211, 2011. DOI: 10.1049/iet-cds.2010.0227
- [9] Duarte-Villaseñor, M.A., Tlelo-Cuautle, E. and de la Fraga L.G., Binary genetic encoding for the synthesis of mixed-mode circuit topologies. *Circuits, Systems, and Signal Processing*, 31(3), pp. 849-863, 2012. DOI: 10.1007/s00034-011-9353-2
- [10] Coello-Coello C.A., Van Veldhuizen, D.A. and Lamont, G.B., *Evolutionary algorithms for solving multi-objective problems*, Second ed, Springer US, 2002. ISBN: 978-0-387-33254-3
- [11] Sheldon, X. and Tlelo-Cuautle, E., Recent development in symbolic analysis: An overview. In: Fakhfakh, M., Tlelo-Cuautle, E. and Fernández, F.V., eds. *Design of analog circuits through symbolic analysis*: Bentham Science Publishers, 2012, pp. 3-18. ISBN: 9781608050956
- [12] McConaghy, T. and Gielen, G.G.E., Globally reliable variation-aware sizing of analog integrated circuits via response surfaces and structural homotopy. *Computer-Aided Design of Integrated Circuits and Systems*, IEEE Transactions on, 28(11), pp.1627-1640, 2009. DOI: 10.1109/TCAD.2009.2030351
- [13] McConaghy, T. and Gielen, G.G., Nonlinear template-free symbolic performance modeling for design and process variation analysis of analog circuits. In: Fakhfakh, M., Tlelo-Cuautle, E. and Fernández, F.V., eds. *Design of analog circuits through symbolic analysis*, Bentham Science Publishers, 2012, pp. 3-18. ISBN: 9781608050956
- [14] Polanco-Martagón, S., Reyes-Salgado, G., De la Fraga, L.G., Tlelo-Cuautle, E., Guerra-Gómez, I., Flores-Becerra, G., et al., Optimal sizing of analog integrated circuits by applying genetic algorithms. In: Ramirez-Muñoz, A. and Garza-Rodríguez, I., eds. *Handbook of Genetic Algorithms: New research*. Mathematics Research Developments, Nova Science Publishers; 2012. ISBN: 978-1-62081-184-9
- [15] McConaghy, T. and Gielen, G., Analysis of simulation-driven numerical performance modeling techniques for application to analog circuit optimization, in *Circuits and Systems, 2005 ISCAS 2005 IEEE International Symposium*; pp. 1298-1301, 2005. DOI: 10.1109/ISCAS.2005.1464833
- [16] Wolfe, G. and Vemuri, R., Extraction and use of neural network models in automated synthesis of operational amplifiers. *Computer-Aided Design of Integrated Circuits and Systems*, IEEE Transactions on, 22(2), pp.198-212, 2003. DOI: 10.1109/TCAD.2002.806600
- [17] Sarmiento, A.T. and Soto, O.C., New product forecasting demand by using neural networks and similar product analysis. *DYNA*. 81(186), pp. 311-317, 2014. DOI: 10.15446/dyna.v81n186.45223
- [18] Lizarazo-Marriaga, J.M. and Gómez-Cortés, J.G., Desarrollo de un modelo de redes neuronales artificiales para predecir la resistencia a la compresión y la resistividad eléctrica del concreto, *Ingeniería e Investigación*, [online]. 27(1), pp. 11-18. 2007. [date of reference Dec. 11th, 2015]. Available at: <http://revistas.unal.edu.co/index.php/ingevin/article/view/14771>
- [19] Boolchandani, D., Ahmed, A. and Sahula, V., Efficient kernel functions for support vector machine regression model for analog circuits' performance evaluation. *Analog Integrated Circuits and Signal Processing*, 66(1), pp. 117-128, 2011. DOI: 10.1007/s10470-010-9476-6
- [20] Kotti, M., González-Echevarría, R., Fernández, F.V., Roca, E., Sieiro, J., Castro-López, R., et al., Generation of surrogate models of Pareto-optimal performance trade-offs of planar inductors. *Analog Integrated Circuits and Signal Processing*, 78(1), pp. 87-97, 2014. DOI: 10.1007/s10470-013-0230-8
- [21] Li, X. and Cao, Y., Projection-based piecewise-linear response surface modeling for strongly nonlinear VLSI performance variations. *Quality Electronic Design*, in 9th International Symposium; pp. 108-113, 2008. DOI: 10.1109/ISQED.2008.4479708
- [22] McConaghy, T., Palmers, P., Gielen, G. and Steyaert, M., Automated extraction of expert knowledge in analog topology selection and sizing, in *Computer-Aided Design, 2008 ICCAD 2008 IEEE/ACM International Conference*, pp. 392-395, 2008. DOI: 10.1109/ICCAD.2008.4681603
- [23] Vladislavleva, E.J., Smits, G.F. and den Hertog, D., Order of nonlinearity as a complexity measure for models generated by symbolic regression via Pareto genetic programming. *Evolutionary Computation*, IEEE Transactions on, 13(2), pp. 333-349, 2009. DOI: 10.1109/TEVC.2008.926486
- [24] Aggarwal, V. and O'Reilly, U-M., Simulation-based reusable posynomial models for MOS transistor parameters, in *Conference on Design, Automation and Test in Europe*, San Jose, CA, USA: EDA Consortium, 2007, pp. 69-74. 2007. DOI: 10.1109/DATE.2007.364569
- [25] McConaghy, T. and Gielen, G.G.E., Template-Free symbolic performance modeling of analog circuits via Canonical-Form functions and genetic programming. *Computer-Aided Design of Integrated Circuits and Systems*, IEEE Transactions on, 28(8), pp. 1162-1175, 2009. DOI: 10.1109/TCAD.2009.2021034
- [26] Koza, J.R., *Genetic programming: On the programming of computers by means of natural selection*. Cambridge: MIT press; 1992. ISBN: 9780262111706
- [27] Poli, R., Langdon, W.B., McPhee, N.F. and Koza, J.R., A field guide to genetic programming [online] Published via <http://lulu.com>; 2008.[consulted, December 10th, 2015]. Available at <http://www.gp-field-guide.org.uk>.
- [28] Hii, C., Searson, D. and Willis, M., Evolving toxicity models using multigene symbolic regression and multiple objectives. *International Journal of Machine Learning and Computing*, 1(1), pp. 30-35, 2011.
- [29] Searson, D.P., Leahy, D.E. and Willis, M.J., GPTIPS: An open source genetic programming toolbox for multigene symbolic regression, in *International Multiconference of Engineers and Computer Scientists*, pp. 77-80, 2010.
- [30] Sallem, A., Benhala, B., Kotti, M., Fakhfakh, M., Ahaitouf, A. and Loulou, M., Application of swarm intelligence techniques to the design of analog circuits: Evaluation and comparison. *Analog Integrated Circuits and Signal Processing*, 75(3), pp. 499-516, 2013. DOI: 10.1007/s10470-013-0054-6
- [31] Polanco-Martagón, S., Reyes-Salgado, G., Flores-Becerra, G., Tlelo-Cuautle, E., de la Fraga, L.G., Guerra-Gómez, I., et al., Selection of MOSFET sizes by fuzzy sets intersection in the feasible solutions space, *Journal of Applied Research and Technology*. [online]. 10(3), pp. 472-483, 2012. [date of reference Dec. 11th, 2015]. Available at: http://www.jart.ccadet.unam.mx/jart/vol10_3/selection_15.pdf
- [32] Santos-Azevedo, M.S., Pérez-Abril, I., León-Benítez, C.D., Cabral-Leite, J. and Holanda-Bezerra, U., Multiobjective optimization of the reactive power compensation in electric distribution systems. *DYNA*, 81(187), pp. 175-183, 2014. DOI: 10.15446/dyna.v81n186.40979
- [33] Ocampo, R.A.B., Flórez, C.A.C. y Zuluaga, A.H.E., Planeamiento multiobjetivo de la expansión de la transmisión considerando seguridad e incertidumbre en la demanda, *Ingeniería e Investigación*, [online]. 29(3), 74-78, 2009. [date of reference December. 11th, 2015]. Available at: <http://www.revistas.unal.edu.co/index.php/ingevin/article/view/15186/34199>
- [34] Zitzler, E., Thiele, L. and Bader, J., On Set-Based multiobjective optimization. *Evolutionary Computation*, IEEE Transactions on, 14(1), pp. 58-79, 2010. DOI: 10.1109/TEVC.2009.2016569
- [35] Mueller-Gritschneider, D., Graeb, H. and Schlittmann, U., A Successive approach to compute the bounded Pareto front of practical multiobjective optimization problems. *SIAM Journal on Optimization*, 20(2), pp. 915-934, 2009. DOI: 10.1137/080729013
- [36] Deb, K., Pratap, A., Agarwal, S. and Meyarivan, T., A fast and elitist multiobjective genetic algorithm: NSGA-II. *Evolutionary Computation*, IEEE Transactions on, 6(2), 182-197, 2002. DOI: 10.1109/4235.996017
- [37] Li, M., Liu, L. and Lin, D., A fast steady-state ϵ -dominance multi-objective evolutionary algorithm. *Computational Optimization and Applications*, 48(1), pp. 109-138, 2011. DOI: 10.1007/s10589-009-9241-x
- [38] Guerra-Gómez, I., Tlelo-Cuautle, E., McConaghy, T. and Gielen, G., Optimizing current conveyors by evolutionary algorithms including differential evolution. In: *16th Electronics, Circuits, and Systems, ICECS 2009 IEEE International Conference on*; 2009, pp. 259-262. DOI: 10.1109/ICECS.2009.5410989
- [39] Fernandez, F.V., Esteban-Muller, J., Roca, E. and Castro-López, R., Stopping criteria in evolutionary algorithms for multi-objective performance optimization of integrated inductors, In *Evolutionary Computation (CEC), 2010 IEEE Congress*; pp. 1-8, 2010. DOI:10.1109/CEC.2010.5586196
- [40] Keramat, M., Kielbasa, R., Modified latin hypercube sampling Monte Carlo (MLHSMC) estimation for average quality index. *Analog*

Integrated Circuits and Signal Processing, 19(1), pp. 87-98, 1999. DOI: 10.1023/A:1008386501079

S. Polanco-Martagón, received a BSc. degree in Computer Science at the Technological Institute of Puebla, Puebla, México, in 2006. He earned his MSc. Degree in Engineering, with honors, in the area of Information Technologies from the Institute of Technology in Puebla, Puebla, México in the year of 2010. He received his Dr. degree in July 2015 from the Centro Nacional de Investigación y Desarrollo Tecnológico (CENIDET), Cuernavaca, Morelos, México. Currently, he is working as a professor and researcher at the Universidad Politécnica de Victoria, Ciudad Victoria, Tamaulipas, México. His main areas of interest are artificial neural networks, fuzzy systems and bio-inspired algorithms. He has published about eight papers in book chapters, journals and conferences.
ORCID: 0000-0001-8473-0534

J. Ruiz-Ascencio, received his BSc. degree in physics from Universidad Nacional Autónoma de México (UNAM), Mexico, in 1971, a MSc. degree in Electrical Engineering in 1973, from Stanford University, USA, and the DPhil. degree in Engineering and Applied Science in 1989, from the University of Sussex, USA. He has been a researcher at the Institute of Applied Mathematics, IIMAS-UNAM, a full-time lecturer at the Autonomous University of Barcelona, automation project leader for Allen-Bradley, a researcher at the Instituto Tecnológico de Monterrey, and an invited scholar at McGill University's Center for Intelligent Machines in 2003 and again in 2010. He joined the Computer Science Department at the Centro Nacional de Investigación y Desarrollo Tecnológico (CENIDET) in 1995, where he is a member of the Artificial Intelligence Group. His current interests are machine vision and intelligent control.
ORCID: 0000-0001-8411-6817

M.A. Duarte-Villaseñor, received his BSc. degree of in Electronics from the Benemérita Universidad Autónoma de Puebla (BUAP), Mexico, in 2005. He received his MSc. and Dr. degree in the National Institute of Astrophysics, Optics and Electronics (INAOE), in 2007 and 2010. He has been a professor and researcher at the Instituto Tecnológico de Tijuana (ITT), Mexico since 2014. He is interested in the design of integrated circuits, artificial intelligence, automation and evolutionary algorithms. He has authored more than 20 works including book chapters, journals and conferences.
ORCID: 0000-0002-8858-8595



UNIVERSIDAD NACIONAL DE COLOMBIA

SEDE MEDELLÍN
FACULTAD DE MINAS

Área Curricular de Ingeniería
de Sistemas e Informática

Oferta de Posgrados

Especialización en Sistemas
Especialización en Mercados de Energía
Maestría en Ingeniería - Ingeniería de Sistemas
Doctorado en Ingeniería- Sistema e Informática

Mayor información:

E-mail: acsei_med@unal.edu.co
Teléfono: (57-4) 425 5365

Thermodynamic properties of moisture sorption in cassava flour

Alfredo A. Ayala-Aponte

Escuela de Ingeniería de Alimentos, Universidad del Valle, Cali, Colombia. alfredo.ayala@correounivalle.edu.co

Received: June 25th, de 2015. Received in revised form: November 6th, 2015. Accepted: December 3rd, 2015

Abstract

Thermodynamic adsorption properties (differential enthalpy and entropy, spreading pressure and net integral enthalpy and entropy) of cassava flour were determined at 25, 30 and 35°C to provide information about the water and energy requirements for the sorption process. The Guggenheim–Anderson–de Boer (GAB), Peleg and Halsey models were found to adequately describe the sorption characteristics, however the Peleg model provided the best fit. The enthalpy and entropy differentials decreased with increased equilibrium moisture content. The spreading pressure increased with increased water activity but decreased with increased temperature. The net integral enthalpy decreased with increased moisture content, with an asymptotic tendency to 0.2920 kJ/mol for a moisture content of 33.4% (d.m). In contrast, the net integral entropy values were negative and increased up to 0.1687 kJ/mol for a moisture content of 29.2% (d.m).

Keywords: adsorption isotherms; mathematical sorption models; thermodynamic sorption properties.

Propiedades termodinámicas de humedad de sorción en harina de yuca

Resumen

Se determinaron propiedades termodinámicas de adsorción (entalpía y entropía diferencial, presión de superficie, entalpía y entropía integral neta) en harina de yuca a 25, 30 y 35°C, para proporcionar información sobre requerimientos de agua y de energía en el proceso de sorción. Los modelos de GAB, Peleg y Halsey fueron los que mejor describieron las características de sorción, siendo el de Peleg el de mejor ajuste. Los diferenciales de entalpía y entropía disminuyeron con el incremento del contenido de humedad de equilibrio. La presión de superficie se incrementó con el aumento de la actividad de agua pero disminuyó con el incremento de temperatura. La entalpía integral neta decreció con el aumento de la humedad con una tendencia asintótica hasta 0.2920 kJ/mol para una humedad de 33.4 % (b.s); mientras que los valores de la entropía integral neta fueron negativos y se incrementaron hasta 0.1687 kJ/mol para una humedad de 29.2 % (b.s).

Palabras clave: isothermas de adsorción; modelos matemáticos de sorción; propiedades termodinámicas de sorción.

1. Introduction

Cassava (*Manihot esculenta*, Crantz) is a root rich in carbohydrates that is frequently consumed in tropical countries. It is processed in many ways for consumption by humans or animals [1]. In its fresh state, it is composed of approximately 35% carbohydrates, 62% water, 1-2% protein, 0.3% fat, 1-2% fiber and 1% minerals [2]. Due to its high water content, it is a perishable vegetable that requires preservation processes. Drying processes, which are widely applied to food to extend its storage stability, can be used as alternatives to preservation [3, 4]. As a preservation and processing alternative, cassava is transformed into flour after drying; however, to prevent the physicochemical and

microbial degradation of the flour, it should be stored under suitable conditions of temperature and relative humidity [1]. Hence, it is important to understand the relationship between a_w and the equilibrium moisture content (EMC) of cassava flour (sorption isotherms) and of food in general in order to predict the changes to their physical, chemical and microbiological properties that occur during storage and processing [4]. The sorption isotherm is applicable to the analysis and design of food drying, mixing and packaging, and can be used to determine optimal storage conditions, product shelf-life and proper packaging material [1]. Furthermore, the determination of sorption isotherms at various temperatures provides information about the thermodynamic sorption properties of the food-water vapor

How to cite: Ayala-Aponte, A.A., Thermodynamic properties of moisture sorption in cassava flour. DYNA 83(197), pp. 139-145, 2016.

system that affects the process, such as the differential enthalpy (isosteric heat of sorption), differential entropy, net integral enthalpy, spreading pressure and net integral entropy. These properties are useful to understand the behavior of water during the food-drying process and to optimize and quantify the energy requirements during mass and heat transfer [2]. They are also useful for providing information about the microstructure, water-sorption kinetics and interpretation of the physicochemical phenomena that occur at the water-solid interface [5]. According to Al-Muhtaseb et al. [6], the thermodynamic properties are also useful for determining the final dehydration point needed to achieve a stable food with the least amount of energy applied.

The differential enthalpy of sorption is useful to estimate the energy requirements for drying, as well as to provide information on the water state in food products. The differential entropy of sorption is proportional to the number of active sorption sites available at a specific energy level [7]. The spreading pressure is associated with the excess of free energy at the surface of the food solid, which means the increase of surface tension in the sorption active sites of water molecules [8,9]. The net integral enthalpy and entropy are used to explain the modes of moisture sorption in food products. The integral enthalpy provides information on the bond strength between water molecules and the solid [9], and the integral entropy describes the degree of disorder or randomness of the movement of water molecules [10]. These thermodynamic properties have been widely studied in many foods, such as potato [8], rice [7], mango powder [11], and starch [12], among others. However, the available literature is scarce for cassava flour. Notable research has been undertaken by Koua et al. [13] and Aviara et al. [2]; these studies were performed at different temperatures and with different varieties than those used in this study. Therefore, the purpose of this study was to determine the thermodynamic sorption properties (differential enthalpy, differential entropy, spreading pressure, net integral enthalpy and net integral entropy) in cassava starch at 25, 30 and 35°C.

2. Materials and methods

2.1. Preparation of the raw material

Cassava (*Manihot esculenta*, Crantz) roots were used, which were washed with potable water, manually peeled with steel knives, chopped and dried in an oven (MLW, Germany) with forced-air circulation at 65°C until constant weight. The dry parts were ground in a ball mill on a laboratory scale. The flour samples were packaged in polyethylene bags and stored in desiccators containing P₂O₅.

2.2. Determination of the adsorption isotherms

The adsorption isotherms were determined using a temperature controlled cabinet (Hot Pack, US), with an accuracy of ±1 °C at the selected temperatures of 25, 30 and 35°C using the gravimetric method, according to the COST 90 project [14]. This method consists of applying saturated saline solutions to maintain a given constant value of a_w of the flour samples, when equilibrium is achieved between the

atmosphere and the food.

Eight saturated saline solutions (LiCl, CHCOOK, MgCl₂, K₂CO₃, NaBr, SrCl₂, KCl and BaCl₂) were used, which were enclosed in hermetic containers and had a_w values that varied between 0.1117 and 0.9017 at various temperatures [15]. Petri boxes with 2.500 ± 0.002 g of cassava flour were introduced into each hermetic container with the corresponding saline solution. Tymol was placed into the containers with saline solutions that had a_w values higher than 0.65 to prevent microbial growth. The samples were periodically weighed every three days until a constant weight was achieved, to ensure equilibrium between the samples and the saline solutions. The samples were removed from the containers between the 19th and 25th day. The duration of the periodical weighing of the samples was less than 1 min to avoid any effect on the results. The EMC of the samples was determined using the oven method by AOAC [16]. The adsorption experiments for each temperature were performed in triplicate. The effect of temperature on the equilibrium moisture content was evaluated using analysis of variance (ANOVA) at a significance level of 95%, using the statistical software SPSS 11.

2.3. Modeling of the sorption isotherms

The Guggenheim, Anderson and Boer (GAB) (Eq. 1), Peleg (Eq. 2) and Halsey (Eq. 3) sorption models were used to fit the experimental EMC values.

$$EMC = \frac{x_0 C K a_w}{(1 - k a_w)(1 + (c - 1) k a_w)} \quad (1)$$

x_0 : moisture content of the monolayer

C : Guggenheim constant, a characteristic of the product and related to the heat of adsorption of the monolayer

K : constant related to the heat of adsorption of the multilayer

$$EMC = k_1 (a_w)^{n_1} + k_2 (a_w)^{n_2} \quad (2)$$

k_1, k_2, n_1, n_2 : model constants

$$EMC = \left[\frac{a}{-Ln(a_w)} \right] \left(\frac{1}{r} \right) \quad (3)$$

a and r : model constants

These models are widely used in the literature on food [8, 13, 17, 18]. The model parameters were determined by nonlinear regression using the program MATLAB v7.0. The mean relative error percentage (MRE, %) (Eq. 4) and the coefficient of correlation (r^2) were used as criteria to determine the best fit for the experimental sorption values. A model is considered acceptable if values for MRE are below 10% [19] and r^2 approaches 1.

$$(\%) MRE = \frac{100}{N} \sum_{i=1}^N \frac{|X_{ei} - X_{ci}|}{X_{ei}} \quad (4)$$

X_{ei} : moisture content (g water/g d.m)

X_{ci} : calculated moisture content (g water/g d.m)

N : number of samples

2.4. Determination of the Thermodynamic Properties

2.4.1. Differential enthalpy of sorption

The differential enthalpy of sorption is also called the isosteric heat of sorption (Q_{st}) and is determined from experimental values using the Clausius-Clapeyron equation (Eq. 5):

$$\left[\frac{\partial \ln(a_w)}{\partial \left(\frac{1}{T} \right)} \right]_{EMC} = -\frac{Q_{st} - \lambda}{R} = -\frac{q_{st}}{R} \quad (5)$$

where q_{st} is the net isosteric heat of sorption, R is the universal gas constant, and λ is the latent heat of vaporization of pure water, calculated for the fixed temperature of 30°C ($\lambda = 46.011$ kJ/mol), corresponding to the average of 25, 30 and 35°C.

By plotting $\ln(a_w)$ versus $1/T$ for various EMC values (between 5 and 35 kg water/kg d.m), linear relationships were obtained, with slopes equal to $-(Q_{st}-\lambda/R)$, from which Q_{st} was calculated.

2.4.2. Sorption entropy

The differential entropy of adsorption (S_d) was calculated by the Gibbs-Helmholtz equation (Eq. 6):

$$S_d = (Q_{st} - G)/T, \quad (6)$$

Where G is the Gibbs free energy, calculated using Eq. 7:

$$G = RT \ln a_w \quad (7)$$

By substituting equation (7) into (6), Eq. 8 is obtained:

$$-\ln(a_w) \big|_{EMC} = \frac{-Q_{st}}{RT} - \frac{S_d}{R}, \quad (8)$$

By plotting $\ln(a_w)$ versus $1/T$ for a constant water content, the S_d value is calculated from the intercept (S_d/R).

2.4.3. Net integral enthalpy of sorption

The value of the net integral enthalpy (Q_{in}) (Eq. 9) is analogous to the differential enthalpy of sorption, but it is calculated at a constant diffusion pressure. The value Q_{in} can be obtained from the slope of the plot of $\ln(a_w)$ vs. $1/T$:

$$\left[\frac{\partial \ln(a_w)}{\partial \left(\frac{1}{T} \right)} \right]_{\phi} \approx \frac{Q_{in}}{R} \quad (9)$$

2.4.4. Spreading pressure

The spreading pressure (ϕ) represents the free surface

energy for sorption and can be considered to be the difference in surface tension between the active adsorption sites in the solid and the water molecules [8, 9]. Its units are (J/m²). The value ϕ is not experimentally determined, but it can be obtained through an analytical procedure related to Eq. 10 [9].

$$\phi = \frac{KT}{A_m} \int_0^{a_w} \frac{\theta}{a_w} da_w \quad (10)$$

Where $\theta = EMC/x_0$ (x_0 : the moisture in the monolayer).

To avoid undefined values of ϕ for $a_w = 0$, the lower limit of $a_w = 0.05$ was taken in Eq. 10. The calculated spreading pressure values were fitted, adding the corresponding value to the range of a_w between 0 and 0.05, which was calculated assuming a linear relation (Henry's law) between EMC and a_w [9, 20]. Hence, by using an analytical procedure and the Halsey equation (Eq. 3) to describe the empirical relation between EMC and a_w , ϕ was calculated (Eq. 11). For this, the Halsey equation was substituted into Eq. 10 [2, 8, 20]:

$$\phi = \frac{KT}{A_m} a^{1/r} \left[\frac{1}{\left(\frac{1}{r} - 1 \right) (-\ln(a_w))^{1/r-1}} \right]_{0.05}^{a_w} \quad (11)$$

Where r and a correspond to the Halsey equation constants (Eq. 3), and ϕ was determined at the three study temperatures (25, 30 and 35°C).

2.4.5. Net integral entropy of sorption

The net integral entropy (S_{in}) describes the degree of disorder and randomness of the movement of water molecules adsorbed during sorption [9]. The S_{in} of a system can be calculated with equation 12 [10], considering that the geometrical mean of activity (a_w)* is used, obtained at a constant spreading pressure at various temperatures:

$$S_{in} = \frac{-Q_{in}}{R} - R \ln(a_w)^* \quad (12)$$

3. Results

3.1. Adsorption isotherms of water

The experimental adsorption isotherms of water from cassava flour at various temperatures are presented in Fig. 1. The samples presented an initial moisture content of 2.621 ± 0.011 g water /g dry matter. The standard deviation of the average EMC values in the adsorption isotherms at the three temperatures varied from 0.02 to 0.08. The isotherms presented a sigmoid shape (inflection point at low a_w values), classified as type II according to Brunauer, Deming and Troller, as reported by Iguedjtal [21].

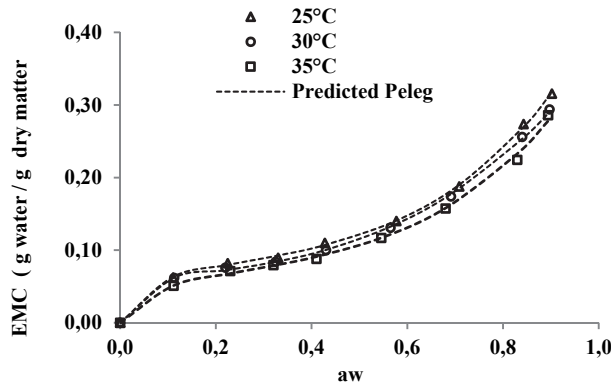


Figure 1. Adsorption isotherms of cassava flour: experimental and modeled using the Peleg model at 25, 30 and 35°C.

Source: The author

ANOVA indicated significant influence ($p < 0.05$) of the temperature on the EMC, indicating a decrease in EMC with increased temperature in the entire range of a_w . This means that cassava flour has lower water retention at higher temperature. In turn, with increased temperature for the same moisture level, a_w increases, which may increase the product deterioration rate [22]. Similar results have been observed in the water-adsorption isotherms in certain cassava products [17,23,24]. This influence of temperature on water retention has been attributed to the reduction or loss of active sites where water molecules interact with the surface of the food [25].

3.2. Modeling of the sorption isotherms

Table 1 presents the adsorption parameters, including the MRE and r^2 value for the GAB, Halsey and Peleg sorption models for the three temperatures. The three models presented good fits, with MRE values lower than 10% and high r^2 values. However, the Peleg model was best because it presented the lowest MRE and MRES values and highest r^2 values.

Table 1. Values of constants and statistical criteria for the fitting of experimental values of adsorption in cassava flour at various temperatures.

Model	Constants	Temperature (°C)		
		25	30	35
GAB	X ₀	0.0731	0.0728	0.0632
	C	34.9600	15.550	30.390
	K	0.8919	0.8329	0.8677
	R ²	0.9929	0.9970	0.9957
	MRE %	5.1541	5.8874	3.1647
Halsey	a	0.0181	0.0161	0.0122
	r	1.7412	1.7640	1.8022
	R ²	0.9857	0.9879	0.9915
	MRE %	3.3289	3.2156	4.1335
Peleg	k ₁	0.1246	0.0905	0.1152
	n ₁	0.3082	0.1758	0.3698
	k ₂	0.2922	0.2789	0.2581
	n ₂	3.8320	2.9840	3.8616
	R ²	0.9994	0.9992	0.9962
	MRE %	1.5137	1.7251	2.5199

Source: The author

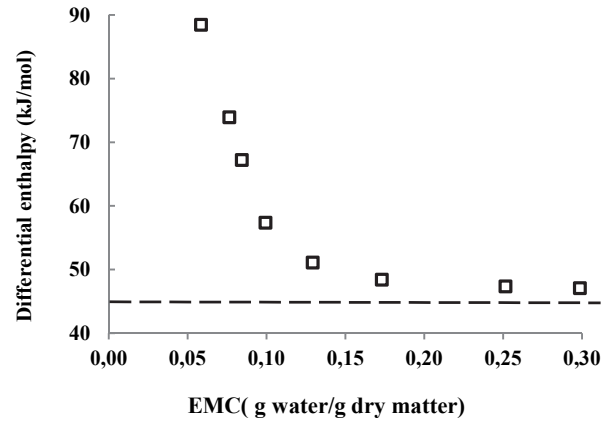


Figure 2. Variation in the differential enthalpy of sorption with equilibrium moisture content in cassava flour.

Source: The author

A comparison of the EMC experimental values for cassava flour and those predicted with the Peleg model is presented in Fig. 1. It indicates the good fit of this model along the entire range of a_w .

The GAB model has the advantage of providing information about the moisture content at the monomolecular layer or monolayer (x_0) (Table 2). The x_0 value indicates the amount of water that is strongly adsorbed in the active sites of the food, and it is considered to be the value at which the food is most stable during storage [1].

3.3. Thermodynamic properties

3.3.1. Differential enthalpy of sorption (isosteric heat)

Fig. 2 shows the change in Q_{st} with EMC. Notice that when EMC decreases, Q_{st} increases, ranging from 30 to 5.8 g water/g dry matter, respectively. This indicates the strong retention of water molecules in the active sites on the solid phase of the food. According to some researchers [26], the energy interaction level between water and the solid phase at low EMC is higher than the interaction between water molecules at high EMC. Similar behavior was observed in sorption isotherms in various cassava products [17,23,24] and in other foods, such as corn [27], yellow pitahaya [28], plantain pulp [29] and orange powder [18].

In the entire EMC range, Q_{st} is higher than the heat of vaporization of pure water ($\lambda_{(30^\circ\text{C})} = 46.011$ kJ/mol), and it has similar values at high EMC. This indicates that the energy of interaction between the water molecules at the adsorption sites (active points) at the surface of the food is higher than the energy at which the molecules of pure water exist in the liquid phase. In the adsorption processes during storage, Q_{st} provides information on the water-solid interaction and the chemical, physical and microbiological state of the food [8].

3.3.2. Differential entropy of sorption

Fig. 3 indicates the strong dependence of the differential entropy (ΔS) on the EMC, with a maximum of 129.0 J/mol·K

at an EMC of 0.058 g water/g d.m, past which the value decreases with increasing moisture. This result has been observed in many products, such as potato flakes [30], potato [8], prickly pear [31] and potato starch [12]. The ΔS value is proportional to the number of active sorption sites at a determined energy level [7]. This means that the number of available sorption sites decreases with increased moisture because these active sites become covered with water molecules, decreasing their mobility. According to certain researchers [32,33], the decrease in the differential entropy with increased moisture is associated with the restriction of the molecular movement of water. According to other researchers [34], the differential entropy can be used to understand the moisture adsorption process associated with dilution and crystallization. Rotstein [35] argues that ΔS can be used in an energy balance, providing information on the energy use in food processing.

3.3.3. Spreading surface

The evolution of spreading pressure (Φ) with temperature is presented in Fig.4. The value Φ increases with increased water activity, and at a certain a_w value, it decreases with rising temperature, although this is not very notable between 25 and 30°C.

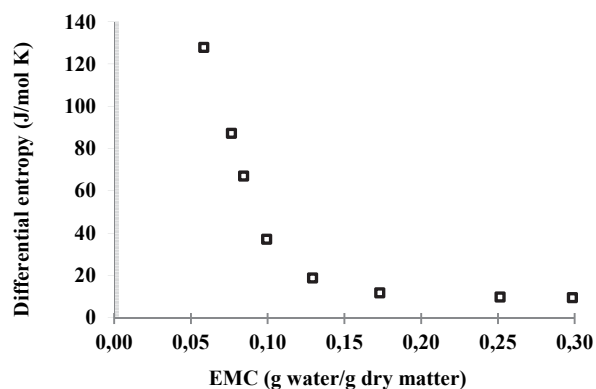


Figure 3. Changes in the differential entropy of sorption with EMC in cassava flour.
Source: The author

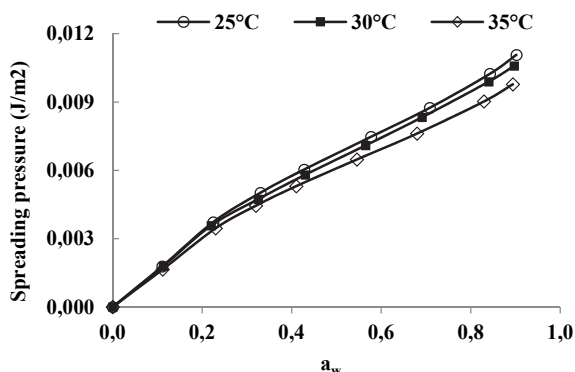


Figure 4. Evolution of the spreading pressure with increased a_w at various temperatures.
Source: The author

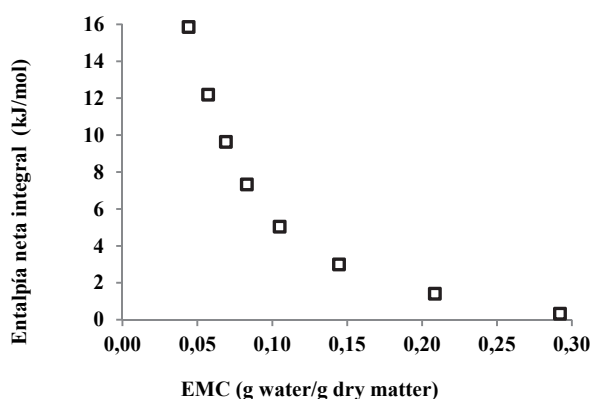


Figure 5. Net integral enthalpy of cassava flour as a function of EMC.
Source: The author

This behavior has been presented in sorption studies of water in cassava [2] and starch [12]. The value Φ represents the excess free energy at the surface, indicating an increase in the surface tension at the sorption sites, caused by the adsorbed molecules [9].

3.3.4. Net integral enthalpy of sorption

The changes in the net integral enthalpy (Q_{in}) with equilibrium moisture content are shown in Fig. 5.

This integral was determined the same way as the differential enthalpy of sorption, but at a constant spreading pressure. The value Q_{in} decreases with increased EMC, varying from 15.86 kJ/mol for an EMC of 0.004 g water/g d.m. to 0.334 kJ/mol for 0.2920 g water/g d.m. This thermodynamic behavior can be attributed to the saturation of the active sorption points (monolayer moisture) [9], as the water is more easily adsorbed in the active sites of the external surface of cassava flour at low moisture contents, reaching the maximum value of Q_{in} . This maximum value is an indication of the greater water-cassava flour interaction, which means that the sorption bond sites are stronger [26]. With the increase in the equilibrium moisture, the decrease in the net integral enthalpy is evident, which is attributed to the lower favorability of the active sorption sites because they are covered with water, forming multilayers [9]. According to these results, the variation in the net integral enthalpy with EMC indicates that the level of the solid-water interaction is higher than the interaction between water molecules. According to Rizvi and Benado [36], the magnitude of the enthalpy with high moisture content reflects the presence of free water. Similar interaction between Q_{in} and EMC has been reported in various foods, such as potato [8], potato starch [12] and macaroni [37]. The determination of integral enthalpy is useful in quantifying the energy required for drying, from an initial moisture content to a final moisture content of the food [26]. It also provides information on the microstructure of the food and the theoretical interpretation of the physical phenomena that occur at the water-food interface [38].

3.3.5. Net integral entropy of sorption

Fig. 6 presents the net entropy of sorption (S_{in}) as a function of the moisture content. All S_{in} values were negative,

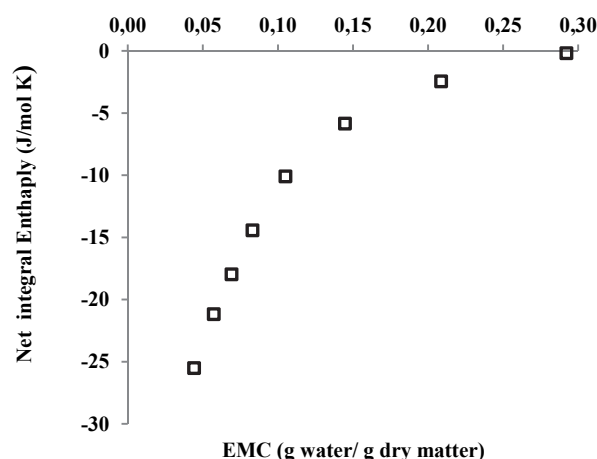


Figure 6. Variation of the net integral entropy of cassava flour with EMC.
Source: The author

which might indicate the existence of chemical adsorption or structural modification of the adsorbent or the external surface of the solid [20]. S_{in} increased with increased equilibrium moisture content, varying from a minimum value of $-25.50 \text{ J/mol} \cdot \text{K}$ for an EMC value of $0.044 \text{ g water/g d.m.}$ to $-0.1618 \text{ J/mol} \cdot \text{K}$ for $0.2920 \text{ g water/g d.m.}$ This result may be associated with the greater mobility of water molecules, promoting the formation of multilayers. Thus, the net integral entropy tends to approach that of free liquid water [39]. Similar behavior was observed in the sorption of water in mango [11] and cassava [2] powder. The minimum net integral entropy value can be considered the maximum point of stability for food during storage [11]. As the water molecules are strongly bonded with the adsorbent at this zone, they are more ordered and less available to participate in deterioration reactions [40]. Therefore, this minimum S_{in} value is found at an EMC of 4.4% (dry matter), a value relatively close to the moisture of the monomolecular layer (x_0) predicted with the GAB model (Table 1), which varied between 6.3 and 7.3% (dry matter).

4. Conclusions

The adsorption isotherms of cassava flour determined at 25, 30 and 35°C demonstrated lower water retention capacity at higher temperature. The GAB, Halsey and Peleg models are appropriate for the prediction of EMC in cassava starch between 25 and 35°C , where the Peleg model provides the best prediction. All thermodynamic properties presented a correlation with the equilibrium moisture content. The differential enthalpy, differential entropy and net integral enthalpy decreased with increased moisture. The net integral entropy increased with increased moisture, and the spreading pressure decreased with temperature and increased with a_w . According to the monolayer moisture and the minimum net integral entropy value, cassava flour presents maximum stability at an EMC between 4.4 and 7.3% (dry matter). These adsorption findings in cassava flour are of interest in the determination of the optimal storage conditions, the prediction of shelf life and the optimization and

quantification of energy requirements during the transfer of mass and heat during drying.

References

- [1] Ayala, A., Estimación de las isothermas de adsorción y del calor isostérico en harina de yuca, *Biología en el Sector Agropecuario y Agroindustrial*, [Online]. 9(1), pp. 88-96, 2011. Available at: <http://www.scielo.org.co/pdf/bsaa/v9n1/v9n1a11.pdf>
- [2] Aviara, N. and Ajibola, O., Thermodynamics of moisture sorption in melon seed and cassava, *Journal of Food Engineering*, [Online]. 55(2), pp. 107-113, 2002. Available at: <http://www.sciencedirect.com/science/article/pii/S0260877402000237>
- [3] Onayemi, O. and Oluwamukomi, M., Moisture equilibria of some dehydrated Cassava and Yam Products, *Journal of Food Process Engineering*, 9(3), pp. 191-200, 1987. DOI: 10.1111/j.1745-4530.1987.tb00124.x
- [4] Moreira, R., Chenlo, F., Torres, M. and Vallejo, N., Thermodynamic analysis of experimental sorption isotherms of loquat and quince fruits, *Journal of Food Engineering*, 88(4), pp. 514-521, 2008. DOI: 10.1016/j.jfoodeng.2008.03.011
- [5] Rizvi, S. and Benado, A., Thermodynamic properties of dehydrated foods, *Food technology*, 38(3), pp. 83-92, 1984.
- [6] Al-Muhtaseb, A., McMinn, W. and Magee, T., Water sorption isotherms of starch powders: Part 1: Mathematical description of experimental data, *Journal of Food Engineering*, 61(3), pp. 297-307, 2004. DOI: 10.1016/S0260-8774(03)00133-X
- [7] Toğrul, H. and Arslan, N., Moisture sorption behaviour and thermodynamic characteristics of rice stored in a chamber under controlled humidity, *Biosystems Engineering*, 95(2), pp. 181-195, 2006. DOI: 10.1016/j.biosystemseng.2006.06.011
- [8] McMinn, W. and Magee, T., Thermodynamic properties of moisture sorption of potato, *Journal of Food Engineering*, 60(2), pp. 157-165, 2003. DOI: 10.1016/S0260-8774(03)00036-0
- [9] Fasina, O., Ajibola, O. and Tyler, R., Thermodynamics of moisture sorption in winged bean seed and gari, *Journal of Food Process Engineering*, 22(6), pp. 405-418, 1999. DOI: 10.1111/j.1745-4530.1999.tb00496.x
- [10] Mazza, G. and LeMaguer, M., Water sorption properties of yellow globe onion (*Allium cepa* L.), *Canadian Institute of Food Science and Technology Journal*, 11(4), pp. 189-193, 1978. DOI: 10.1016/S0315-5463(78)73269-4
- [11] Cano, D., Villa, H., Telis, J., Vázquez, H. and Nicoletti, V., Influence of alternative drying aids on water sorption of spray dried mango mix powders: A thermodynamic approach, *Food and Bioprocess Processing*, 93, pp. 19-28, 2015. DOI: 10.1016/j.fbp.2013.10.005
- [12] Al-Muhtaseb, A., McMinn, W. and Magee, T., Water sorption isotherms of starch powders. Part 2: Thermodynamic characteristics, *Journal of Food Engineering*, 62(2), pp. 135-142, 2004. DOI: 10.1016/S0260-8774(03)00202-4
- [13] Koua, B., Koffi, P., Gbaha, P., & Toure, S. Thermodynamic analysis of sorption isotherms of cassava (*Manihot esculenta*), *Journal of food science and technology*, 51 (9), pp. 1711-23, 2014. <http://dx.doi.org/10.1007/s13197-012-0687-y>
- [14] Wolf, W., Spiess, W. and Jung, G., Standardization of isotherm measurements. In: Simatos, D. and Multon, J.L. (Eds), *Properties of Water in Foods*. Martinus Nijhoff Publishers: Dordrecht, Netherlands, 1985, pp. 661-677.
- [15] Greenspan, L., Humidity fixed points of binary saturated aqueous solutions, *Journal of Research of the National Bureau of Standards Section A: Physics and Chemistry*, 81A (1), pp. 89-96, 1977. DOI: 10.6028/jres.081A.011
- [16] AOAC, Official methods of analysis of the Association of Official Analytical Chemists International. Moisture in dried fruits. Method 934.06, Arlington, USA. 1990. pp. 911-912.
- [17] Sanni, L., Atere, C. and Kuye, A., Moisture sorption isotherms of fufu and tapioca at different temperatures, *Journal of Food Engineering*, 34 (2), pp. 203-212, 1997. DOI: 10.1016/S0260-8774(97)00072-1
- [18] Sormoli, M. and Langrish, T., Moisture sorption isotherms and net isosteric heat of sorption for spray-dried pure orange juice powder,

- LWT - Food Science and Technology, 62(1), pp. 875-882, 2015. DOI: 10.1016/j.lwt.2014.09.064
- [19] Lomauro, C., Bakshi, A. and Labuza, T., Evaluation of food moisture sorption isotherm equations Part I: Fruit, vegetable and meat products, LWT - Food Science and Technology, [Online]. 18(2), pp. 111-117, 1985. Available at: <http://congttythienquang.com/data/news/Lomauro1985%20Sorption%20Isotherm%20Fruit%20Vegetable%20Meat.pdf>
- [20] Iglesias, H. and Chirife, J., A model for describing the water sorption behavior of foods, Journal of Food Science, 41(5), pp. 984-992, 1976. DOI: 10.1111/j.1365-2621.1976.tb14373.x
- [21] Iguedjal, T., Louka, N. and Allaf, K., Sorption isotherms of potato slices dried and texturized by controlled sudden decompression, Journal of Food Engineering, 85(2), pp. 180-190, 2008. DOI: 10.1016/j.jfoodeng.2007.06.028
- [22] Van den Berg, C. and Bruin, S., Activity and its estimation in food systems. In: L.B. Rockland, F. and Stewart. (Eds.), Water Activity: Influence on Food Quality, Academic Press, New York, 1981. pp. 147-177.
- [23] Adebawale, A., Sanni, L., Awonorin, S., Daniel, I. and Kuye, A., Effect of cassava varieties on the sorption isotherm of tapioca grits, International Journal of Food Science & Technology, 42(4), pp. 448-452, 2007. DOI: 10.1111/j.1365-2621.2007.01261.x
- [24] Ikhu-Omoregbe, D., Comparison of the sorption isotherm characteristics of two Cassava Products, International Journal of Food Properties, 9(2), pp. 167-177, 2006. DOI: 10.1080/10942910600592026
- [25] Mazza, G., Thermodynamic considerations of water vapour sorption by horseradish roots, Lebensmittel-Wissenschaft & Technologie, 13, pp. 13-17, 1980.
- [26] Velázquez, S., Figueira, A., Rodríguez, M., Román, A., Carrillo, H. and Pérez, C., Sorption isotherms, thermodynamic properties and glass transition temperature of mucilage extracted from chia seeds (*Salvia hispanica* L.), Carbohydrate Polymers, 121, pp. 411-419, 2015. DOI: 10.1016/j.carbpol.2014.11.068
- [27] Samapundo, S., Devlieghere, F., De Meulenaer, B., Atukwase, A., Lamboni, Y. and Debevere, J., Sorption isotherms and isosteric heats of sorption of whole yellow dent corn, Journal of Food Engineering, 79(1), pp. 168-175, 2007. DOI: 10.1016/j.jfoodeng.2006.01.040
- [28] Ayala-Aponte, A., Serna-Cock, L. and Rodríguez-de La Pava, G., Moisture adsorption isotherms in yellow pitahaya (*Selenicereus megalanthus*), DYNA, [Online]. 78(170), pp. 7-14, 2012. Available at: http://www.scielo.org.co/scielo.php?pid=S0012-73532011000600001&script=sci_arttext
- [29] Ciro, H., Osorio J.A. and Cortés A., Determination of the isosteric heat to plantain pulp (*Musa paradisiaca*) by sorption isotherms, DYNA, [Online]. 75 (156), pp. 127-134, 2008. Available at: <http://dyna.unalmed.edu.co/es/ediciones/156/articulos/a12v75n156/a>
- [30] Lago, C., Liendo, M. and Zapata, C., Thermodynamic sorption properties of potato and sweet potato flakes, Food and Bioprocess Processing, 91(4), pp. 389-395, 2013. DOI: 10.1016/j.fbp.2013.02.005
- [31] Hassini, L., Bettaieb, E., Desmorieux, H., Sandoval, S. and Touil, A., Desorption isotherms and thermodynamic properties of prickly pear seeds, Industrial Crops and Products, 67, pp. 457-465, 2015. DOI: 10.1016/j.indcrop.2015.01.078
- [32] Goneli, A., Corrêa, P., Oliveira, G. and Botelho, F., Water desorption and thermodynamic properties of okra seeds, American Society of Agricultural and Biological Engineers, 53(1), pp. 191-197, 2010. DOI: 10.13031/2013.29486
- [33] Eim, V., Rosselló, C., Femenia, A. and Simal, S., Moisture sorption isotherms and thermodynamic properties of carrot, International Journal of Food Engineering, 7(3), pp. 1-14, 2011. DOI: 10.2202/1556-3758.1804
- [34] Aviara, N., Ajibola, O. and Dairo, U., PH—Postharvest technology: Thermodynamics of moisture sorption in sesame seed, Biosystems Engineering, 83(4), pp. 423-431, 2002. DOI: 10.1006/bioe.2002.0124
- [35] Rotstein, E., The exergy balance: A diagnostic tool for energy optimization, Journal of Food Science, 48(3), pp. 945-950, 1983. DOI: 10.1111/j.1365-2621.1983.tb14936.x/pdf
- [36] Rizvi, S. and Benado, A., Thermodynamic analysis of drying foods, Drying Technology, 2(4), pp. 471-502, 1983. DOI: abs/10.1080/07373938408959849
- [37] Arslan, N. and Toğrul, H., Modelling of water sorption isotherms of macaroni stored in a chamber under controlled humidity and thermodynamic approach, Journal of Food Engineering, 69(2), pp. 133-145, 2005. DOI: 10.1016/j.jfoodeng.2004.08.004
- [38] Rizvi, S., Thermodynamics of foods in dehydration. In: Rao, M.A., Rizvi, S.S.H. and Datta, A.K. eds., Engineering Properties of Foods, CRC Press, New York, 1986, pp. 133-214.
- [39] Benado, A. and Rizvi, S., Thermodynamic properties of water on rice as calculated from reversible and irreversible isotherms, Journal of Food Science, [Online]. 50(1), pp. 101-105, 1985. Available at: <http://onlinelibrary.wiley.com.bd.univalle.edu.co/doi/10.1111/j.1365-2621.1985.tb13286.x/epdf>
- [40] Beristain, C., Azuara, E. and Vernon, E., Effect of water activity on the stability to oxidation of spray-dried encapsulated orange peel oil using mesquite gum (*Prosopis juliflora*) as wall material, Journal of Food Science, [Online]. 67(1), pp. 206-211, 2002. Available at: <http://onlinelibrary.wiley.com.bd.univalle.edu.co/doi/10.1111/j.1365-2621.2002.tb11385.x/epdf>

A.A. Ayala-Aponte, received his BSc. Eng in Agricultural Engineering in 1993 from the Universidad del Valle, Cali, Colombia and a PhD in Science and Food Technology in 2011 from the Universidad Politécnica de Valencia, España. He is a professor in the area of Food Technology and Engineering, in the Universidad del Valle, Cali, Colombia. His research interests include: preservation and food processing.
ORCID: 0000-0003-0310-3577



UNIVERSIDAD NACIONAL DE COLOMBIA

SEDE MEDELLÍN
FACULTAD DE MINAS

Área Curricular de Ingeniería
Química e Ingeniería de Petróleos

Oferta de Posgrados

Maestría en Ingeniería - Ingeniería Química
Maestría en Ingeniería - Ingeniería de Petróleos
Doctorado en Ingeniería - Sistemas Energéticos

Mayor información:

E-mail: qcaypet_med@unal.edu.co
Teléfono: (57-4) 425 5317

Cr/CrN bilayer coating behavior under erosion-corrosion degradation by jet impingement testing

Johanna Alegría-Ortega ^a, Luz Marina Ocampo-Carmona ^a, Fernando Rodríguez ^b & Elena Forlerer ^b

^a Grupo de Ciencia y Tecnología de Materiales, Facultad de Minas, Universidad Nacional de Colombia, Medellín, Colombia. jaalegri@unal.edu.co

^b Laboratorio de Recubrimientos y Tribología, Gerencia de Materiales, Comisión Nacional de Energía Atómica, Buenos Aires, Argentina. forlerer@cnea.gov.ar

Received: July 29th, 2015. Received in revised form: November 19th, 2015. Accepted: December 10th, 2015

Abstract

Cathodic arc technique was used to deposit a Cr/CrN bilayer coating on AISI 440C stainless steel. The morphology and composition of the coating were verified by Scanning Electron Microscope and X-Ray Diffraction. Hardness and adhesion of the coating were also studied. In the erosion-corrosion wear, a slurry jet with impingement angle 90° at 4 and 8 m/s was used to impact upon bare and coated steel. All the wear mechanisms were evaluated according to ASTM G-119.

The results indicate that the synergism wear rate is the most aggressive wear component, exceeding 40% of the total mass loss for coating steel and 60% of mass loss for bare steel. An increase in the impact velocity increases mass loss for both bare and coated steel. In comparison, the Cr/CrN coating had better corrosion and erosion-corrosion resistance.

Keywords: Multilayer coating, Stainless Steel, Cathodic Arc, erosion-corrosion.

Comportamiento de un recubrimiento bicapa de Cr/CrN bajo degradación por erosión-corrosión por ensayo de impacto con chorro

Resumen

La técnica de deposición catódica fue utilizada para depositar un recubrimiento bicapa de Cr/CrN sobre acero AISI 440C. La morfología y composición del recubrimiento fueron evaluadas con microscopía electrónica de barrido y difracción de rayos-x. La dureza y adhesión del recubrimiento también fueron estudiadas.

Se siguió el procedimiento de la norma ASTM G-119 para evaluar el desgaste por erosión-corrosión usando un chorro de lodo impactando a 90° a una velocidad de 4 y 8 m/s para impactar sobre el acero desnudo y recubierto.

Los resultados muestran que el sinergismo es la componente de desgaste más agresiva, ésta excede en un 40% la pérdida de masa total para el acero recubierto y el 60% de la pérdida de masa total para el acero desnudo. El incremento en la velocidad de impacto incrementa la pérdida de masa tanto en el acero desnudo como en el recubierto. Comparando, el recubrimiento de Cr/CrN presentó mayor resistencia a la corrosión y al desgaste por erosión-corrosión.

Palabras claves: Recubrimientos multicapas, Acero inoxidable, Arco catódico, erosión-corrosión.

1. Introduction

In aqueous conditions the presence of ions (CO_3^{2-} , Cl^- , SO_4^{2-}) and particles harder than the surface can cause rapid deterioration of metal parts, which could be mitigated by employing stainless steel or potentiostatic protection. However, a high flux of certain ions added with harder particles impacting

the surface promote the breakdown of the protective layer.

Stainless steels and corrosion inhibitors enable the reduction of the mass loss associated with the pure corrosion effect, allowing erosion loss to be the principal component of wear, although their protective capacity lasts until the flow velocity and turbulence affect the integrity of the surface. Alternatively, ceramic coatings have been used because their

How to cite: Alegría-Ortega, J., Ocampo-Carmona, L.M., Rodríguez, F. and Forlerer, E., Cr/CrN bilayer coating behavior under erosion-corrosion degradation by jet impingement testing. DYNA 83(197), pp. 145-152. June, 2016 Medellín. ISSN 0012-7353 Printed, ISSN 2346-2183 Online

high toughness and good adhesion can protect the surface against erosion, they could also extend the life of the part under corrosion due to their superior chemical resistance [1].

In general, erosion-corrosion (EC) studies reveal that the increase in the loss of material due to EC is greater than the sum of pure erosion and pure corrosion, which verifies the existence of a combined action [2-5]. It has also been found that for some metallic materials, the passivation zone promotes a decrease in the erosion rate, i.e. passive films generate some protection against erosive particles.

Stack et al. [3] suggest that under EC wear, wear rates are limited to four stages: corrosion stage 1, corrosion stage 2, increased wear by pure corrosion, and pure erosion. In the last stage, the effect is due to an increase in temperature and erosion velocity. For tests in which the mass loss is measured gravimetrically, corrosion stage 1 is characterized by mass loss associated with the dissolution or degradation of the natural protective Cr_2O_3 oxide; corrosion stage 2 shows a mass gain due to the formation of new oxides; in the increased wear by pure corrosion and pure erosion stages, the mass losses are more pronounced, due to the convective and diffusive flow action and the kinetic energy of the particles impinging on the surface, causing the generation of subsurface cracks. In order to decrease the mass loss rate due to stages 3 and 4, the deposition of ceramic coatings has been proposed as an alternative because of their high hardness, and their chemical and thermal inertia, which generate lower mass loss rates [1,5-6].

According to ASTM G-119 [2] the mass loss rate by E/C in aqueous media is described by the eq. (1):

$$TML = E_0 + C_0 + S \quad (1)$$

In which, TML: total mass loss rate, E_0 : Mass loss rate by pure erosion, C_0 : Mass loss rate by pure corrosion and S: erosion-corrosion synergism.

The synergism component can be seen as the sum of the interactions between the processes, where ΔC_w that is the change in corrosion rate due to wear and ΔW_c is the change in wear rate due to corrosion as is described by eq. (2):

$$TML = E_0 + C_0 + \Delta E_c + \Delta C_w \quad (2)$$

The additive effect, ΔC_w , refers to the change in corrosion rate due to wear, a positive additive effect are related with the generation of rugosity, these stressed zones behave like electrochemical cells increasing material loss; when a negative additive effect occurs it means there are compressive stresses due to the impact of the erosive particles, this stress state alters the surface electrochemical behavior triggering a lower mass loss rate [3-6].

The synergistic effect, ΔW_c , refers to the enhancement of wear due to corrosion, when this term is negative (antagonism) the corrosion products during wear are providing protection to the surface.

With PVD techniques, there has been success in obtaining coatings at low temperatures (below 400°C), high deposition rates, and the possibility to deposit multilayers or super-networks that exhibit better performance due to the combination of the properties of their components and the laminated structure [7-10]. Indeed, with the diversity of

techniques that have emerged for the deposition of hard films, currently there are EC wear studies being carried out on various nitrides and carbides of transition elements, as well as on micro and nanometric multilayers of these, in order to propose better alternatives for the mitigation of the wear of metallic parts [5,6,11-13].

2. Materials and methods

The experimental procedure used can be divided into two steps: coating deposition and evaluation of coating performance under erosion-corrosion experiments.

The first step can be described as follows: Heat treatment of the steel, substrate characterization, coating deposition. The details of each stage will be presented below.

2.1. Heat treatment of AISI 440C steel

Samples of AISI 440C of 12.7 mm in diameter and 5 mm in thickness were subjected to cryo-quenching and tempering treatments (Fig. 1) in the Coatings and Tribology Laboratory (DRyT) of CNEA, to improve hardness and wear resistance. A cryo-quenching process was employed in order to reduce the retained austenite [14], which is metastable at room temperature and could transform to new martensite during quenching after coating, resulting in volume change, which generates internal stresses, and consequently unacceptable dimensional changes and delayed cracking [15,16].

2.2. Coating deposition

The samples were polished with abrasive paper from 80-800 mesh size. Some of the samples were used for the deposition of coatings and other bare samples were used as a reference for erosion-corrosion tests.

Two bilayers of Cr/CrN were deposited with cathodic arc equipment Nissin Electric CO, LTD RMP-R3-J, in the Coatings and Tribology Laboratory (DRyT) of CNEA. For the deposition of the coatings, the vacuum back-pressure was 3×10^{-5} Torr (4×10^{-3} Pa) using a two-stage rotative mechanical pump for the rough vacuum, coupled with a diffusion pump. The target-substrate working distance was

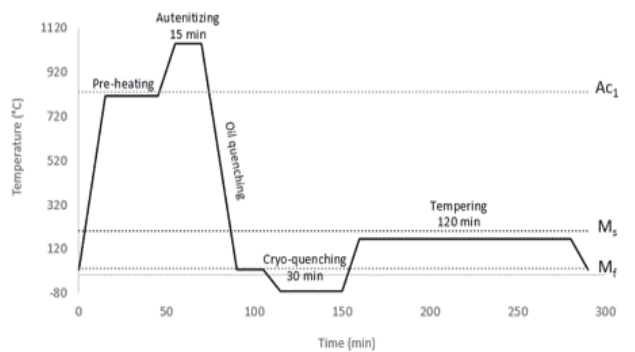


Figure 1 Heat Treatment of AISI 440C steel. Ac1: initiation of austenite transformation temperature, Ms: initiation of martensitic transformation temperature start, Mf: end of martensitic transformation temperature.

Source: The authors.

15 cm. The ionic cleaning was made with Vbias: -800V with Ar gas for 15 min. Then, the deposition of two bilayers of Cr + CrN, each layer under Vbias: -400V for 1.5 min with a flow rate of 40 cm³ of Ar gas in the Cr layer deposition and Vbias: -125V for 15 min with a mixture of N₂ (36 cm³) and Ar (2 cm³) gases in the CrN layer deposition, was carried out. Between the two bilayers, the voltage sources were turned off without breaking the vacuum.

2.3. Evaluation of coating performance under erosion-corrosion experiments

The jet impingement test was executed in a cell that holds 5 L of solution and it was recirculated by means of a peristaltic pump. The nozzle is made of Nylon and has an outlet diameter of 2 mm, at 90° from the sample.

The cell, Fig. 2, was coupled to a potentiostat Gamry 600 to generate potentiodynamic polarization curves in order to evaluate the corrosion rate. The reference electrode (RE) of saturated calomel (SCE) was put into a glass shell with a Vycor tip, which acts as a membrane and facilitates electrochemical measurements without any contamination of the electrode. The auxiliary electrode (AE) of platinum was located near the working electrode (sample), to reduce the ohmic drop during the measurement.

Potentiodynamic curves were initiated once the open circuit potential (OCP) was stabilized. The exposure area of the sample was 78.54 mm². A sweep was carried out from -0.6 to 0.6 V vs SCE using a scan rate of 1 mV/s. In all the tests, the samples were immersed for 30 min in the solution or in the slurry under the action of the jet. The electrochemical parameters were calculated with Tafel extrapolation using E-Chem Analysis-Gamry software.

The corrosion tests were performed in a 3.5%wt. NaCl solution, while in the erosion and EC tests slurry composed

Table 1.

Nomenclature used for corrosion-erosion test.

Material type	Test type	Impact velocity
a: 440C steel	TML: Total material loss by erosion-corrosion	4: 4m/s
R: Cr/CrN Bilayer coating	W ₀ : pure erosion C ₀ : pure corrosion C _w : electrochemical corrosion rate during the erosion – corrosion test	8: 8m/s

Source: The authors.

of a solution with 10% wt. of silica particles of 0.300-0.212 mm (AFS 50/70) size was used. In the erosion test, a cathodic voltage of -1V vs E_{corr} was supplied in order to avoid mass loss associated with corrosion. In the slurry tests, a paddle stirrer was used in order to facilitate the suspension of the silica particles and supply a homogeneous solution to the pump.

The jet impingement velocities were 4 and 8 m/s. The test was performed at room temperature (25-28°C). The Nylon sample holder was of the Avesta cell type in order to avoid crevice corrosion. The sample was placed 5 mm from the nozzle. The mass loss erosion and EC test were calculated gravimetrically.

The nomenclature used in the samples was as follows: Table 1.

3. Results and Discussion

The characterization of the substrates after the heat treatment, the coating and the electrochemical measurements of mass loss of the coated and uncoated samples after completing the tests, as well micrographs will be reported.

3.1. Characterization of substrates

The monograph in Fig. 3 shows that the steel was etched with Beraha attack [17] and that there is a dark matrix consistent with very fine martensitic matrix exhibited by

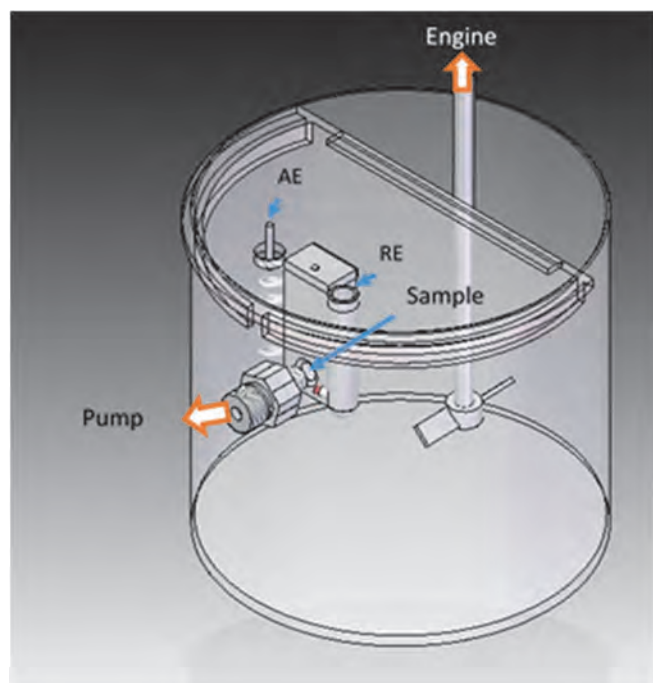


Figure 2 Erosion corrosion cell.

Source: The authors.

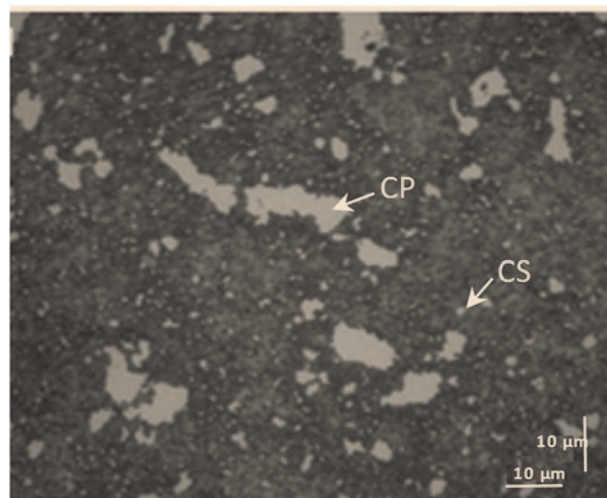


Figure 3 Bright field micrograph of 440C steel after cryo-quenching and tempering treatment. CP: primary carbides, CS: secondary carbides. Etched with Beraha.

Source: The authors.

440C steel thermal treated [18], and bimodal size distribution of white areas, which correspond to carbides. A martensitic matrix was checked by measuring the microhardness, by means of a microdurometer Akashi MVK-H2, which gave a result of (679 ± 71) HV0.025.

3.2. Coating characterization

The evaluation of the properties of the coating was performed on a standard sample for each batch of coatings, using the equipment of the CTL-CNEA.

The hardness measurement was (1300 ± 80) HV0.025. Fig. 4 shows a micrograph of thickness coating measured with CaloTest equipment CSEM S/N 03-134. The partial and full thickness measured was $3,2 \mu\text{m}$ and $6,6 \mu\text{m}$ respectively.

The scratch test was performed with CSEM Scratch-Tester S-N 26-393 equipment with a diamond type C Rockwell indenter. The indenter was moved at a speed of 10 mm/min, linearly varying the load from 0 to 60 N with a table speed of 100 N/mm. The critical load was determined by optical microscope, and the lineal relation of load vs. table speed was $Lc_1: (38 \pm 10)$ N.

To identify the phases present and their preferred orientation, diffraction measurements were performed on an X-ray diffractometer X'pert Panalytical with Cu-K α radiation. The diffraction pattern, Fig. 5, shows the presence of a CrN phase (JCPDF00-011-0065). It can be noticed that the film has a strong texture at CrN [220] orientation. This is consistent with the information presented by many authors [19-22] for the combination of parameters used in the deposition of this coating, since the texture of the films is affected by the Vbias and nitrogen partial pressure used in the deposition process [19-21]. The Cr peaks have shifted to the left indicating that the crystalline structure has compression microstresses that help to counteract subsurface tangential stresses during erosion preventing pitting and chipping of the coating. Microstresses alter the lattice spacing and broadening the diffraction peak [23]. It does not present evidence of Cr peaks.

Fig. 6a shows the surface microstructure of the coating, in which some surface defects like pinholes and droplets,

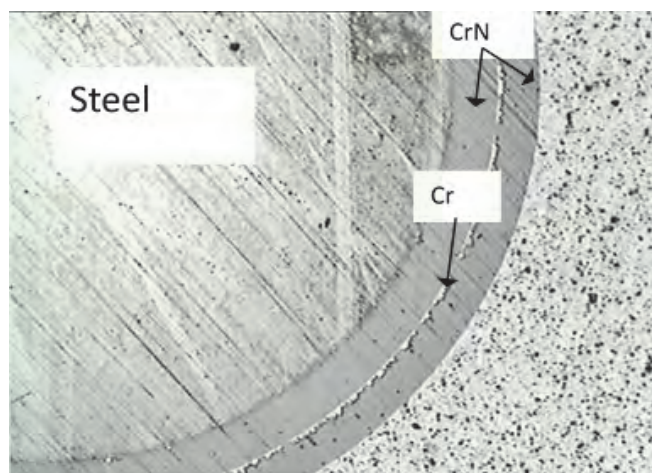


Figure 4 Optical micrograph at 50X of ball crater thickness test for Cr/CrN bilayer coating.

Source: The authors.

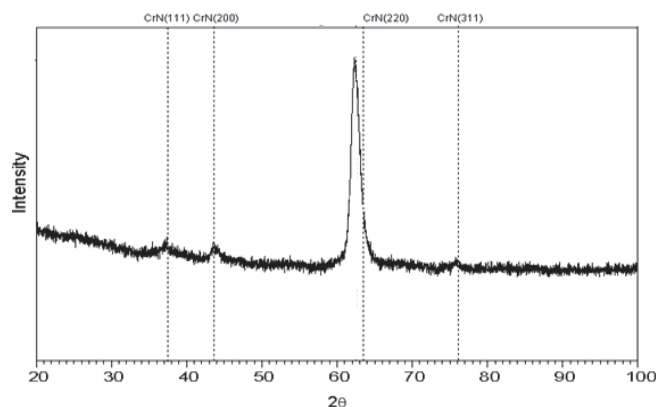


Figure 5 Diffractogram of Cr/CrN coating on AISI 440C steel.

Source: The authors.

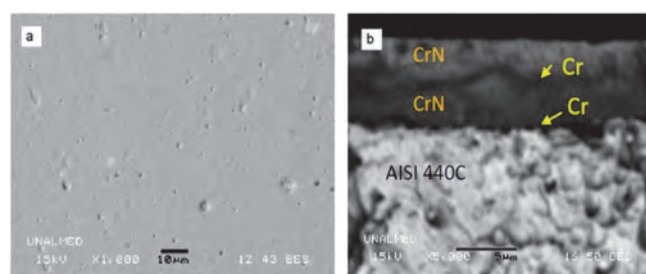


Figure 6 SEM micrographs of Cr/CrN coating. a) Surface. b) Transversal.

Source: The authors.

characteristic of process deposition, are seen. The micrographs were performed with SEM JEOL JSM 5910 LV at UNal laboratories.

Fig. 6b is a transversal micrograph of the coating, and it was obtained by previously fracturing a sample immersed in liquid nitrogen and etching the multilayer coating with a solution of HNO₃ and HF in the ratio 1:9 for 30 min. This preferentially attacks the Cr, and so can display the stratification layers.

3.3. Electrochemical measurements

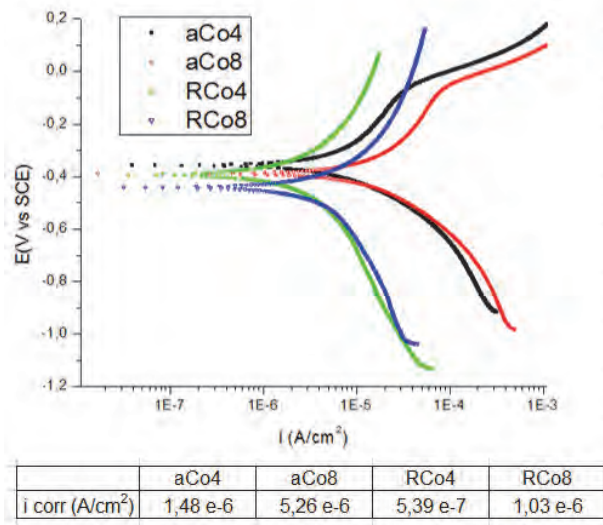
The polarization curves for pure corrosion (C_0) and erosion-corrosion (C_w) carried out on 440C steel and on a Cr/CrN bilayer coating in function of the impact velocity are shown in Fig. 7.

For all conditions of wear, the velocity increase causes a higher corrosion current density. Although the corrosion potential of coated samples are very close to bare samples, coating protects the surface under conditions of pure corrosion as well as EC, as is observed in Fig. 7, since corrosion current density is lower for coated samples.

3.4. Mass losses

Table 2 shows the individual contributions of erosion, corrosion and synergism for each of the evaluated conditions obtained, following the procedure described in ASTM G-119-09 [2]. Generally, increasing the velocity produces a

(a)



(b)

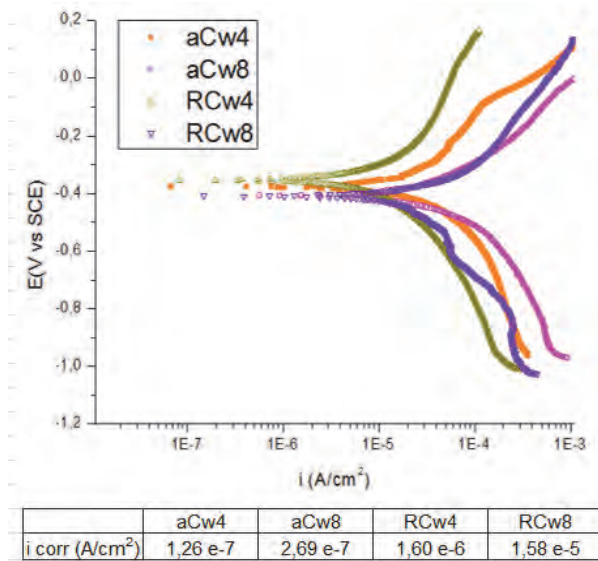


Figure 7. Potentiodynamic polarization curves made at different velocities on the 440C steel and the coating of Cr/CrN in a) pure corrosion b) erosion-corrosion.

Source: The authors.

higher mass loss rate, which is in accordance with the theories of erosion. Also, it should be noted that there is a greater loss of mass for both materials in erosion-corrosion conditions, the case of martensitic stainless steel being more drastic.

The pure corrosion mechanism (C_0) has little effect on the coating and on stainless steel under any of the test conditions. The pure erosion mechanism (W_0) was the greatest wear mechanism, mass loss being higher for uncoated steel. The protective effect of the coating for pure erosion decreases with increasing impact velocity. Moreover, the synergism effect in erosion-corrosion is remarkable for all other conditions, and even more so for

the uncoated samples (TML). Under the conditions tested and with the values reported in Table 2, Fig. 8 shows the percentage contribution of mass loss; for both the CrN/Cr coating as well as the 440C steel the synergism exceeds 40% of the total mass loss, which means that the mechanisms of interaction between erosion and corrosion are very important for the wearing down of both materials. For all conditions, the coating had a higher resistance, represented by a mass loss rate of about half compared with that obtained for the bare 440C steel. Fig. 9 shows the SEM micrographs, which confirm that the Cr/CrN bilayer coating exposed to pure erosion undergoes less material detachment at low velocity, without any observable EDS peaks from the substrate. For the eroded surfaces under EC, Fig. 9c and d, the material detachment is greater and a large and deep scar area in the center of the samples can be seen. Debris were observed scattered around the scar coating. In Fig. 9d, for high velocity EC it can be observed that there is a degradation of the substrate, which shows plastic deformation with the formation of valleys generated by the impact of the particles as observed in metals like aluminum. The direction range of valleys comes from the interaction between the particles in the jet due to their high concentration, which generates changes in the trajectories, which is reinforced by the particle size distribution in the jet [24]. The EDS of Fig. 9d was taken in the central zone of the sample, it shows Fe peaks and due to the height of the Fe/Cr peaks is close to that typical of bare steel, it seems that the substrate surface has been uncovered. Therefore, it can be confirmed that the fracture toughness and hardness of the coating is critical for erosion resistance, since the coating has a mode of failure by brittle fracture. For all cases, the synergistic effect (ΔW_c) is positive, being of a higher value for bare steel. This situation suggests that the surface does not produce corrosion products that help protect the surface [2-4, 25-26], which is confirmed in the potentiodynamic polarization curves of corrosive flux particles, which are shown in Fig. 6. It can be seen that there is no characteristic curve with anodic formation of passive layers. The coating was also unable to produce corrosion products that protect the surface, but the ΔW_c data are less positive, which provides protection twice that of 440C stainless steel.

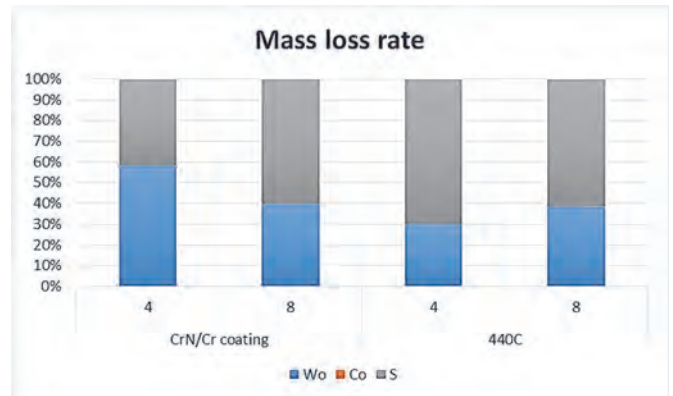


Figure 8 Percentage contribution of pure erosion, pure corrosion and synergism over total mass loss rate by erosion-corrosion wear.

Source: The authors.

Table 2.

Mass loss rate in the EC test. TML: Total material loss rate; W_0 : mechanical wear rate; C_0 : electrochemical corrosion rate; C_w : electrochemical corrosion rate during the corrosive wear process; S: Synergism; ΔW_c : Synergistic effect; ΔC_w : Additive effect.

Sample material	CrN/Cr coating		440C	
Impact velocity (m/s)	4	8	4	8
TML	9,3728	15,6213	28,4530	36,9889
W_0	5,4674	6,2485	8,5359	14,2265
C_0	6,8910E-06	1,3214E-05	1,7179E-05	6,1240E-05
S	3,9053	9,3727	19,9171	22,7623
ΔC_w	1,3513E-05	1,8903E-04	1,0905E-04	2,0800E-04
ΔW_c	3,9053	9,3726	19,9170	22,7621

Source: The authors.

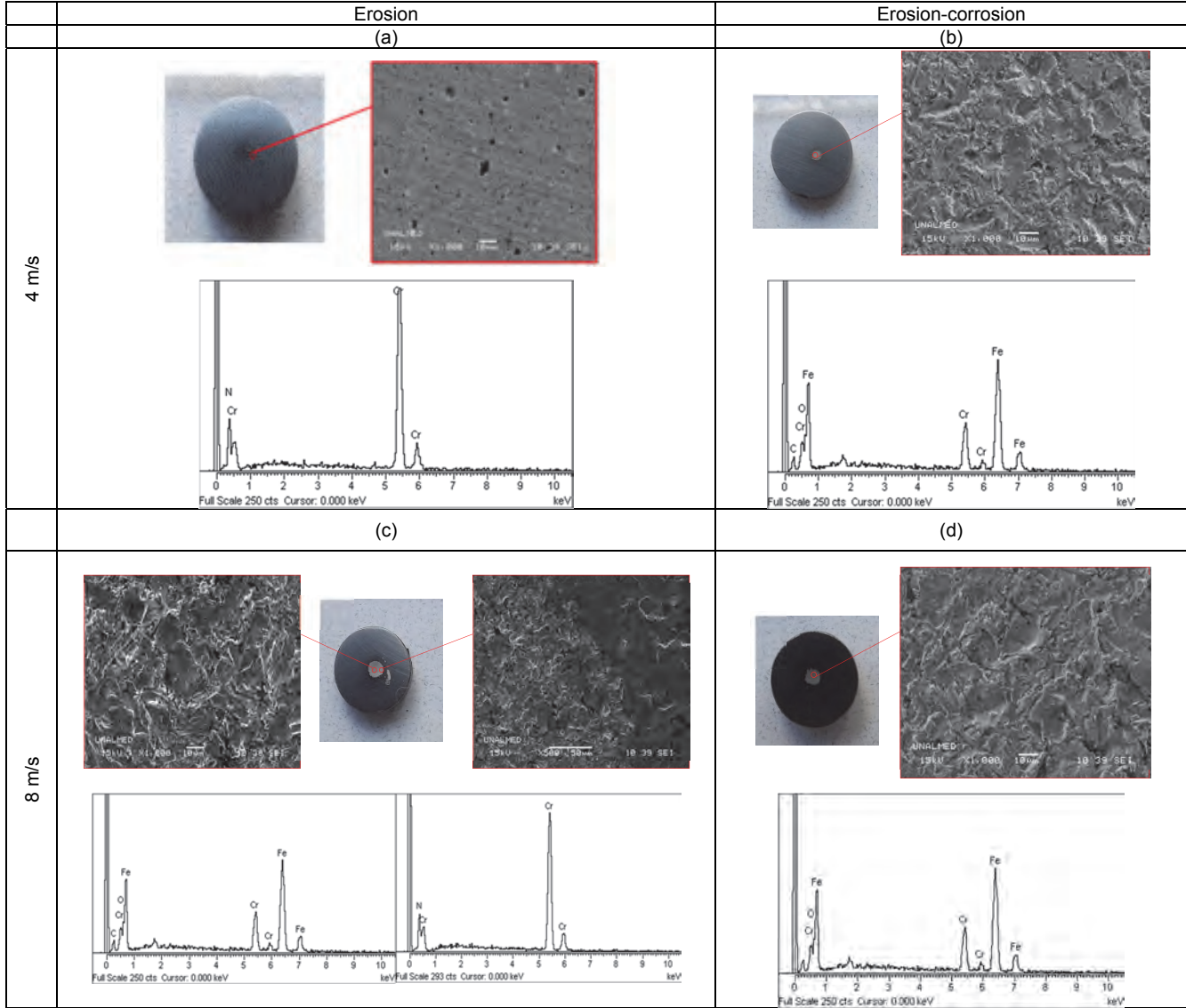


Figure 9. SEM micrographs of coatings after pure erosion and erosion-corrosion tests.

Source: The authors.

Different authors [5-6,11-13], who have studied some alloys and coatings, reported that synergism is an important component of the degradation in EC. Synergism in the case of alloys that undergo passivation is linked to the breakdown of the passive film by the action of the particles. For the case

of multiple layers, as in the case of ceramic coatings, the EC mechanisms and the synergism may be more complex, due to the presence of galvanic between Cr and Fe.

4. Conclusions

The bilayer coating of Cr/CrN offered resistance to erosion, corrosion and erosion-corrosion under the impact of 4 to 8 m/s compared to the bare surface.

The corrosion potential of potentiodynamic polarization curves is very similar to the bare surface, which can be attributed to the electrolyte penetrating the porosity of the coating.

Under the action of jet impact with slurry at 8 m/s, the coating fractured and detached from the substrate. Hence it is recommended that the adhesion of the coating be increased in order to prevent detachment, and also to heat treat the coating after processing in order to achieve a better toughness.

The pure corrosion rate compared to pure erosion is negligible, due to the high solid content (10% wt.) in the case of erosion, which generates a greater amount of impact, causing greater wear on surfaces through the mechanical effect.

At higher impact velocity, mass loss increased in terms of erosion, flow corrosion and EC for both materials (440C steel, bare and with coating).

The EC wear for all conditions tested was higher than the sum of the rates of pure erosion and pure corrosion, except for the case of steel subjected to erosion at 4 m/s.

It was possible to establish that the synergism exceeds 40% of the mass loss for all cases tested, and therefore is an important component of low mass loss under EC wear.

The 440C steel surface showed a mass loss rate of 304% under an impact of 4 m/s and 237% under an impact of 8 m/s compared to the coating mass loss rates.

Acknowledgments

The authors would like to thank Hugo Estupiñan, Ph.D., for his assistance in the electrochemical measurements. We are grateful to the Argentine Atomic Energy Commission (CNEA) and the National University of Colombia (UNal) for the use of the equipment and the laboratories, at CNEA (Coatings and Tribology Laboratory, metallography, X-ray diffraction) and at UNal (metallography laboratories and testing of materials, advanced microscopy). Thanks go to Dr. Norma Mingolo (CNEA), Mery Arrubla, Martin Griffiths (CNEA) and Manuel Iribarren (CNEA), for their help in conducting the tests, and to COLCIENCIAS and the Directorate of Research, National University of Colombia-Medellín (DIME) for the financial support given through the notice of Opening for Young Researchers No. 510 (2010).

References

- [1] Mattox, D.M., Handbook of physical vapor deposition (PVD) processing: film formation Adhesion, Surface Preparation and Contamination Control (Materials Science and Process Technology). USA: Noyes Publication, 1998. DOI: 10.1016/B978-081551422-0.50010-3
- [2] ASTM G119-09, Standard guide for determining synergism between wear and corrosion. ASTM International, West Conshohocken, PA, 2003, DOI: 10.1520/G0119-09. DOI: 10.1520/G0119-09
- [3] Stack, M.M., Lekatos, S. and Stott, F.H., Erosion-corrosion regimes: Number, nomenclature and justification? *Tribology International*, 28, pp. 445-451, 1995. DOI: 10.1016/0301-679X(95)00009-S
- [4] Stack, M.M., Zhou, S. and Newman, R.C., Identification of transitions in erosion-corrosion regimes in aqueous environments. *Wear*, 186-187, pp. 523-532, 1995. DOI: 10.1016/0043-1648(95)07175-X
- [5] Stack, M.M. and Wang, W.H., The slurry erosive wear of physically vapour deposited TiN and CrN coatings under controlled corrosion. *Tribology Letters*, 6, pp. 23-36, 1999. DOI: 10.1023/A:1019186918225
- [6] Stack, M.M. and Wang, H.W., Simplifying the erosion-corrosion mechanism map for erosion of thin coatings in aqueous slurries. *Wear*, 233, pp. 542-551, 1999. DOI: 10.1016/S0043-1648(99)00216-1
- [7] Kot, M., Rakowski, W.A., Major, L., Major, R. and Morgiel, J., Effect of bilayer period on properties of Cr/CrN multilayer coatings produced by laser ablation. *Surface & Coatings Technology*, 202, pp. 3501-3506, 2008. DOI: 10.1016/j.surfcoat.2007.12.036
- [8] Kaciulis, S., Mezzi, A., Montesperelli, G., Lamastra, F., Rapone, M., Casadei, F., Valente, T. and Gusmano, G., Multi-technique study of corrosion resistant CrN/Cr/CrN and CrN:C coatings. *Surface & Coatings Technology*, 201, pp. 313-319, 2006. DOI: 10.1016/j.surfcoat.2005.11.128
- [9] Marulanda, D.M., Olaya, J.J., Piratoba, U., Mariño, A. and Camps, E., The effect of bilayer period and degree of unbalancing on magnetron sputtered Cr/CrN nano-multilayer wear and corrosion. *Thin Solid Films*, 519, pp. 1886-1893, 2011. DOI: 10.1016/j.tsf.2010.10.010
- [10] Aperador, W., Ramirez, C. and Caicedo, J.C., The effect of Ti(CN)/TiNb(CN) coating on erosion-corrosion resistance. *Ingeniería e Investigación*, 32(2), pp. 6-11, 2012.
- [11] Purandare, Y., Stack, M.M. and Hovsepian, P., Velocity effects of erosion-corrosion of CrN/Nb superlattice PVD coatings. *Surface & Coatings Technology*, 201, pp. 361-370, 2006. DOI: 10.1016/j.surfcoat.2005.11.143
- [12] López, D., Sanchez, C. and Toro, A., Corrosion-Erosion behavior of TiN-coated stainless steels in aqueous slurries. *Wear*, 258, pp. 684-692, 2005. DOI: 10.1016/j.wear.2004.09.015
- [13] López, D., Congote, J.P., Cano, J.R., Toro, A. and Tschiptschin, A.P., Effect of particle velocity and impact angle on the corrosion-erosion of AISI 304 and AISI 420 stainless steels. *Wear*, 259, pp. 118-12, 2005. DOI: 10.1016/j.wear.2005.02.032
- [14] Barlow, L., The effect of austenitising and tempering parameters on the microstructure and hardness of martensitic stainless steel AISI 420. MSc. Thesis. University of Pretoria. Pretoria, South Africa, 2009.
- [15] Das, D., Dutta, A.K. and Ray, K.K., Influence of varied cryotreatment on the wear behavior of AISI D2 steel. *Wear*, 266, pp. 297-309, 2009. DOI: 10.1016/j.wear.2008.07.001
- [16] Jha, A.K., Swathi-Kiranmayee, M., Ramesh-Narayanan, P., Sreekumar, K. and Sinha, P.P. Metallurgical analysis of ball bearing seized during operation. *Journal of Materials Engineering and Performance*, 21, pp. 1076-1084, 2012.
- [17] Vander-Voort, G.F., Manilova, E.P., Michael, J.R. and Lucas, G.M., Study of selective etching of carbides in steel. *www.dgm.de. Buehler Ltd.* [Online]. [date of reference September 16th of 2013]. Available at: <www.dgm.de/past/2004/metallographie/download/686-60.pdf>
- [18] ASM Handbook., Vol. 4. Heat Treatment. USA, 1991.
- [19] Grant, W.K., Loomis, C., Moore, J.J., Olson, D.L., Mishra, B. and Perry, A.J., Characterization of hard chromium nitride coatings deposited by cathodic arc vapor deposition. *Surface and Coatings Technology*, 86-87, pp. 788-796, 1996. DOI: 10.1016/S0257-8972(96)03071-X
- [20] Oden, M., Ericsson, C., Håkansson, G. and Ljungcrantz, H., Microstructure and mechanical behavior of arc-evaporated Cr-N coatings. *Surface and Coatings Technology*, 114, pp. 39-51, 1999. DOI: 10.1016/S0257-8972(99)00019-5
- [21] Ahn, S.H., Choi, Y.S., Kim, J.G. and Han, J.G., A study on corrosion resistance characteristics of PVD Cr-N coated steels by electrochemical method. *Surface and Coatings Technology*, 150, pp. 319-326, 2002. DOI: 10.1016/S0257-8972(01)01529-8

- [22] Chipatecua-Godoy Y.L., Marulanda-Cardona D.M. and Olaya-Flórez J.J., Evaluating the corrosion resistance of UBM-deposited Cr/CrN multilayers. *Ingeniería e Investigación*, 31(1), pp. 16-25, 2011.
- [23] Noyan, C. and Cohen, J.B., *Residual stresses - Measurement by diffraction and Interpretation*. USA: Springer-Verlag, 1987.
- [24] Efird, K.D., *Jet impingement testing for flow accelerated corrosion*. Corrosion 2000. NACE International, 2000.
- [25] Chang, Z.K., Wan, X.S., Pei, Z.L., Gong, J. and Sun, C., Microstructure and mechanical properties of CrN coating deposited by arc ion plating on Ti6Al4V substrate. *Surface & Coatings Technology*, 205, pp. 4690-4696, 2011. DOI: 10.1016/j.surfcoat.2011.04.037
- [26] Warcholinski, B. and Gilewicz, A., Tribological properties of CrNx coatings. *Journal of Achievements in Materials and Manufacturing Engineering*. 37, pp. 498-504, 2009.

J. Alegria-Ortega, received her BSc. in Mechanical Engineering in 2009 and her MSc. degree in Materials and Manufacture Process in 2012, both from the Universidad Nacional de Colombia. Medellin, Colombia. From 2008 to 2011, she worked as quality management system coordinator in a Materials Laboratory of Universidad Nacional de Colombia. Medellin, Colombia. Currently, she is working as a project engineer in a consulting engineering company.
ORCID: 0000-0002-2934-5470

L.M. Ocampo-Carmona, received her BSc. in Chemical Engineering in 1994, her MSc degree in Metallurgical and Materials Engineering in 2000, and a Dr. degree in Metallurgical and Materials Engineering in 2005, all from the Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil. She worked as a teacher in chemistry, fluid mechanics, and strength of materials in the Universidad de Antioquia from 1996 to 1998. She is a full professor since 2006 in Department of Materials and Minerals, Facultad de Minas, Universidad Nacional de Colombia, Medellin, Colombia. She teaches chemistry metallurgy, physical metallurgy, metallurgical electrochemistry, electrometallurgy and corrosion, and materials science at undergraduate level and Fundamentals of corrosion at the Masters and Doctorate level. Her research interests include Corrosion and protection of metals, recycling of materials, bio-materials, and electrochemistry.
ORCID: 0000-0002-8117-1391.

F.M. Rodríguez, is Technician in 1978. He started working for the National Atomic Energy Commission, Argentina in 1984 and was involved in Vacuum technology projects and in Plasma coatings in 1997. He obtained a plasma materials process specialization in Nagoya University, Japan in 1997. He has been involved in the Arc-PVD process for more than 15 years.
ORCID: 0000-0001-9952-2420

E. Forlerer, received her BSc in Physical Science in 1976 and a PhD. in 1994 in Physics, from the Universidad Nacional de Buenos Aires, Argentina. She taught materials science at the Instituto Sabato of the Universidad Nacional de San Martin between 1994-2004 at the MSc. degree level. Since 2001, she has been Master's thesis and Doctoral advisor in materials science. She is currently Head of the Coatings and Tribology Laboratory at the Argentine National Atomic Energy Commission. Currently, her research is focused on thin film coatings as barriers for corrosion, wear, thermal transference, and diffusion.
ORCID: 0000-0003-3309-4544.



UNIVERSIDAD NACIONAL DE COLOMBIA

SEDE MEDELLÍN

FACULTAD DE MINAS

Área Curricular de Ingeniería
Geológica e Ingeniería de Minas y Metalurgia

Oferta de Posgrados

Especialización en Materiales y Procesos
Maestría en Ingeniería - Materiales y Procesos
Maestría en Ingeniería - Recursos Minerales
Doctorado en Ingeniería - Ciencia y Tecnología de
Materiales

Mayor información:

E-mail: acgeomina_med@unal.edu.co

Teléfono: (57-4) 425 53 68

Validation of an inertial sensor-based platform to acquire kinematic information for human joint angle estimation

Mauro Callejas-Cuervo ^{ab}, Andrés Felipe Ruíz-Olaya ^c & Rafael M. Gutiérrez ^b

^a Facultad de Ingeniería, Universidad Pedagógica y Tecnológica de Colombia, Tunja, Colombia. mauro.callejas@uptc.edu.co

^b Doctorado en Ciencia Aplicada, Universidad Antonio Nariño, Bogotá, Colombia. Rafael.gutierrez@uan.edu.co

^c Facultad de Ingeniería Electrónica y Biomédica, Universidad Antonio Nariño, Bogotá, Colombia. andresru@uan.edu.co

Received: May 11th, 2015. Received in revised form: January 17th, 2016. Accepted: May 5th, 2016.

Abstract

This paper shows the results of a set of experiments aimed at calibrating and validating an inertial sensor-based motion capture system that is used to capture and analyze the elbow joint flexion/extension motion. An experimental platform was constructed that provides accurate angular position information for reference purposes. Results obtained have an average error of 2.14 degrees when the arm is guided by a servomotor that rotates at 1 radian per second. The system also has an RMSE of 3.3, 4.9, 6.4 and 7.7 degrees for speeds of 2, 3, 4 and 5 radians per second respectively. Results show that the errors are acceptable to use a kinematic information capture platform with inertial sensors that is focused on monitoring the recovery of the motor function in the upper limbs through physical therapy.

Keywords: Motion capture, biomechanical signals, inertial sensors, potentiometer, upper limb, telerehabilitation.

Validación de una plataforma basada en sensores inerciales para adquirir información cinemática de estimación del ángulo articular humano

Resumen

El artículo presenta los resultados de una serie de experimentos para calibrar y validar un sistema de captura de información cinemática asociado al movimiento de flexión/extensión de la articulación del codo, con sensores inerciales, para lo cual se construyó una plataforma con un potenciómetro de precisión usado como referencia de posición angular. Los resultados muestran que los datos generados por el sistema presentan un promedio de error de 2.14 grados cuando el brazo guiado por el servomotor funciona a una velocidad de 1 radian por segundo. El sistema también tiene un RMSE de 3,3, 4,9, 6,4 y 7,7 grados para velocidades de 2, 3, 4 y 5 radianes por segundo, respectivamente. Los resultados muestran que los errores son aceptables para utilizar una plataforma de captura de información cinemática con sensores inerciales, orientada a controlar la recuperación de la función motora en los miembros superiores través de la terapia física.

Palabras clave: captura de movimiento, señales biomecánicas, sensores inerciales, potenciómetro, extremidad superior, telerehabilitación.

1. Introduction

The motion capturing system to measure flexion/extension, pronation/supination and abduction/adduction through the use of inertial sensors [1-2] that adhere to body segments is a broad field of research. It is found in topics such as: motion analysis [3-5]; active

videogame development [6-7]; high performance athlete analysis [8-10]; physical rehabilitation on human beings [11-13]; among others. This shows how important it is to analyze locomotive and manipulative behavior as well as to identify strategies for their improvement. In spite of technological advances there are a restricted number of systems that are able to make accurate, quantitative and low-cost calculations of the

How to cite: Callejas-Cuervo, M., Ruíz-Olaya, A.F. and Gutiérrez, R.M., Validation of an inertial sensor-based platform to acquire kinematic information for human joint angle estimation. DYNA 83(197), pp. 154-159, 2016.

angle value between the different body segments or joints.

This work shows the outcomes from a validation process of a motion capture system that is based on inertial sensors. The system allows the angle of any upper limb joint to be estimated. The validation experiments have been guided in order to obtain the flexion/extension angle from the elbow.

To calibrate the motion capture system, an experimental platform has been implemented that provides reference information on the angle between two mechanical segments: one fixed segment and one mobile segment. A precision potentiometer provides information on the reference angle. The experiments were made by mounting the inertial sensors onto segments in the experimental platform. The angle estimate is carried out by merging information from inertial sensors (a 3-axis accelerometer and a 3-axis gyroscope). The angle has been validated for the following different angular velocities: 1, 2, 3, 4 and 5 rad/s.

The structure of this article is the following: section 2 includes an introduction to orientation kinematics. Section 3 presents the experimental methods that include the construction of a test platform, programming and adjusting of the data capture software, reference system calibration, and data acquirement and outcomes analysis. Section 4 presents the results, and finally the conclusions are presented.

2. Orientation kinematics

Within a joint angle estimation context, orientation kinematics relates to the calculation of the relative orientation of a body relative to a global coordinate system. It is useful to attach a coordinate system to the body frame as well as another one to the reference when both, the reference and the body frames, have the same fixed origin. Fig. 1 presents a system with two coordinate systems: the body frame (Oxyz) and the reference frame (OXYZ).

Taking X,Y, Z as a reference coordinate system, and x, y, z as a body coordinate system, the rotation matrices are given by eq. (1) - (4):

$$\text{Rotation } \phi \text{ about X-axis, } R(X, \phi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & -\sin\phi \\ 0 & \sin\phi & \cos\phi \end{bmatrix} \quad (1)$$

$$\text{Rotation } \theta \text{ about Y-axis, } R(Y, \theta) = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix} \quad (2)$$

$$\text{Rotation } \psi \text{ about Z-axis, } R(Z, \psi) = \begin{bmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3)$$

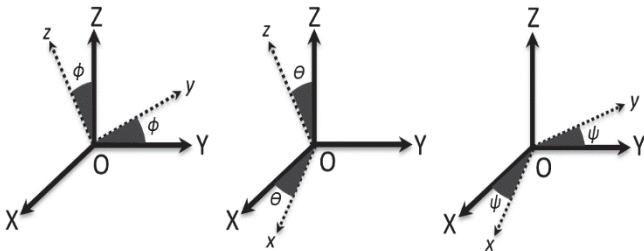


Figure 1. Coordinate transformation between the reference frame and the body.

Source: The authors.

The total rotation matrix, in the order ZYX, could be written as M_t , which means turning the body coordinates with respect to the reference coordinates [14]:

$$M_t = R(Z, \psi) R(Y, \theta) R(X, \phi)$$

$$M_t = \begin{bmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & -\sin\phi \\ 0 & \sin\phi & \cos\phi \end{bmatrix}$$

$$M_t = \begin{bmatrix} \cos\psi\cos\theta & -\sin\psi\cos\theta + \cos\psi\sin\theta\sin\phi & \sin\psi\sin\theta + \cos\psi\sin\theta\cos\phi \\ \sin\psi\cos\theta & \cos\psi\cos\theta + \sin\psi\sin\theta\sin\phi & -\cos\psi\sin\theta + \sin\psi\sin\theta\cos\phi \\ -\sin\theta & \cos\theta\sin\phi & \cos\theta\cos\phi \end{bmatrix} \quad (4)$$

3. Experimental methods

The following subsections are the stages used to develop the calibration processes as well as the validation of the inertial platform for human motion capture and analysis.

3.1. Experimental platform construction

The tests platform is a reliable and steady environment that allows data capture instead of reference. This platform is composed of an articulate segment together with a servomotor that provides movement and a high precision potentiometer, such as the one in Fig. 2.

Moreover, the platform integrates a control and processing unit, in this case, a microcontroller within an Arduino's board. The chosen sensor for application is an MPU-9150 from Invensense [15], which has 9 degrees of freedom and combines a three axis gyroscope, a three axis accelerometer and a three axis magnetometer; a Xbee wireless communication system that connects Arduino with the graphic interface from the developed system, which can be accessed from different devices (PC, Tablet, Smartphone and others); a 6v-2600mAh battery as a power supply. The MPU9150 sensor provides digital information through an I2C communication. Finally, the connection of the wireless communication modules is made through an Xbee shield for Arduino and a RF Xbee module.

3.2. Software/firmware for sensors data capture

After a test environment, it is necessary to program the data acquirement software. In this stage, some computational algorithms are created, proved and adjusted, and these are

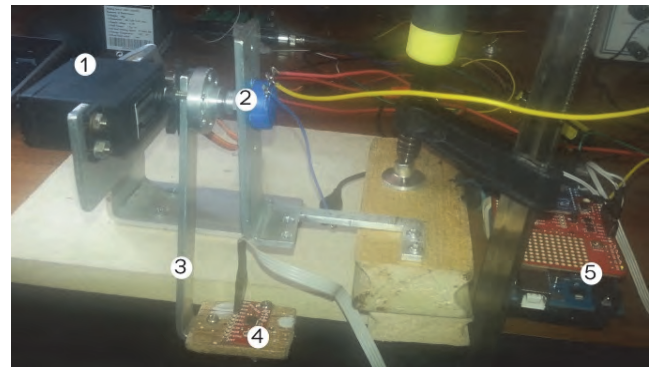


Figure 2. Test platform. ① Servomotor, ② Potentiometer, ③ Mechanical arm, ④ Sensor, ⑤ Arduino platform.

Source: The authors.

implemented on the capture system, the purpose of which is to collect, process, and distribute data from the whole motion capture system.

The firmware within the Arduino's board microcontroller is responsible for generating the communication link between sensors and users or data processing software, as well as containing the necessary functions to carry out the corresponding reading, setup, and processing of the information from the inertial sensors that are connected to it. Inside the developed software, there are several libraries containing basic functions that collect information from sensors in addition to mathematical algorithms that process the collected data. The following libraries are structured as follows [16]:

MPU9150 Library: This part of the firmware contains the functions necessary to execute the fusion of the inertial sensor's nine axis, through data obtention, the correlation of magnitudes, their variables and the implementation of calculation algorithms that deliver as results values in different units of measurement (quaternions, total and axis Euler angle, differential angles and raw values from the sensor).

Calibration Library: This contains the code related to the magnetometer and accelerometer calibration (the gyroscope does not require calibration). It also disposes of specialized algorithms for necessary specific data collection during the calibration, for example the auto select of the most relevant and accurate information that sensor can give.

Motion Driver: This contains selected and modified files from Embedded SDK v5.1 provided by InvenSense, and it is necessary to manage the sensor hardware.

I2CDev: This contains communication utilities between hardware and software through I2C protocol. It was originally created by Jeff Rowberg with an open source license. It has been slightly modified for its use in this application.

DueFlash: EEPROM library for AVR Arduinos. In this case, it is used to permanently write the outcomes of the calibration within EEPROM memory into the Arduino's board microcontroller.

Thus, there are tools to suitably capture and visualize data to later undertake the respective calculation and analysis.

3.3. Experimental platform calibration

The reference system used is based on a precision potentiometer, which is adjusted with manual tools that provide the angle information for different angular positions.

Maximum and minimum values of the potentiometer are read in order to find out the range of operation and correlate this with the information provided by a protractor. These values are computed by magnitude in voltage to correspond to an image in degrees. This allows a direct comparison between the potentiometer and the sensor information.

Several data collections of the mechanical segment motion were made in order of establish the correlation between voltage value and the angular displacement by the potentiometer. The outcome of the comparison with angles (180°, 135°, 90°, 45° and 0°) can be seen in Table 1.

Table 1.

The potentiometer calibration process

Angular displacement (degrees)	Potentiometer output (volts)
180	1.06
135	1.98
90	2.91
45	3.90
0	4.76

Source: The authors.

3.4. Data acquisition

In this stage, hardware and software are merged to collect necessary information from the experiments by taking into account different factors such as synchronization and reliability from the data recorded on different devices in the tests.

The conditioned potentiometer signal provides an analog voltage that is proportional to the angular displacement received through mechanical arm.

Information provided by the inertial sensor and magnetometer, relates to the transformation process of signals from acceleration, angular velocity and the magnetic field.

Information provided by the angular sensor relates to the transformation process of signals from the inertial sensor (acceleration, angular velocity and magnetic field). This angular information (Euler angles) was computed from equations (1) to (3) and corresponds to elbow joint flexion-extension. Further movements can be calculated from the rotation matrix in equation (4).

4. Results and Discussion

The results are carried out in four steps that involve the statistical analysis of the data, the linearization equation, data collection graphics generation, and quantitative analysis of variables.

4.1. Data statistical analysis

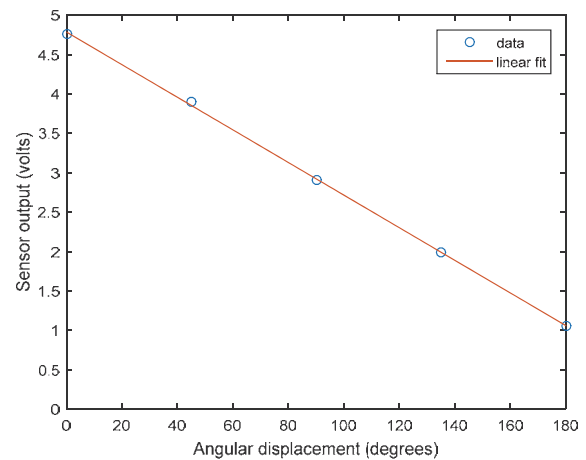


Figure 3. Data correlation and function linearization.

Source: The authors.

From the data acquired, batches of information were processed in a specialized software (OriginLab Demo) [17], thus, its intercept, percentage deviation, pendant, statistical correlation and median were calculated. This information allowed for an arithmetic linearization that corrects bugs from all the processed values by means of averaging. For the purpose of complementing the analysis, data correlation and function linearization were graphed, thus it was corroborated that the method used is correct, and the system modeling to a linear equation was simplified, as can be seen in Fig. 3.

4.2. Equation linearization

By using linear regression, the best-fitting straight line through the points was found. The equation for the line in Fig. 3 is:

$$Y = 4.78 - 0.02X \quad (5)$$

Equation (5) linearly correlates to the value of the potentiometer and shows the value in degrees, which represents the reference.

4.3. Graphical outcomes

Once data batches collected in test platforms were converted and adjusted through the mathematical modeling described above, their graphics were created in order to ease the theoretical analysis and provide the most relevant information about tests.

With an angular velocity of 1 rad/s, a high performance of the sensor can be seen in terms of the potentiometer. This is the system reference, as can be seen in Fig. 4.

A 1 rad/s velocity is of interest for this research as the upper-limb motion of a human body is usually placed within this velocity range. However, some tests were made with velocities of 2, 3 and 4 rad/s (Fig. 5), which presented similar behavior to the motion tracing but with an increase in the bug by the inertial sensor due to drift bias from the gyroscope [18-19] and white noise added by magnetic materials to the magnetometer signals [20].

Finally, the behavior of the system was analyzed with a higher velocity of 5 rad/s (see Fig. 6) supposing an abrupt movement infrequent in human upper limb biomechanics. However, from a technical point of view, it is interesting to analyze this in order to determine the effective response of the system, even in extreme cases or when some kind of abuse in its correct use can be presented.

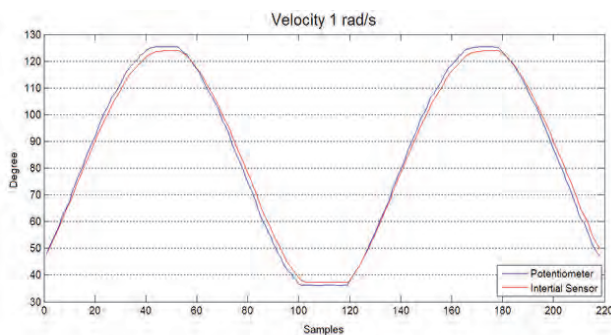


Figure 4. Lowest angular velocity – 1 rad/s.
Source: The authors.

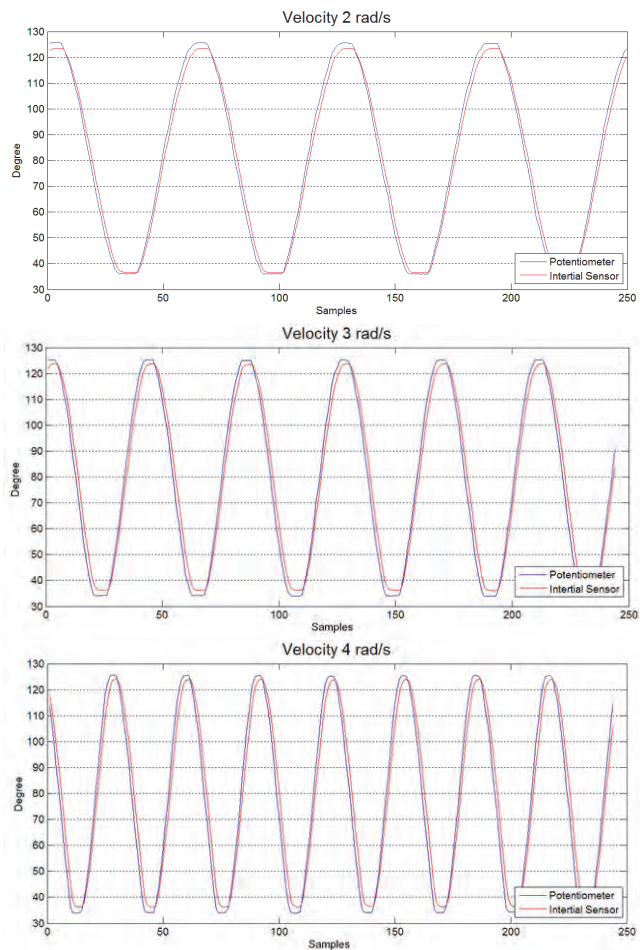


Figure 5. Behavior with angular velocities of 2, 3 and 4 rad/s.
Source: The authors.

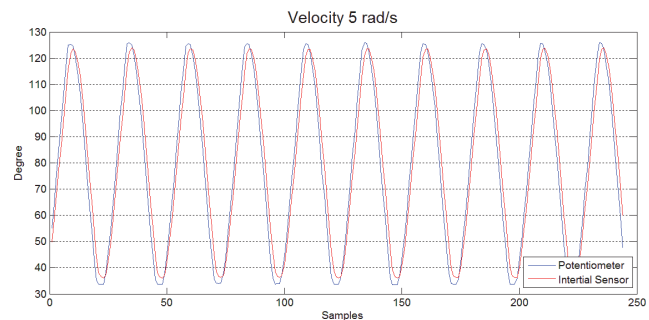


Figure 6. Highest angular velocity - 5 rad/s.
Source: The authors.

Furthermore, graphic and mathematical information can be merged in real time to determine other system behaviors, such as parasitic noises, deviations in measurements or even sensor calibration according to the variables imposed by the environment.

4.4. Quantitative analysis of variables

The root mean square error (RMSE) was used to quantify the difference of the value of the angles between the reference information (potentiometer) and the estimate using the inertial sensor. This can be seen in eq. (6) and (7).

Table 2
Outcomes

Standard	Velocity				
	1 rad/s	2 rad/s	3 rad/s	4 rad/s	5 rad/s
Number of samples	218	353	244	244	244
Potentiometer maximum value	125.6°	125.6°	125.2°	125.6°	125.9°
Potentiometer minimum value	36°	35.5°	33.9°	33.9°	33.5°
Sensor maximum value	123.9°	123.6°	123.7°	124.2°	123.9°
Sensor minimum value	37.3°	36.3°	36.12°	36.2°	36°
RMSE	2.14°	3.3°	4.9°	6.4°	7.7°

Source: The authors.

$$e(n) = \theta_{potentiometer}(n) - \theta_{sensor}(n) \quad (6)$$

$$RMSE = \sqrt{\frac{1}{k} \sum_{n=1}^k e^2(n)} \quad (7)$$

where k represents the number of samples taken with both systems and n , represents a sample in a time slot.

Table 2 summarizes the collected data using the tests platform that allowed the information and the performance in both systems in different circumstances to be characterized.

5. Conclusions

According to the results obtained in the tests, and taking into account their exactness in terms of the reference system, it becomes possible to identify that the system provides high efficiency in normal movements (from 1 and 2 rad/s), showing stability and reliability. However, projecting the implementation of a filter or estimator is necessary as it allows for the drift problem of the inertial sensors to be controlled, as well as the removal of white noise added to the signal during capture.

The results obtained through calibration and validation are sufficiently accurate, in terms of the data generated by an MPU-9150 sensor, to continue developing the capture and biomechanical data analysis platform that is used for telerehabilitation purposes. This is proposed in [5]. Also, the technology used is ergonomically enhanced, accurate, and has an easy handling option for any technical medical user.

References

- [1] Shaeffer, D.K., MEMS Inertial Sensors: A tutorial overview. IEEE Commun. Mag., 51(54), pp. 100-109, 2013. DOI: 10.1109/MCOM.2013.6495768.
- [2] Yazdi, N., Ayazi, F. and Najafi, Micromachined inertial sensors. Proceedings of the IEEE. 86(8), pp. 1640-1659, 1998. DOI: 10.1109/5.704269.
- [3] Choudhury, T.T., Rahman, M.M., Khorshidtalab, A. and Khan, M.R., Modeling of human arm movement: A study on daily movement. Fifth Int. Conf. Comput. Intell. Model. Simul, 2013. pp. 63-68. DOI: 10.1109/CIMSim.2013.19.
- [4] Tao, G., Huang, Z., Sun, Y., Yao, S. and Wu, J., Biomechanical model-based multi-sensor motion estimation. IEEE Sensors Appl. Symp. Proc., 2013. pp. 156-166. DOI: 10.1109/SAS.2013.6493577.
- [5] Callejas-Cuervo, M., Ruiz-Olaya, A.F. and Gutierrez, R., Biomechanical motion capture methods focused on tele-

- physiotherapy. Pan American Health Care Exchanges, 2013. pp. 1-6. DOI: 10.1109/SAS.2013.6493577.
- [6] Chiang, I.-T., Tsai, J.-C. and Chen, S.-T., Using Xbox 360 Kinect games on enhancing visual performance skills on institutionalized older adults with wheelchairs. IEEE Fourth Int. Conf. Digit. Game Intell. Toy Enhanc. Learn., 2012. pp. 263-267. DOI: 10.1109/DIGTEL.2012.69.
- [7] Lange, B., Koenig, S., McConnell, E., Chang, C.-Y., Juang, R., Suma, E., Bolas, M. and Rizzo, A., Interactive game-based rehabilitation using the Microsoft Kinect. IEEE Virtual Real, 2012. pp. 171-172. DOI: 10.1109/VR.2012.6180935.
- [8] Coyte, J.L., Stirling, D., Ros, M. and Gray, A., Displacement profile estimation using low cost inertial motion sensors with applications to sporting and rehabilitation exercises. IEEE/ASME Int. Conf. Adv. Intell. Mechatronics, 2013. pp. 1290-1295. DOI: 10.1109/AIM.2013.6584272.
- [9] Charry, E., Hu, W., Umer, M., Ronchi, A. and Taylor, S., Study on estimation of peak ground reaction forces using tibial accelerations in running. IEEE Eighth Int. Conf. Intell. Sensors, Sens. Networks Inf. Process., 2013. pp. 288-293. DOI: 10.1109/ISSNIP.2013.6529804.
- [10] Jung, P.-G., Lim, G. and Kong, K., A mobile motion capture system based on inertial sensors and smart shoes. IEEE Int. Conf. Robot. Autom., 2013. pp. 692-697. DOI: 10.1109/ICRA.2013.6630648.
- [11] Zhang, S., Naghdy, F., Stirling, D. and Gray, A., Ankle injury assessment using inertial 3D data. IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM), 2013. pp. 810-815. DOI: 10.1109/AIM.2013.6584193.
- [12] Bruckner, H.-P., Nowosielski, R., Kluge, H. and Blume, H., Mobile and wireless inertial sensor platform for motion capturing in stroke rehabilitation sessions. 5th IEEE Int. Work. Adv. Sensors Interfaces IWASI, 2013. pp. 14-19. DOI: 10.1109/IWASI.2013.6576085.
- [13] Zhang, Z.-Q., Ji, L.-Y., Huang, Z.-P. and Wu, J.-K., Adaptive information fusion for human upper limb movement estimation. IEEE Trans. Syst. Man, Cybern. - Part A Syst. Humans, 42(5), pp. 1100-1108, 2012. DOI: 10.1109/TSMCA.2012.2189876.
- [14] Álvarez, D., Álvarez, J.C., González, R.C. and López, A.M., Upper limb joint angle measurement in occupational health, Comput. Methods Biomech. Biomed. Engin. pp. 1-12, 2015. DOI: 10.1080/10255842.2014.997718.
- [15] Invensense. Technology Motion, [online]. [Date of reference: June 2015]. Available at: <http://www.invensense.com/technology/motion/>
- [16] Arduino. Libraries of programming, [online]. [Date of reference: June 2015]. Available at: <http://playground.arduino.cc/Main/MPU-9150>
- [17] OriginaLab. Software Demo, [online]. [Date of reference: June 2015]. Available at: <http://www.originlab.com/demodownload.aspx>
- [18] Olivares, A., Olivares, G., Mula, F., Górriz, J.M. and Ramírez, J., Wagyromag: Wireless sensor network for monitoring and processing human body movement in healthcare applications. J. Syst. Archit., 57(10), pp. 905-915, 2011. DOI: 10.1016/j.sysarc.2011.04.001.
- [19] Vaccaro, R. and Zaki, A., Statistical modeling of rate gyros. IEEE Trans. Instrum. Meas., 61(3), pp. 673-684, 2012. doi: 10.1109/TIM.2011.2171609.
- [20] Butta, M. and Sasada, I., Sources of noise in a magnetometer based on orthogonal fluxgate. IEEE Trans. Magn., 48(4), pp. 1508-1511, 2012. DOI: 10.1109/TMAG.2011.2173177.

M.Callejas-Cuervo, received his BSc. in System Engineering in 1997 and is currently undertaking his PhD in Applied Science, both from the Antonio Nariño University, in Bogotá, Colombia. He is also a PhD student in Energy and Process Control at Oviedo University, Spain. He received an MSc in Computer Science from the Institute of Technology in Monterrey, Mexico. He is a full-time professor in the Faculty of Engineering at the Pedagogical and Technological University of Colombia, and a member of the Software Research Group. His research interests include motion capture with inertial sensor projects in telerehabilitation and development of video game actives, as well as software engineering and business intelligence.
ORCID: 0000-0001-9894-8737

A.F. Ruiz-Olaya, received his BSc. in Electronic Engineering from the University of Valle, Cali, Colombia and his PhD from Carlos III University of Madrid, Spain. He is a full-time professor in the Faculty of Electronics and Biomedical Engineering at the Antonio Nariño University, Bogotá,

Colombia, institution in which he is a member of the Bioengineering Research Group. His research interests include robotics rehabilitation, multimodal human-robot interaction, bioinspired control architectures and modeling of human motor control. Currently, he is working on several projects oriented towards the development of neurorehabilitation systems to aid individuals with motor disabilities. Prof. Ruiz-Olaya is a member of IEEE. He has been author and co-author of multiple papers and has worked as reviewer of several important international journals and conferences.
ORCID: 0000-0002-5883-5786

R.M. Gutierrez, received his BSc in Physics from the Universidad Nacional de Colombia, his MSc in Condensed Matter from Montreal University, Mexico, and his PhD in Applied Sciences from New York University, USA. He is Director of an Applied Science PhD Program and Head of the Complex System Research Group at the Antonio Nariño University in Bogotá, Colombia. He has also been vice director of the Scientific and Technology Development Program at the Colombian Institute for the Development of Science And Technology "Francisco José De Caldas" – Colciencias, March 2006 - March 2007 and Director of Basic Sciences on a National Program at the Colombian Institute for the Development of Science And Technology "Francisco José De Caldas" – Colciencias, October 2004 - March 2006. He is author of more than 130 indexed publications in various interdisciplinary subjects.
ORCID: 0000-0001-5734-6174



UNIVERSIDAD NACIONAL DE COLOMBIA

SEDE MEDELLÍN
FACULTAD DE MINAS

Área Curricular de Ingeniería
Eléctrica e Ingeniería de Control

Oferta de Posgrados

Maestría en Ingeniería - Ingeniería Eléctrica

Mayor información:

E-mail: ingelcontro_med@unal.edu.co
Teléfono: (57-4) 425 52 64

Study of the evolution of the residual stress state in thermal barrier coatings sprayed on AISI 304 stainless steel

Pedro Yáñez-Contreras ^a, José Dolores Oscar Barceinas-Sánchez ^b, Carlos Agustín Poblano-Salas ^c,
José Martín Medina-Flores ^d, Adrián Luis García-García ^b & Iván Domínguez-López ^b

^a Universidad Politécnica de Guanajuato, Guanajuato, México. pyanez@upgto.edu.mx

^b Centro de Investigación en Ciencia Aplicada y Tecnología Avanzada (CICATA-IPN), Querétaro, México. obarceinas@ipn.mx, agarcia@ipn.mx, idominguezl@ipn.mx

^c Centro de Tecnología Avanzada (CIATEQ A.C.), Querétaro, México. carlos.poblano@ciateq.mx

^d Departamento de Ingeniería Mecánica, Instituto Tecnológico de Celaya (ITC), Guanajuato, México. martin.medina@itcelaya.edu.mx

Received: June 9th, 2015. Received in revised form: January 27th, 2016. Accepted: February 11th, 2016

Abstract

Thermal barrier coatings (TBC) were manufactured by thermal spraying of CoNiCrAlY and YSZ powders, which constituted the bond coat (BC) and top coat (TC), respectively. The BC was deposited on stainless steel substrates by a high velocity oxygen fuel (HVOF) system whilst the TC was deposited on the BC by an atmospheric plasma spraying (APS) gun. TBCs were heat treated at 1200 °C for 8 and 16 h in order to determine the evolution of the residual stress state as a function of time. The residual stress profiles were obtained by employing the Modified Layer Removal Method (MLRM) and an adaptation of this for bilayer coatings called Modified Layer Removal Method for Duplex Coatings (MLRMD). It was observed that as the exposure time was increased, the compressive residual stresses increased, which was more noticeable in the BCs. The TBCs without heat treatment revealed tensile residual stresses.

Keywords: thermal barrier coating, residual stress, modified layer removal method, thermal spraying, Modified Layer Removal Method for Duplex Coatings.

Estudio de la evolución del perfil de esfuerzos residuales en recubrimientos barrera térmica depositados sobre acero inoxidable AISI 304

Resumen

Se fabricaron recubrimientos barrera térmica depositando polvos de CoNiCrAlY y YSZ, los cuales constituyeron las capas de enlace y superior, respectivamente. La primera se depositó sobre el acero inoxidable AISI 304 mediante un sistema de oxígeno combustible de alta velocidad; mientras que la segunda se depositó mediante un sistema de rociado por plasma atmosférico. Los recubrimientos se trataron térmicamente a 1200 °C por 8 y 16 h con el fin de determinar la evolución del estado de esfuerzos residuales en función del tiempo. Los perfiles de esfuerzos residuales se determinaron empleando el método de remoción de capa modificada y una adaptación de éste para recubrimientos bicapa, llamado método de remoción de capa modificada para recubrimientos dúplex. Se encontró que conforme el tiempo de exposición se incrementó los esfuerzos residuales de compresión incrementaron, siendo más notorios en las capas de enlace. Los recubrimientos sin tratamiento térmico revelaron esfuerzos residuales de tensión.

Palabras clave: recubrimiento barrera térmica, esfuerzos residuales, método de remoción de capa modificada, rociado térmico, método de remoción de capa modificada para recubrimientos dúplex.

1. Introduction

Thermal barrier coatings (TBC) are commonly employed in hot sections of aeronautic and land turbines to provide protection against hot corrosion and oxidation of critical components [1,2].

A TBC is formed by two layers of different materials deposited on a metallic substrate. The external layer, also known as the top coat (TC), is commonly formed by yttria stabilized zirconia (YSZ), a ceramic with low thermal conductivity, good thermal-shock resistance and a relatively low thermal expansion coefficient [3-

How to cite: Yáñez-Contreras, P., Barceinas-Sánchez, J.D.O., Poblano-Salas, C.A., Medina-Flores, J.M., García-García, A.L. and Domínguez-López, I., Study of the evolution of the residual stress state in thermal barrier coatings sprayed on AISI 304 stainless steel. DYNA, 83(197), pp. 160-166, 2016.

5]. On the other hand, the layer in contact with the substrate, known as the bond coat, is normally fabricated from a MCrAlY alloy ($M = \text{Co, Ni, or Co+Ni}$). Such a layer promotes good adherence of the TC to the substrate, as it has a thermal expansion coefficient similar to that of the substrate [7]. It also provides protection against oxidation [5,8] and hot corrosion [2,8] by forming a stable alumina thermally grown oxide (TGO) at the interface with the TC. TBCs can be processed entirely by thermal spraying; the BC is normally produced by employing the HVOF process, whereas the ceramic TC is fabricated by high temperature processes, such as atmospheric plasma spraying (APS). Both processes are cost-effective and produce good quality TBCs[9].

Residual stresses are always present in thermal spray coatings. They are produced by two known mechanisms: i) quenching of splats after impact the substrate and restricted contraction of them by adherence to the substrate; and ii) differential thermal expansion of the coating and the substrate during cooling. The stress state of a coating is important as it has a strong influence on its mechanical properties, which, in turn, affect its performance and durability [10, 11]. According to Widjaja *et al.* [12], the final stress state of a TBC is influenced not only by quenching of splats and thermal mismatch but also by phase transformation during coating deposition.

In the HVOF process the particles projected onto the substrate reach high velocities, and temperatures lower than those measured in the APS process, which in contrast produces a low velocity and high temperature plume, adequate for the deposition of ceramic powders. Despite those differences, it has been suggested that the coatings processed by each process to make a TBC, i.e. the BC and TC, respectively, have a compressive residual stress state [13], which opposes to the tensile stresses that are responsible for coating delamination during service [14].

Another important variable when processing a TBC by thermal spraying is substrate roughness. A substrate having high roughness promotes a compressive residual stress state in the BC, which improves the cracking resistance of the TBC [15]. Therefore, it is important to promote a compressive stress state in a TBC as a tensile stress state increases the coating susceptibility to delamination and cracking [16].

During the last decade, different techniques have been proposed to study the final stress state of TBCs that are produced by thermal spraying: x-ray diffraction [6, 16], indentation [17], finite element simulation [18, 20], material removal [21,22], hole drilling [23], 3D imaging correlation [24], and curvature evolution [17]. In this work, the residual stress state was evaluated by employing the Modified Layer Removal Method (MLRM) [25]. The method is based on the monitoring of strains in the coating after removing thin layers by fine grinding. This technique was developed for mono-layer systems; however, the method has also been employed in duplex systems but only considering the physical properties of one material in the calculations [21]. Here, the MLRM technique was modified in order to calculate the residual stress state of a TBC employing the physical properties of the BC and TC substrates. This modified procedure was named as the Modified Layer Removal Method for Duplex Coatings (MLRMDC). A comparison of the residual stress state of a TBC, with and without heat treatment at 1200 °C, obtained by employing the MLRM and MLRMDC methods, is presented in this work.

2. Materials and Methods

In this investigation, the MLRM and MLRMDC methods were used to calculate the residual stresses profiles of TBCs prior and after heat treatment at 1200 °C for 8 and 16 h. The following section describes each experimental stage.

2.1. Deposition of TBCs

Three TBCs were deposited on stainless steel AISI 304 substrates of 25.4 mm X 25.4 mm X 6.35 mm in size. The BC and TC were obtained by deposition of powders of a CoNiCrAlY alloy (AMDRY 9954, Co32Ni21Cr8Al0.5Y), and yttria stabilized zirconia (NS-204, YSZ), respectively: both from Sulzer Metco. Before deposition, the substrates were cleaned with acetone and grit blasted by an air stream carrying alumina particles that impacted at a 45° angle at a distance of 200 mm, they were then re-cleaned with acetone; the surface roughness (R_a) ranged from 3 to 7 μm . The BC was deposited by a HVOF system, model DJH2700 by Sulzer, with a powder feed rate of 38 g/min and a distance of 203 mm. The mean thickness of this layer was $180 \pm 5 \mu\text{m}$. On the other hand, the TC was deposited by an APS system, model 9MB by Sulzer, with a powder feed rate of 45 g/min and a distance of 120 mm. The potential and current of the gun were set at 67 V and 600 A, respectively. The mean thickness of this layer was $120 \pm 3 \mu\text{m}$. Each layer was built-up applying 16 passes at a gun lateral displacement speed of 1.5 m/s. Table 1 shows the pressures and flow rates of the gases employed.

2.2. Heat treatment of coatings

Once the TBCs were deposited, two samples were heat treated at 1200 °C using a closed muffle (by Nabertherm) in argon atmosphere, one for 8 h and the other for 16 h, Table 2. The heating rate was 6 °C/min. The sample without heat treatment (WT) was used for comparison.

2.3. Fixing and soldering of strain gauges

After heat treatment, the back faces of the samples were carefully ground and cleaned to eliminate oxide and dirt, thus ensuring an adequate attaching of the strain gauges. These devices were supplied by Vishay Precision Group Inc.

Table 1.
Pressure and flow rate of the gases.

HVOF	Pressure (bar)	Flow rate (SLPM)
Oxygen	10	154
Propane	8.9	38
Air	6.8	294
APS	Pressure (bar)	Flow rate (SLPM)
Argon	5.2	42
Hydrogen	3.4	7

Source: The authors

Table 2.
Heat treatment of samples.

Sample	Temperature (°C)	Time (h)
8H	1200	8
16H	1200	16
WT	Without treatment	

Source: The authors

(model CEA-09-125UT-350). The gauges were joined to connecting wires by applying a tin-based solder. The connection verification and strain measurements were carried-out by means of a strain meter model P3500, also provided by the same supplier.

2.4. Residual stresses determination by MLRMD

The method followed for the removal of thin layers of the TBC coating is described elsewhere [21,25]. The physical properties of all materials involved were considered when determining the residual stresses; stainless steel AISI 304 for substrate, CoNiCrAlY for BC, and YSZ for TC, Table 3.

The mathematical analysis of the stresses was performed using only a quarter of the sample, as shown in Fig. 1. The force and momentum acting on the removed layer along and around the X axis are F_x and M_x , respectively. Thickness of the removed layer, remaining layer, substrate, BC, and TC are h , h' , H , h_{BC} y h_{TC} , respectively. The strain gauge was attached on samples' back face; b_x and b_y are half of both length and width of the actual sample. Plane located at $z = 0$ is the interface substrate/BC.

Strain changes of the remaining part of the sample, after removing a thin layer by fine grinding, are given by:

$$\varepsilon_x = \varepsilon_{x0} + k_{xz} \quad \text{and} \quad \varepsilon_y = \varepsilon_{y0} + k_{yz} \quad (1)$$

Where k_{xz} and k_{yz} are sample curvatures.

For a plain stress state, the stress-deformation equation is given by:

$$\begin{Bmatrix} \sigma_x \\ \sigma_y \end{Bmatrix} = E' \begin{bmatrix} 1 & \nu \\ \nu & 1 \end{bmatrix} \begin{Bmatrix} \varepsilon_x \\ \varepsilon_y \end{Bmatrix} \quad (2)$$

Where E' is a relationship that involves the Young modulus, and ν the Poisson ratio of the materials.

E_s is the Young modulus of the substrate, E_{BC} of the BC, and E_{TC} of the TC:

$$E_s = \frac{E_s}{(1-\nu_s)}, E'_{BC} = \frac{E_{BC}}{(1-\nu_{BC})}, \text{ and } E'_{TC} = \frac{E_{TC}}{(1-\nu_{TC})} \quad (3)$$

Where E_t is the overall Young modulus of the TBC:

$$E_t = E_{BC} + E_{TC} \quad (4)$$

If $\sum h \geq h_{TC}$, then $E'_{TC} = 0$, but if $\sum h \geq h_{TC} + h_{BC}$ then $E'_{TC} = E'_{BC} = 0$.

Table 3.
Physical properties of the TBC materials.

Property	CoNiCrAlY [26]	YSZ [10]	AISI 304 [27]
Young modulus (GPa)	200	70	200
Poisson ratio	0.30	0.23	0.29

Source: The authors

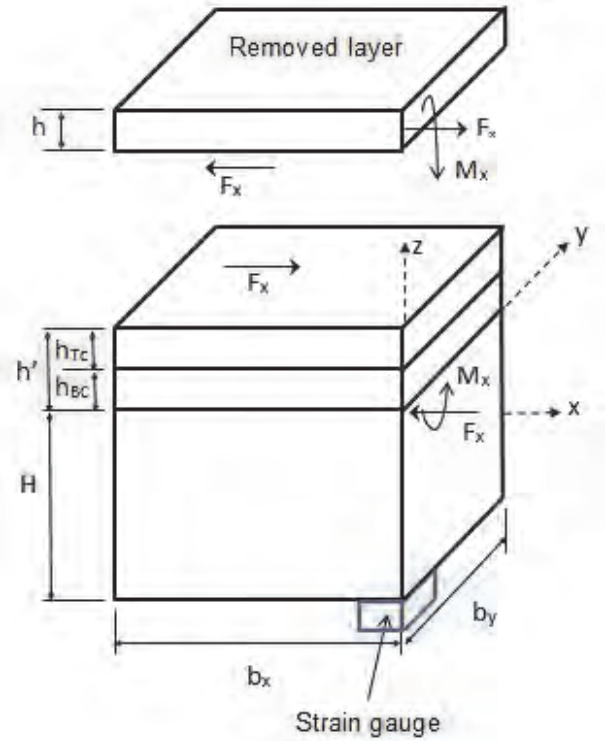


Figure 1. Schematic representation of the TBC indicating the removed layer.
Source: The authors

Where $\sum h$ is the sum of the thickness of removed layers.

The resulting forces and momenta along and around x and y axes are F'_x and F'_y , and M'_x and M'_y , respectively, which are related with the stresses by means of eq. 5:

$$\begin{Bmatrix} F'_x \\ F'_y \end{Bmatrix} = \int_{-H}^0 \begin{Bmatrix} \sigma_x \\ \sigma_y \end{Bmatrix} dz + \int_0^{h'} \begin{Bmatrix} \sigma_x \\ \sigma_y \end{Bmatrix} dz \quad \text{and} \quad \begin{Bmatrix} M'_x \\ M'_y \end{Bmatrix} = \int_{-H}^0 \begin{Bmatrix} \sigma_x \\ \sigma_y \end{Bmatrix} z dz + \int_0^{h'} \begin{Bmatrix} \sigma_x \\ \sigma_y \end{Bmatrix} z dz \quad (5)$$

Substituting equations (1) and (2) in (5) results:

$$\begin{Bmatrix} F'_x \\ F'_y \\ M'_x \\ M'_y \end{Bmatrix} = \begin{bmatrix} A_{11} & A_{12} & B_{11} & B_{12} \\ A_{12} & A_{22} & B_{12} & B_{22} \\ B_{11} & B_{12} & D_{11} & D_{12} \\ B_{12} & B_{22} & D_{12} & D_{22} \end{bmatrix} \begin{Bmatrix} \varepsilon_{x0} \\ \varepsilon_{y0} \\ k_x \\ k_y \end{Bmatrix} \quad (6)$$

Where:

$$\begin{aligned} A_{11} &= A_{22} = E_s H + E_t h'; \quad A_{12} = \nu_s E_s H + (\nu_{BC} + \nu_{TC}) E_t h' \\ B_{11} &= B_{22} = ((E_t - E_s) H h') / 2; \quad B_{12} = ((\nu_{BC} + \nu_{TC}) E_t h' - \nu_s E_s H h') / 2 \\ D_{11} &= D_{22} = ((E_s H / 12) * (H^2 + 3h'^2)) + ((E_t h' / 12) * (h'^2 + 3H^2)) \\ D_{12} &= ((\nu_s E_s H / 12) * (H^2 + 3h'^2)) + (((\nu_{BC} + \nu_{TC}) E_t h' / 12) * (h'^2 + 3H^2)) \end{aligned}$$

Using eq. (1) and the deformations measured by the strain gauges, equations for ε_{x0} and ε_{y0} can be derived:

$$\varepsilon_{x0} = \Delta\varepsilon_{xG} + k_x \left(\frac{h'-H}{2} \right) \quad \text{and} \quad \varepsilon_{y0} = \Delta\varepsilon_{yG} + k_y \left(\frac{h'-H}{2} \right) \quad (7)$$

Substituting eq. (7) in (6) to eliminate ε_{x0} and ε_{y0} ; thus eq. (8) is obtained:

$$\begin{Bmatrix} F'_x \\ F'_y \\ M'_x \\ M'_y \end{Bmatrix} = - \begin{Bmatrix} \sigma_{xL} h \\ \sigma_{yL} h \\ \sigma_{xL} h \left(\frac{h+h'-H}{2} \right) \\ \sigma_{yL} h \left(\frac{h+h'-H}{2} \right) \end{Bmatrix} \quad (8)$$

Where σ_{xL} and σ_{yL} are the stresses of the removed layer.

Equations (6) and (8) are substituted in (7) in order to obtain 4 equations with 4 variables, 2 stresses and 2 curvatures. This is undertaken in terms of deformation changes and sample properties. Thus, the following is obtained:

$$\begin{bmatrix} h & 0 & A_{11} \left(\frac{h'-H}{2} \right) + B_{11} & A_{12} \left(\frac{h'-H}{2} \right) + B_{12} \\ 0 & h & A_{12} \left(\frac{h'-H}{2} \right) + B_{12} & A_{22} \left(\frac{h'-H}{2} \right) + B_{22} \\ \left(\frac{h+h'-H}{2} \right) h & 0 & B_{11} \left(\frac{h'-H}{2} \right) + D_{11} & B_{12} \left(\frac{h'-H}{2} \right) + D_{12} \\ 0 & \left(\frac{h+h'-H}{2} \right) h & B_{12} \left(\frac{h'-H}{2} \right) + D_{12} & B_{22} \left(\frac{h'-H}{2} \right) + D_{22} \end{bmatrix} \begin{Bmatrix} \sigma_{xL} \\ \sigma_{yL} \\ k_x \\ k_y \end{Bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{12} & A_{22} \\ B_{11} & B_{12} \\ B_{12} & B_{22} \end{bmatrix} \begin{Bmatrix} \Delta\varepsilon_{xG} \\ \Delta\varepsilon_{yG} \end{Bmatrix} \quad (9)$$

Equation (9) shows the equation system to be solved for the stresses of the removed layer, σ_{xL} and σ_{yL} , and changes in the curvatures, k_x and k_y . Equation (7) is used to calculate ε_{x0} and ε_{y0} . Then, the residual stresses on the remaining part of the sample can be calculated by eq. (2).

3. Results and Discussion

Different residual stresses are developed during the fabrication of TBCs by thermal spraying. A compressive stress state is first developed on the substrate front face when its surface is prepared by grit blasting. As the BC is deposited, a tensile residual stress is developed by particle quenching and restricted contraction on the substrate [5], and finally, when the TC is deposited, a mixed stress state is developed due to the significant difference in thermal expansion coefficients between both layers. Compressive residual stresses are developed in the TC, whereas the BC develops tensile residual stresses [28].

The residual stress profiles that were calculated by the procedure described in section 2.4 are presented in Figs. 2 and 3. Fig. 2 shows the residual stress profiles, obtained by MLRMDC, for the samples that were heat treated at 1200 °C for 8 and 16 h, and for that without heat treatment. The plots show how the residual stress varies from the TC surface to the substrate.

For the sample without heat treatment (WT-DC), an almost zero residual stress was measured on the TC surface, which agrees with what was reported by Lima et al. [21] and Weyant et al. [6]. This stress state resulted from the balance between the compressive and tensile stresses in the coating. The former stress is promoted by the large difference of thermal expansion coefficients of YSZ and CoNiCrAlY, whereas the latter is produced by rapid quenching of flying particles when impacting the substrate and previously deposited layers. At an approximate depth of 250 μm , within the BC, a residual tensile stress of 20 MPa was measured. Such behavior has also previously been reported by other authors [21, 29]. In this case, a dominant effect of quenching stress in the bond coat was evident. As the measurements approach the substrate, the stress state tended, once again, to be zero, and the residual stress measured within the substrate was compressive. Although the results obtained by the MLRMDC method in this work showed a similar behavior to those reported in the literature employing the MLRM procedure [21], the magnitude of the measured stresses was different for each technique. This behavior could be due to the different processing conditions employed in each case (substrate, fuel and spraying distance).

The residual stress profile of the TBC that was heat-treated at 1200 °C for 8 h (8H-DC, Fig. 2) also showed a near zero stress at the TC surface. From 50 μm only compressive stresses were measured. The highest residual stress (-19 MPa) in the coating was located within the BC, whereas a further increase, up to -40 MPa, was measured in the substrate. The residual stress profile of the TBC that was heat-treated at 1200 °C for 16h (16H-DC) revealed a stress of -18 MPa on the TC surface, which increased as the depth was increased. The profile of this sample showed the same trend as that of the sample treated for 8 h, but the magnitude of the stresses was higher; this behavior has also been reported by Khan et al. [3]. These authors reported an increase in compressive residual stresses of a TBC as the exposure time at high temperature was increased. The TBC that was heat-treated for 16 h also showed a maximum compressive residual stress (-58 MPa) within the BC at a depth of 250 μm . The substrate showed a larger compressive residual stress of -78 MPa. A BC having a large residual compressive stress promotes good adherence of the TBC to the substrate; therefore, an increased adherence is expected in heat-treated samples compared to that of the as sprayed counterpart. Also, a TBC having a compressive residual stress state shows higher resistance to cracking and delamination in service [15]. The operating temperature of aeronautic and land turbines are higher than the testing temperature employed in this work (1200 °C). This testing temperature was selected as in a future investigation the residual stress profiles of TBCs, affected by chemical attack of calcium magnesium aluminosilicate (CMAS), will be calculated by MLRMDC. CMAS attack normally occurs between 1100 and 1200 °C on TBCs in aeronautic turbines exposed to volcanic flying ashes or SiO₂-based sand.

Fig.3 shows the evolution of residual stress calculated by MLRM and MLRMDC as a function of depth. A similar behavior of the evolution of residual stress was observed for both methods for the substrate and BC. However, the stress

magnitude for the TC was slightly different. Fig. 4 shows a magnified view of the residual stress profile calculated for the TC. The residual stresses on the surface of the samples without heat treatment and heat-treated for 8 h showed a difference smaller than 1 MPa, whereas for the sample heat-treated for 16 h the difference was about 2 MPa, which represents a dissimilarity of 15 % between both methods. The difference of the residual stress profiles, calculated by both methods for the TC, can be explained by the fact that the MLRMDC procedure takes into account the physical properties of the substrate, BC, and TC, whereas the MLRM only considers the physical properties of the substrate and TC.

This work should be regarded as an exploratory study aiming to provide an initial insight into the results obtained by MLRM and MLRMDC. This is why only one sample per condition was studied. A larger number of samples need to be studied in order to assess the error involved in each measurement and unambiguously determine if there is a statistically valid difference between both methods when calculating the residual stress profile through the TC. Nonetheless, it seems that the MLRM underestimates the magnitude of the stresses within the TC.

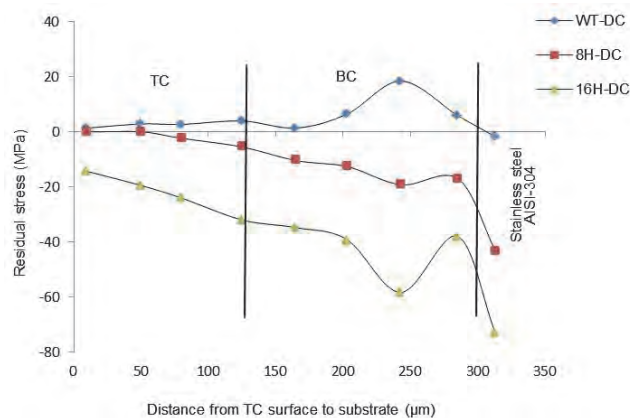


Figure 2. Evolution of residual stress, obtained by MLRMDC as a function of depth for samples with and without heat treatment.
Source: The authors

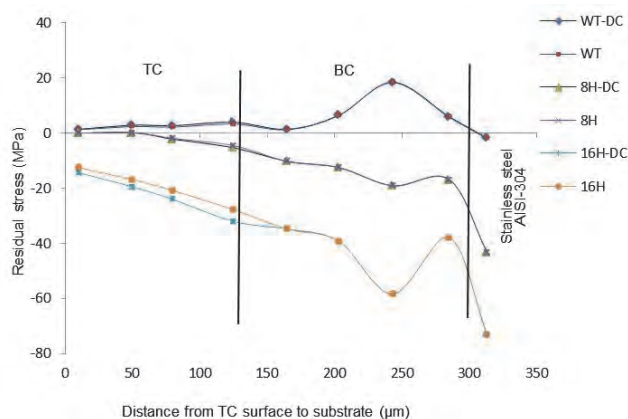


Figure 3. Residual stress profiles for the TBC obtained by MLRM and MLRMDC.
Source: The authors

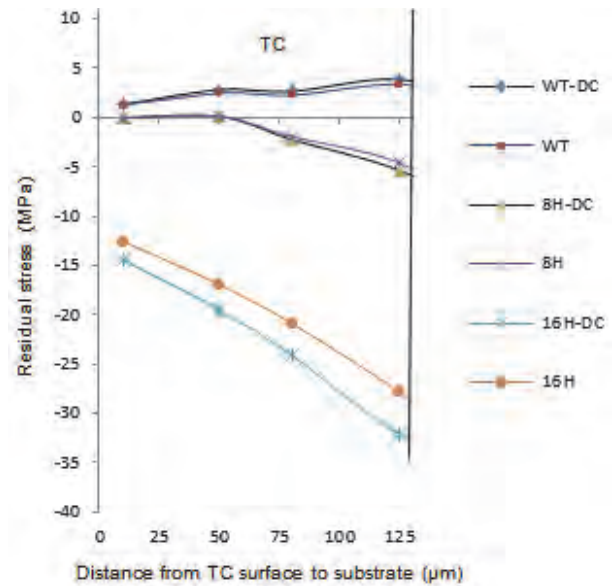


Figure 4. Comparison of the residual stress profiles in the TC calculated by MLRM and MLRMDC.
Source: The authors

4. Conclusions

The evolution of residual stress state in the TBCs (YSZ/CoNiCrAlY), fabricated by thermal spraying, was obtained by employing the modified layer removal method for duplex coatings (MLRMDC). This method resulted in residual stress profiles similar to those obtained by MLRM, which is a procedure regularly reported in the literature to determine the residual stresses of coatings manufactured by thermal spraying. The results obtained employing both methods are in good agreement with those reported in the literature. The effect of heat treatment at 1200 °C on the evolution of the residual stress profiles was also studied here. It was also demonstrated that an increase in soaking time at high temperature eliminates all tensile stresses in the TBCs, which results in compressive stresses after heat treatment. This, in turn, may enhance the adherence and delamination resistance of the coating in service. Until this point, it seems that the MLRM may be underestimating the magnitude of the compressive stresses within the TC.

Acknowledgements

The main author gratefully acknowledges the scholarship granted by PROMEP that was used to carry out his Doctoral studies. Also, thanks to CICATA-IPN Querétaro, CIATEQ, NACOBRE and Universidad Politécnica de Guanajuato for providing the facilities to perform the experimental work. Finally, the authors acknowledge the financial support of the Secretaría de Investigación y Posgrado from the Instituto Politécnico Nacional, through Project Grant SIP 20121606.

References

- [1] Loganathan, A. and Gandhi, A.S., Effect of phase transformations on the fracture toughness of γ' yttria stabilized zirconia. Materials

- Science & Engineering A 556 pp. 927-935, 2012. DOI: 10.1016/j.msea.2012.07.095
- [2] Osorio, J.D., Toro, A. and Hernández-Ortiz, J.P., Thermal barrier coatings for gas turbine applications: Failure mechanisms and key microstructural features. DYNA, [Online]. 79(176), pp. 149-158, 2012. Available at: <http://www.revistas.unal.edu.co/index.php/dyna/article/view/28328/43558>
- [3] Khan, A.N., Lu, J. and Liao, H., Effect of residual stresses on air plasma sprayed thermal barrier coatings. Surface and Coatings Technology 168, pp. 291-299, 2003. DOI: 10.1016/S0257-8972(03)00207-X.
- [4] Osorio, J.D., Lopera-Valle, A., Toro A. and Hernández-Ortiz, J.P. Phase transformations in air plasma-sprayed yttria-stabilized zirconia thermal barrier coatings. DYNA, [Online]. 81(185), pp. 13-18, 2014. Available at: <http://www.redalyc.org/articulo.oa?id=49631031002>
- [5] Rajendran, R., Gas turbine coatings – An overview. Engineering Failure Analysis, 26, pp. 355-369, 2012. DOI: 10.1016/J.ENGFAILANAL.2012.07.007
- [6] Weyant, C.M., Almer, J. and Faber, K.T., Through-thickness determination of phase composition and residual stresses in thermal barrier coatings using high-energy X-rays. Acta Materialia 58, pp. 943-951, 2010, DOI: 10.1016/j.actamat.2009.10.010
- [7] Hyung-Jun, J., Dong-Ho, P., Yeon-Gil, J., Jung-Chel, J., Sung-Churl, C. and Ungyu, P., Mechanical characterization and thermal behavior of HVOF-sprayed bond coat in thermal barrier coatings (TBCs). Surface & Coatings Technology 200 pp. 4355-4436, 2006, DOI: 10.1016/j.surfcoat.2005.02.170
- [8] Zhao, L. and Lugscheider, E., High velocity oxy-fuel spraying of a NiCoCrAlY and an intermetallic NiAl-TaCr alloy. Surface and Coatings Technology 149, pp. 231-235, 2002. DOI: 10.1016/S0257-8972(01)01444-X
- [9] Gómez-García, J., Poza, P. y Utrilla, V., Crecimiento y caracterización de recubrimientos cerámicos con aplicaciones como barreras térmicas. Bol. Soc. Esp. Cerámica y Vidrio 45(2), pp. 70-74, 2006.
- [10] Teixeira, V., Andritschky, M., Fischer, W., Buchkremer, H.P. and Stöver, D., Effects of deposition temperature and thermal cycling on residual stress state in zirconia-based thermal barrier coatings. Surface and Coatings Technology (120–121) pp. 103-111, 1999. DOI: 10.1016/S0257-8972(99)00341-2
- [11] Hashmi, M.S.J., Pappalettere, C. and Ventola, F., Residual stresses in structures coated by a high velocity oxy-fuel technique. Journal of Materials Processing Technology, 75 pp. 81-86, 1998. DOI: 10.1016/S0924-0136(97)00295-1
- [12] Widjaja, S., Limarga A.M. and Yip, T.H., Modeling of residual stresses in a plasma-sprayed zirconia y alumina functionally graded-thermal barrier coating. Thin Solid Films 434 pp. 216-227, 2003. DOI: 10.1016/S0040-6090(03)00427-9
- [13] Bansal, P., Shipway, P.H. and Leen, S.B., Residual stresses in high-velocity oxy-fuel thermally sprayed coatings – Modelling the effect of particle velocity and temperature during the spraying process. Acta Materialia 55, pp. 5089-5101, 2007. DOI: 10.1016/j.actamat.2007.05.031
- [14] Lima, R.C. and Guilemany, J.M., Adhesion improvements of thermal barrier coatings with HVOF thermally sprayed bond coats. Surface & Coatings Technology, 201 pp. 4694-4701, 2007. DOI: 10.1016/j.surfcoat.2006.10.005
- [15] Zhang, X., Watanabe, M. and Kuroda, S., Effects of residual stress on the mechanical properties of plasma-sprayed thermal barrier coatings. Engineering Fracture Mechanics, 110 pp. 314-327, 2013. DOI: 10.1016/j.engfractmech.2013.08.016
- [16] Totemeier, T.C. and Wright, J.K., Residual stress determination in thermally sprayed coatings a comparison of curvature models and X-ray techniques. Surface & Coatings Technology, 200 pp. 3955-3962, 2006. DOI: 10.1016/j.surfcoat.2005.06.003
- [17] Chen, X., Yan, J. and Karlsson, A.M., On the determination of residual stress and mechanical properties by indentation. Materials Science and Engineering. A 416, pp. 139-149, 2006. DOI: 10.1016/j.msea.2005.10.034
- [18] Ng, H.W. and Gan, Z., A finite element analysis technique for predicting as-sprayed residual stresses generated by the plasma spray coating process. Finite Elements in Analysis and Design, 41, pp. 1235-1254, 2005. DOI: 10.1016/j.finel.2005.02.002
- [19] Wang, L., Wang, Y., Sun, X.G., He, J.Q., Pan, Z.Y. and Wang, C.H., Finite element simulation of residual stress of double-ceramic-layer La2Zr2O7/8YSZ thermal barrier coatings using birth and death element technique. Computational Materials Science, 53, pp. 117-127, 2012. DOI: 10.1016/j.commatsci.2011.09.028
- [20] Jiang Y., Xu, B.S., Wang, H.D. and Lu, Y.H., Finite element modeling of residual stress around holes in the thermal barrier coatings. Computational Materials Science, 49, pp. 603-608, 2010. DOI: 10.1016/j.commatsci.2010.05.057
- [21] Lima, C.R.C., Nin, J. and Guilemany, J.M., Evaluation of residual stresses of thermal barrier coatings with HVOF thermally sprayed bond coats using the Modified Layer Removal Method (MLRM). Surface & Coatings Technology, 200, pp. 5963-5972, 2006. DOI: 10.1016/J.SURFCOAT.2005.09.016
- [22] Jiang, L.M., Peng, J., Liao, Y.G., Zhou, Y.C., Liang, J., Hao, H.X. and Lu, C., A modified layer-removal method for residual stress measurement in electrodeposited nickel films. Thin Solid Films, 519, pp. 3249-3253, 2011. DOI: 10.1016/j.tsf.2011.01.260
- [23] Santana, Y.Y., La Barbera-Sosa, J.G., Staia, M.H., Lesage, J., Puchi-Cabrera, E.S., Chicot, D. and Bemporad, E., Measurement of residual stress in thermal spray coatings by the incremental hole drilling method. Surface & Coatings Technology, 201, pp. 2092-2098, 2006. DOI: 10.1016/j.surfcoat.2006.04.056
- [24] Zhu, J., Xie, H., Hu, Z., Chen, P. and Zhang, Q., Residual stress in thermal spray coatings measured by curvature based on 3D digital image correlation technique. Surface & Coatings Technology, 206 pp. 1396-1402, 2011. DOI: 10.1016/J.SURFCOAT.2011.08.062
- [25] Rybicki, E.F., Shadley, J.R., McGrann, R.T.R., Savarimuthu, A.C. and Graving, D., An ASM recommended practice for Modified Layer Removal Method (MLRM) to evaluate residual stress in thermal spray coatings. [Online]. 2000. Available at: www.asminternational.org/.../dvd?...an-asm-recomm
- [26] Mao, W.G., Zhou, Y.C., Yang, L. and Yu, X.H., Modeling of residual stresses variation with thermal cycling in thermal barrier coatings. Mechanics of Materials, 38 pp. 1118-1127, 2006. DOI: 10.1016/j.mechmat.2006.01.002
- [27] Meza, J.M., Franco, E.E., Farias, M.C.M., Buiocchi, F., Souza, R.M. y Cruz, J., Medición del módulo de elasticidad en materiales de ingeniería utilizando la técnica de indentación instrumentada y de ultrasonido. Revista de Metalurgia [Online], 44(1), pp. 52-65, 2008. Available at: <http://revistademetalurgia.revistas.csic.es/index.php/revistademetalurgia/article/viewFile/95/94>
- [28] Zhang, X.C., Xu, B.S., Wang, H.D. and Wu, Y.X., Optimum designs for multi-layered film structures base on the knowledge on residual stresses. Applied Surface Science, 253, pp. 5529-5535, 2007. DOI: 10.1016/J.APSUSC.2006.12.076
- [29] Kesler, O., Matejcek, J., Sampath, S., Suresh, S., Gnaeupel-Herold, T., Brand, P.C. and Prask, H.J., Measurement of residual stress in plasma-sprayed metallic, ceramic and composite coatings. USA. Materials Science and Engineering, A257. pp. 215-224, 1998. DOI: 10.1016/S0921-5093(98)00860-0
- [30] Teixeira, V., Andritschky, M., Fischer, W., Buchkremer, H.P. and Stöver, D., Analysis of residual stresses in thermal barrier coatings. Journal of Materials Processing Technology, 92-93, pp. 209-216, 1999. DOI: 10.1016/S0924-0136(99)00157-0

P. Yáñez-Contreras, received his BSc. Eng. in Mechanical Engineering in 1996 from the Instituto Tecnológico de Celaya, Mexico. He completed his MSc in the same Technological Institute in 2002. He is studying a PhD at the Instituto Politécnico Nacional in Querétaro, México. From 2000 to 2003 he worked for CIATEQ, A.C. designing machinery and equipment. He is currently a Full-Professor at the Universidad Politécnica de Guanajuato, México (UPGTO). His main research activities are mechanical properties of Thermal Barrier coatings.
ORCID: 0000-0003-1124-6092

J.D.O. Barceinas-Sánchez, received his BSc. Eng. in Chemical Metallurgy in 1988 from the Universidad Autónoma de Querétaro, Mexico, his MSc in Metallurgy in 1992 from the Instituto Tecnológico de Saltillo, and his PhD

degree in Materials Science in 1998 from the University of Sheffield, UK. From 1998 to 2008, he worked for CIATEQ, A.C. and since 2008 for the Instituto Politécnico Nacional as a Research Fellow in the Tribology Laboratory. His research interests include: friction and lubrication analysis, erosive wear of materials, and deposition and characterization of coatings. ORCID: 0000-0003-2059-5016

C.A. Poblano-Salas, is a BSc. in Metallurgical Engineering who graduated from the National University of Mexico (UNAM), Mexico. He completed his MSc in the same University in 1997. He received the "Alfonso Caso" medal in 1998, awarded for obtaining the highest score achieved for MSc studies in Metallurgy. He obtained a PhD degree in Materials Engineering from the University of Sheffield, UK, in 2005. He is member of the National Research System (SNI), level I, and has been since 2009. Currently, he works for CIATEQ A.C. in the Materials Department. His main research activities are focused on the production and characterization of functional coatings processed by different thermal spray techniques such as atmospheric plasma spray and HVOF. He also has broad experience in the processing of alloys by powder metallurgy and in the assessment of electrochemical properties of different materials. ORCID: 0000-0001-9375-6832

J.M. Medina-Flores, has a Ph.D. degree from the Department of Mechanical Engineering of the University of Guanajuato, México. He worked as performance engineer for three years for CIAT-GE at the Queretaro facility in México. He worked as Professor-Researcher at the Universidad Politécnica de Guanajuato, Mexico, for six years in the Design of Heat Transfer Equipment for Optimal use of Energy. His Research interests include the Robust Design of Thermo-Mechanics Systems for the efficient use of energy. Currently, he is involved in research with the thermal systems group at the Mechanical Engineering Department of the Instituto Tecnológico de Celaya, Guanajuato, México. ORCID: 0000-0001-7393-9033

A.L. García-García, is graduated from Ohio University, USA, with a PhD. in 1994. Currently, he is Full-Professor at the Instituto Politécnico Nacional, Centro de Investigación en Ciencia Aplicada y Tecnología Avanzada (CICATA), Mexico, institution which he led from April 2001 through November 2006. Prior to his current position, he has held full-time professor positions at the Autonomous University of Morelos State, Mexico and the Optical Communications Research Institute, Mexico. He has also had a full-time Research Associate position at the North Carolina State University, USA, at Raleigh, in conjunction with the University of California, at Berkeley, and carried out materials research at the Lawrence Berkeley National Laboratory, developing X-ray microscopes for the Advanced Light Source synchrotron. ORCID: 0000-0002-5121-9032

I. Domínguez-López, received his PhD in Physics in 1994 the Universidad Nacional Autónoma de México (UNAM), Mexico. From 1994 to 1998, he worked for the UNAM at the Institute of Physics; he worked from 1997 to 1999 as a Postdoctoral fellow; for two years at the Lawrence Berkeley National Laboratory, USA; and for one year at the University of Nevada, USA. From 2000 to 2003 he worked as a scientific coordinator at the Centro Nacional de Metrología in México at the Time and Frequency Department, and collaborated in the Atomic Fountain project at the physikalisch-technische bundesanstalt (PTB), in Germany. Currently, he is a Full Professor in the Materials Processing and Manufacturing group at the Centro de Investigación en Ciencia Aplicada y Tecnología Avanzada in Querétaro, México. His research interests include: Applied laser light to monitor wear processes and surface texturing to improve tribological behavior of knee prosthesis. ORCID: 0000-0001-8084-8711



UNIVERSIDAD NACIONAL DE COLOMBIA

SEDE MEDELLÍN
FACULTAD DE MINAS

Área Curricular de Ingeniería Mecánica

Oferta de Posgrados

Maestría en Ingeniería - Ingeniería Mecánica

Mayor información:

E-mail: acmecanica_med@unal.edu.co
Teléfono: (57-4) 4259262

Burst erasure correction using LDPC codes constructed on base matrices generated by matched groups, nested polygons and superposed circulant matrices

Cassio André Sousa-da Silva ^a & Evaldo Pelaes ^b

^aFacultad de Matematica, Universidad Federal de Oeste de Pará, Santarém, Brasil. cassio@ufpa.br

^bDepartamento de Engenharia Elétrica, Universidade Federal de Pará, Belém, Brasil. pelaes@ufpa.br

Received: July 4th, 2015. Received in revised form: January 18th, 2016. Accepted: February 18th, 2016.

Abstract

This article proposes procedures for the construction of base matrices grounded in algebra and geometry. These base matrices serve as a platform to generate the parity check matrices for debugging in bursts erasure through LDPC codes by superposing the base matrices and movements of circulant matrices. The construction of the matrices is performed by concatenation as it is easy to implement and has a lower randomness. To demonstrate the potential of the technique, we developed a number of simulations using low complexity encoding as well as the sum-product algorithm. Several LDPC codes (matrices) were generated and the results were compared with other approaches. We also present the outcomes of erasure recovery simulations that result from the transmission of an image through a noisy channel.

Keywords: Low-density parity-check codes, burst erasure correcting codes, burst erasure channels, erasure-correcting codes, matrix by superposition, Information and Communications Technologies (ICT's).

Corrección de borrado en ráfaga utilizando códigos LDPC construidos sobre matrices generadas por grupos combinados, polígonos anidados y matrices circulantes superpuestas

Resumen

En este artículo son propuestos procedimientos para la construcción de matrices base embazado en el álgebra moderna y en la geometría. Estas matrices sirven de plataforma para generar las matrices de verificación de paridad en la corrección de borrado en ráfaga a través de códigos LDPC, por medio de superposición en las matrices base y movimientos de las matrices circulantes. La construcción de las matrices es realizada por concatenación, siendo de fácil implementación y de menor aleatoriedad. Para demostrar el potencial de la técnica, fue elaborado un conjunto de simulaciones que utiliza codificación de baja complejidad, bien como algoritmo suma y producto. Fueron generados varios códigos LDPC (matrices) y los resultados obtenidos comparados con otros abordajes. Son también presentados los resultados de la simulación de la recuperación de borrados resultantes de la transmisión de una imagen a través de un canal ruidoso.

Palabras clave: Códigos de baja densidad, Códigos de corrección de borrado en ráfaga, Canales de borrado de ráfaga, Códigos de La corrección de errores, matrices superpuestas, Tecnologías de la Información y Comunicaciones (TIC's)

1. Introduction

Currently, information technologies (ICTs) are active in all the social events that use media as a form of

communication. From large corporate business [1-3] to hydrographic ecosystems analysis [4], ICTs are used as a great way to communicate and present scientific conclusions. The huge explosion of the use of these electronic media is

How to cite: Sousa-da Silva, C.A. and Pelaes, E., Burst erasure correction using LDPC codes constructed on base matrices generated by matched groups, nested polygons and superposed circulant matrices. DYNA 83(197), pp. 167-174, 2016.

partly due to the popularity of HDTV [5] and the transmission of data via the internet. Over the last few years a new modality has been becoming more popular in these transmissions and contributes to a real communication: real-time transmissions. The goal of a communication system is to transmit a message through a communication channel so that the receiver can retrieve the message with a given criterion of fidelity. Nevertheless, a real communication system faces various problems, especially certain disturbances that are introduced by the media. These generate erasure during transmission across the channel due to several physical variables imposed by the channel [1].

Recent research suggests that on a variety of networks, packet losses occur in bursts [7-10]. As such, many bits lost during the transmission of messages through the channel cause long decoding delays and other message processing in the receiver. The application of LDPC codes to correct these lost and unrecognizable bits in burst has been widely used [6,11-12].

To detect unrecognizable symbols produced by deleting channels, block codes are typically used as error detection codes in memoryless channels. Recently, LDPC codes have been used for erasure-correcting memory [15-16]. Low-Density Parity-Check code (LDPC) is a linear block code $C(n, k)$ with a rate k/n that is defined by a sparse parity check matrix $H(n-k) \times n$. This uses the iterative algorithm for decoding known as belief propagation or sum-product (SP) [16] that is currently being used to correct erasure in bursts [17,11-14].

In this work, we consider a burst erasure channel (BuEC) in channels with memory, as is defined in [7]. The transmitter sends n binary symbols through the channel, but in the receiver, the received symbols r_i ($0 \leq i \leq n-1$) can be deleted. Since we assume an ideal detector for burst noise, these burst noises give rise to burst erasures. We can measure the effectiveness of a given burst noise code by a single parameter: the maximum resolvable erasure-burst length, L_{max} . This parameter is defined as the maximum length of deletion, and as so the iterative decoding is able to retrieve it regardless of its position within the codeword [25]. Furthermore, we assume that the deletions occur in a single burst. Yang and Ryan [8] developed an efficient algorithm to determine L_{max} in terms of bits. The *efficiency* $\eta(H)$ of an n -length code with rate k/n is defined as the ratio between its L_{max} and the information transmitted, namely:

$$\eta(H) = \frac{L_{max}}{n-k} \quad (1)$$

In this paper we investigate the possibility of constructing an LDPC code capable of performing a successful decoding of erasure and bursts through the construction of a parity-check matrix H less random. In order to do this, the burst erasure efficiency of the check matrix $\eta(H)$ is determined from the concatenation of new base matrices superposed by circulant matrices. In Section 2, we present the mathematical fundamentals used in the construction of the binary matrix that will be used as a platform matrix. In section 3, we present the procedures for constructing the platform matrix and the algorithms that generate the parity-check matrix used for decoding LDPC codes. Finally, Section 4 details findings and

conclusions found in the analysis and performance of LDPC codes that were implemented by the proposed matrices.

2. Base matrices

The base matrix is a sparse binary matrix that uses the superposition of other matrices in its non-zero entries to construct the parity-check matrix of a code [19]. To decode burst errors, in this work, we propose two base matrices that are grounded on the construction of geometrically uniform signs matched with groups [20], [22] and on the formation of Nested Polygons based on Davis [23].

A signal constellation S is any discrete subset in \mathbb{R}^n . The elements of a signal constellation S are called signal points. A *Euclidian-space code* is a subset of, S^I where $I \subseteq \mathbb{Z}$.

The diversity of a communication system can be increased by using specific signal constellations by means of ASK modulation diversity MPSK M-QAM, MFSK [20]. This diversity can be defined as the minimum number of distinct components of the two vectors in an S n -dimensional signal constellation, or a minimum Hamming distance in S . Geometrically, the action of a rotation in S constellation characterizes the modulation diversity so that there are a maximum number of distinct components.

The signal constellations obtained via rotation are known as rotated signal constellations. In n -dimensional Euclidean-spaces the constellations can be characterized as a lattice in the cubic form of the type \mathbb{Z}_n . Thus, an x point of the rotated constellation is obtained by the action of an \mathbb{R} -matrix in the u -vector, that is, the set of points $\{x = u\mathbb{R}, u \in \mathbb{R}_n\}$.

Loeliger in [20] and Forney in [22] proposed algebraic procedures to obtain 2PSK 3PSK, 4PSK signal constellations matched to additive groups from the additive structure of the signal fields that are matched to groups $(\Delta_3, \Delta_4, \mathbb{Z}_6)$. These procedures were based on the classical results of modern algebra (Lattices). It is assumed that the reader is familiar with notions like *group theory*, *subgroups*, *lateral classes*, *homomorphisms*, etc., which comprise the theoretical background that is necessary to be able to comprehend this text.

A signal constellation S is *geometrically uniform* [19], if $s_1, s_2 \in S$ exist $\varphi \in Isom(\mathbb{P}^n)$ such that $\varphi(s_1) = s_2$ and $\varphi(s) = s$. If $\Gamma(S) = \{\varphi \in Isom(\mathbb{P}^n) : \varphi(s) = s\}$, then S is an orbit of any point $s_0 \in S$ under $\Gamma(S)$, that is,

$$S = \{\varphi(s_0) : \varphi \in \Gamma(S)\} = \bigcup_{\varphi \in \Gamma(S)} \{\varphi(s)\}.$$

A block code with the same type of energy [23] is any finite signal constellation on a sphere that generates \mathbb{P}^n as a vector space, which we denote by α . The number of signal points or codeword end d is the quadratic minimum distance of S . In particular, when an S spherical code is geometrically uniform we say that S is a uniform constellation. This is introduced by Slepian in [23] who uses the *group codes* name for the Gaussian channel; it is generalized by Forney [19] for any signal constellation.

A signal constellation S is matched with a group \mathcal{G} if there is a μ -mapping μ of \mathcal{G} over S , such that [17]:

$$d(\mu(g), \mu(h)) = d(\mu(e), \mu(g^{-1}h)), \quad \forall g, h \in \mathcal{G} \quad (2)$$

where $d(.,.)$ is the quadratic Euclidean distance and “ e ” describes the unit of G .

The μ -mapping is called a *matched mapping*. When μ -mapping is a *matched labelling*, that is, if G is isomorphic to $G(S)$, then μ is an *isometric labelling*. Let be C a linear code over Z_m of length n , [17] then we can define the following:

$$\phi: C \rightarrow R^n, \phi((c_1, \dots, c_n)) \square \sum_{j=1}^n A_j B_j \quad (3)$$

where,

$$A_j = \left[r_j \cos\left(\frac{2\pi c_j}{m}\right), r_j \sin\left(\frac{2\pi c_j}{m}\right) \right], B_j = \begin{bmatrix} b_{2j-1} \\ b_{2j} \end{bmatrix} \quad (4)$$

and $\{b_1, b_2, \dots, b_{2n}\}$ is an orthonormal base in P^2 . Matrices A_j and B_j are coordinates of constellation points. Note that r_j is just an energy parameter that depends on j but not on the codeword. In general, we take $r_j = 1$, for every $j = 1, \dots, n$. The mapping ϕ is called canonical mapping.

The first base matrix (5) is constructed by connecting the geometric signal points of the vertices of the antiprism matched with Z_6 . [20] shows this algebraic geometric association starting from steps linking the points of polyhedra (prism and antiprism) to the d-chain developed from a subgroup of G . The matrix (5) expresses, in general, the mapping applied between the matched antiprism with a signal constellation (remainder classes or dihedrons) and a canonical base of P^n . Fig. 1 shows the matched antiprism with the Z_6 group.

$$[B] = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \quad (5)$$

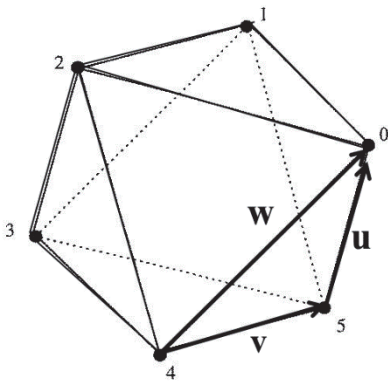


Figure 1. Antiprism matched with Z_6

Source: The authors

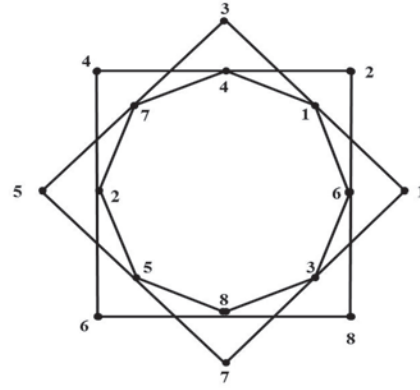


Figure 2 Nest formed by two squares and one hexagon.
Source: The authors

The second matrix is based on the construction of a "nest" formed by an octagon and two squares. The idea of "nest" comes from the proportional division of the sides of a polygon. Davis [21] constructs these nested polygons by connecting the dots of this division; the second polygon is formed by connecting the points generated by the first polygon, the third polygon is constructed analogously, and so on. Fig. 2 shows the nest formed by two squares and one octagon, the division ratio of which is $\frac{1}{2}$. The matrix is associated with that nest.

$$A = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix} \quad (6)$$

Take the matrix $B=2A$ and then construct a $C_{8 \times 6}$ matrix that operates at the B columns. Using the notation C_j , $j = \{1, \dots, 6\}$, we denote the six columns of the matrix C . Do: $C_1 = B_1$, $C_3 = B_4$, $C_5 = B_7$, $C_6 = B_8$ e $B_2 + B_3 = C_2$, $B_5 + B_6 = C_4$ and obtain the following matrix:

$$C = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix} \quad (7)$$

Deleting the sixth column and the last three rows of this matrix gives the following binary matrix:

$$[B]_2 = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \quad (8)$$

3. Construction of a Platform Matrix and a Check Matrix

The check-parity matrix is the key element to determine the L_{max} [14] when correcting burst erasure. The construction of the parity-check matrix made by Galague [16] is almost random, but parity matrices consisting of circulant matrices are systematic. They have been studied by several authors to design LDPC erasure codes in bursts with good efficiency [12], [24], [11]. In this section, we present a procedure to design LDPC codes to correct erasures in bursts using a superposition of matrices [19]. A combination of three types of sparse matrices is used: the base matrix, the platform matrix and the circulant matrix.

Let $M = (c_1, c_2, \dots, c_v)$ a circulant matrix of order v . When $c_1 = 1$ and the other values are zero, the M circulant is the identity matrix of order v . Thus, the vector $I_v = (1, 0, \dots, 0)$, with $v \in N$ will be the identity matrix generator. Using the notation $I_v^{(m)}$ with $m \in N$, it is possible to describe $\text{mod } v$ circulating movements in the identity matrix I_v . The matrix 0 represents the zero matrix of order v . The following matrix $I_8^{(6)}$ represents a circulant identity matrix of size 8 and movement $m = 6$. The zero-elements were discarded.

The platform matrix is an auxiliary matrix that is used to construct the parity-check matrix. The technique developed for the construction of the platform matrix is that proposed by [14], which is made of superposed sub-matrices in non-zero entries of the binary matrix. In this paper the sub-matrices are called client matrices and the platform matrix is developed from the composition of two client matrices: the circulant matrix and the zero matrix, both of order v .

A 4-cycle occurs in H if two columns of H contain non-zero entries in the same two rows. Thus, concatenating and alternating circulating matrices and null matrices in the final construction of the parity-check matrix, the following algorithms will give us 4-cycle free matrices. Avoiding 4-cycles can improve the performance of the decoded interaction and therefore the decoding burst [11].

Example 1: The following is the platform matrix H_p that was constructed from the random superposition of the client matrix $I_8 = (1, 0, 0, 0, 0, 0, 0, 0)$ in the base matrix H_b in a 5-dimension formed by the elements: $H_b(i, j) = 1$ for $i = j$ with

$1 \leq i, j \leq 5$ in the first diagonal and $H_b(2,1)=H_b(3,2)=H_b(4,3)=H_b(5,4)=1$ in the diagonal just below and other null elements. In the first diagonal of H_p $m \in \{2, 3, 4, 5, 6\}$, matrices $I^{(2)}$ and $I^{(6)}$ represent the movements $2 \bmod 8$ and $6 \bmod 8$ circulant of I_8 . Zero elements were discarded.

$$H_p = \begin{bmatrix} I^{(2)} & & & & \\ I & I^{(3)} & & & \\ & I & I^{(4)} & & \\ & & I^{(1)} & I^{(5)} & \\ & & & I & I^{(6)} \end{bmatrix}_{40 \times 40}$$

We will now present two algorithms that generate parity-check matrices using the binary matrices developed in section II. The first algorithm was developed to analyze and check performance of the following codes: LDPC C1(500,250); C2(1800,1200); C3(3000,1500); C4(4158,3465); C5(3750,3125), and the second algorithm was developed to construct the following codes: LDPC C6(1500,1200); C7(4200,3600). The algorithm choice is subject to the length of the code. Algorithm 2 is mainly used for codes with multiple dimension values of 5. Algorithm 1 is mainly used for other dimensions.

Algorithm-1. This algorithm constructs a parity-check matrix of dimension Npv and rate $\approx (p-1)/p$ from the concatenation of p copies of the base matrix (5) of dimension N . Given a sequence $\{B_1, B_2, \dots, B_p\}$ of p copies (5), the first platform H_1 is created by superposition in B_1 , the second H_2 by superposition on B_2 , and so on.

Step 1: Consider $\{B_1, B_2, \dots, B_p\}$ binary matrices of (5), with dimension n

Step 2: Construction of platform matrix H_1 :

- A) In non-zero entries of the main diagonal of B_1 substitute by I_v
- B) In non-zero entries of the diagonal immediately below the main diagonal of B_1 substitute by I_v and make $I_v^{(1)}$ in at least one of the elements

Step 3: Construction of platform matrices H_2, \dots, H_p :

- A) In non-zero entries of the main diagonal superpose I_v^m m is selected so that the sum of the movements of I_v^m is a multiple of N .
- B) In non-zero entries of the diagonal immediately below the main diagonal, I repeat step 2.B

Step 4: Superpose the matrix O_v in the zero elements of $\{B_1, B_2, \dots, B_p\}$

Step 5: In the free element in matrices $\{B_1, B_2, \dots, B_p\}$, superpose I_v^m with any movement.

Step 6: Construct parity-check matrix H_{chq} free of 4-cycles, dimension $N \times pN$ and rate $\approx (p-1)/p$ concatenating $\{H_1, H_2, \dots, H_p\}$

Table 1.

Client matrix and Random Movements - Algorithm 1

Platform Matrix	D ₁ , D ₂ , D ₃ , D ₄ , D ₅	S ₁ , S ₂ , S ₃ , S ₄	Free Element (F)
H ₁	I, I, I, I, I	I, I ⁽¹⁾ , I, I	I
H ₂	I ⁽²⁾ , I ⁽³⁾ , I ⁽⁴⁾ , I ⁽⁵⁾ , I ⁽⁶⁾	I, I, I ⁽¹⁾ , I	I
H ₃	I ⁽²⁾ , I ⁽³⁾ , I ⁽⁴⁾ , I ⁽⁵⁾ , I ⁽⁶⁾	I, I, I ⁽¹⁾ , I ⁽¹⁾	I
H ₄	I ⁽⁶⁾ , I ⁽⁵⁾ , I ⁽⁴⁾ , I ⁽³⁾ , I ⁽²⁾	I, I, I ⁽¹⁾ , I ⁽¹⁾	I ⁽⁵⁾
H ₅	I ⁽⁸⁾ , I ⁽⁶⁾ , I ⁽⁵⁾ , I ⁽⁴⁾ , I ⁽²⁾	I, I, I ⁽¹⁾ , I ⁽¹⁾	I ⁽⁶⁾
H ₆	I ⁽⁹⁾ , I ⁽⁷⁾ , I ⁽⁵⁾ , I ⁽³⁾ , I ⁽¹⁾	I, I, I, I ⁽¹⁾	I
H ₇	I ⁽⁷⁾ , I ⁽⁵⁾ , I ⁽⁴⁾ , I ⁽³⁾ , I ⁽¹⁾	I, I ⁽¹⁾ , I, I	I

Source: The authors

Example 2: Next, we have H_i, with $\{1 \leq i \leq 7\}$ platform matrices for $N=5$. The D₁D₂D₃D₄D₅ notation represents the main diagonal of H_i, while S₁S₂S₃S₄ represents the diagonal immediately below the main one, F indicates the free element. Zero elements were discarded. Table 1 shows some movements that were proposed in the client matrix I.

In the construction of the LDPC code C1(500,250) the concatenations H₁H₂, H₁H₃ and H₁H₄ ($v = 50$, $p = 2$, rate 0.5) and H₁H₂H₃H₄ ($v = 25$, $p = 4$) are produced. For the construction of code C2(1800,1200) ($v = 120$, $p = 3$, rate 0.66) we used the concatenations H₁H₂H₃ and H₁H₃H₄ and in the construction of code C3(3000,1500) we used H₁H₄ ($v = 300$, $p = 2$, rate 0.5). With $N=5$ and $N=3$ we construct analogously with small variations in client matrices, codes LDPC C5(3750,3125) ($v = 125$, $p = 6$, rate 0.833), and C4 (4158,3465) ($v = 231$, $p = 6$, rate = 0.833) respectively. Next, we have check matrices H_{b1} = [H₁H₄] for LDPC code C1(500,250) and H_{b2}=[P₁P₂] (P₁, P₂ auxiliaries matrices) for LDPC code C4 (4158,3465). Zero elements were discarded.

$$H_{b1} = \begin{bmatrix} I & & & I & I^{(6)} & & I^{(5)} \\ I & I & & & I & I^{(5)} & \\ & I^{(1)} & I & & & I & I^{(4)} \\ & & I & I & & I^{(1)} & I^{(3)} \\ & & & I & I & & I^{(1)} & I^{(2)} \end{bmatrix}$$

$$P_1 = \begin{bmatrix} I & & I & I^{(2)} & & I & I^{(2)} & I \\ I & I & & I & I^{(3)} & & I & I^{(4)} \\ & I^{(1)} & I & & I^{(1)} & I^{(4)} & & I^{(1)} & I^{(6)} \end{bmatrix}$$

$$P_2 = \begin{bmatrix} I^{(6)} & & I^{(5)} & I^{(8)} & & I^{(6)} & I^{(1)} & I \\ I & I^{(5)} & & I & I^{(5)} & & I & I^{(3)} \\ & I^{(1)} & I^{(4)} & & I^{(1)} & I^{(2)} & & I^{(1)} & I^{(5)} \end{bmatrix}$$

Algorithm 2's sequence of steps is analogous to Algorithm 1's, differing only in the fact that the platform matrix 8 has a dimension of 5 and has a larger degree of freedom in choosing the movements of client matrices.

Algorithm-2. This algorithm constructs a parity-check matrix with dimension Npv and rate $\approx (p-1)/p$ from the concatenation of p copies of the platform matrix (8). Given

a sequence $\{B_1, B_2, \dots, B_p\}$ of p copies of (8), we have:

Step 1: Take $\{B_1, B_2, \dots, B_p\}$ p copies of binary matrices

of (8).

Step 2: Construct the platform matrices H_1, H_2, \dots, H_p

attributing matrices $sI_v^{(m)}$ by superposition in $\{B_1, B_2, \dots, B_p\}$ so that the sum of random movements is mod5.

Step 3: Superpose matrix O_v in the null elements of $\{B_1, B_2, \dots, B_p\}$

Step 4: Construct the parity-check matrix H_{cheq} free of 4-cycles, dimension $5 \times 5p$ and rate $\approx (p-1)/p$ concatenating $\{H_1, H_2, \dots, H_p\}$

Example 3: Next, we have five platform matrices G_i built from algorithm 2, whose elements are denoted by letters A,B,C,D,E,F,J,L,M, which also designate the superposed client matrices. Zero elements were discarded, see Table 2. The code LDPC C6(1500,1200) was constructed from the following concatenations: $G_1G_1G_1G_1G_1$, $G_1G_2G_2G_2G_2$ and $G_2G_2G_2G_2G_2$ ($v=60$, $p=5$, rate 0.8) and code LDPC C7(4200,3600) by using the concatenations $G_1G_1G_2G_2G_3G_3G_3$ and $G_1G_2G_3G_4G_5G_5G_4$ ($v=24$, $p=7$, rate 0.857).

$$G_i = \begin{bmatrix} A & C & & & & & & \\ & & E & J & & & & \\ B & & & & & L & & \\ & D & F & & & & & \\ & & & K & M & & & \end{bmatrix}$$

4. Simulation and comparison of results

The purpose of this section is to show that the LDPC codes developed from algorithms 1 and 2 perform well when correcting burst erasure in BuEC and also to show that the parity-check matrices developed from the constructions have an excellent performance in correcting these erasures.

4.1. Simulations in BuEC

The motivation for considering a burst erasure channel to model channels with memory is their simplicity as well as the fact that the bit loss of information, often sent, occurs in them. Any system where the receiver is capable of distinguishing long losses in bits can treat this period of failures as a burst erasure channel. Simulations were produced in BuEC with values between -2 and 20 decibels. An SP algorithm with soft decision was used with an error probability $p = 0.01$, with 100 interactions. To analyze the performance of the LDPC codes, the parity-check matrices used were constructed from algorithms 1 and 2 from examples 2 and 3. For code C1(500,250), rate 0.5, we have four concatenations proposed by algorithm1: H₁H₂, H₁H₃, H₁H₄ and H₁H₂H₃H₄ with L_{max} 240, 241, 248 and 118 bits, respectively. It is observed that H₁H₄ has the best efficiency (99.2 %). It can be seen that H₂, H₃, H₄ have few differences in the movements of client matrices contained in the main diagonal, however, H₄ has a

Table 2.

Client matrix and Random Motions – algorithm 2

Platform Matrix	A	B	C	D	E	F	J	K	L	M
G ₁	I ⁽³⁾	I ⁽¹⁾	I ⁽²⁾	I ⁽³⁾	I ⁽⁴⁾	I ⁽⁵⁾	I ⁽⁵⁾	I ⁽⁵⁾	I ⁽⁶⁾	I ⁽⁶⁾
G ₂	I ⁽⁶⁾	I ⁽⁶⁾	I ⁽⁶⁾	I ⁽⁶⁾	I ⁽⁶⁾	I ⁽⁶⁾	I ⁽⁶⁾	I ⁽⁶⁾	I ⁽⁶⁾	I ⁽⁶⁾
G ₃	I ⁽⁴⁾	I ⁽³⁾	I ⁽²⁾	I ⁽¹⁾	I ⁽⁵⁾	I ⁽⁶⁾	I ⁽⁷⁾	I ⁽⁹⁾	I ⁽²⁾	I ⁽¹⁾
G ₄	I ⁽⁵⁾	I ⁽⁵⁾	I ⁽⁵⁾	I ⁽⁵⁾	I ⁽⁵⁾	I ⁽⁵⁾	I ⁽⁵⁾	I ⁽⁵⁾	I ⁽⁵⁾	I ⁽⁵⁾
G ₅	I ⁽⁶⁾	I ⁽⁵⁾	I ⁽⁴⁾	I ⁽³⁾	I ⁽²⁾	I	I ⁽³⁾	I ⁽²⁾	I	I

Source: The authors

Type code	Length code	Rate	Lmax (bits)	Efficiency
[11, N=5, p=2, v=50, pp 648]	500	0.5	248	0.992
[13, Table I]	2040	0.515	509	0.51
[12, Table IV]	2000	0.5	786	0.786
[14, Margulis-Table I]	2640	0.5	1033	0.782
[14, PEG IRA-Table I]	2000	0.5	403	0.403
[11, Table I]	3000	0.5	1468	0.978
Algorithm-1				
N=5, p=2, v=50	500	0.5	248	0.992
N=5, p=2, v=300	3000	0.5	1475	0.983
Algorithm-2				
p=2, v=300	3000	0.5	1498	0.999
[12, Table VI]	500	0.7	121	0.345
[13, Table I]	8176	0.753	1021	0.50
Algorithm-1,				
N=5, p=4, v=25	500	0.75	180	0.994
N=5, p=4, v=408	8180	0.75	2037	0.996
[11, N=5, v=300]	1500	0.8	291	0.97
[13, PEG regular-Table I]	4608	0.8752	287	0.499
[11, Table I]	4158	0.8333	682	0.927
[11, Table I]	16500	0.9	1648	0.999
Algorithm-1,				
N=3, p=6, v=231	4158	0.8333	682	0.927
Algorithm-2				
p=5, v=60	1500	0.8	296	0.986
p=5, v=120	3000	0.8	592	0.986
p=6, v=100	3000	0.833	490	0.980
p=6, v=550	16500	0.833	2748	0.999

Table 3.

Burst Correction Properties of Selected Codes

Source: The authors

significantly different movement in its spare element $F=I^{(5)}$ relative to the free elements of H_2 and H_3 , i.e. $F=I$.

Fig. 3 shows the code efficiency proposed in algorithm 1 when correcting burst erasure.

For code C2(1800, 1200) we proposed the concatenations of the matrices $H_1H_2H_3$, $H_1H_2H_4$ developed in example 2 with L_{max} 572 and 576 bits respectively. The best yield observed is the concatenation $H_1H_2H_4$ which, in turn, has the matrix $I^{(5)}$ as a client matrix in its free element. In a preliminary conclusion, it may be said that the most "abrupt" movement of the free element is responsible for better code performance.

For code C6(1500, 1200), the concatenations proposed were the ones developed in example 3: $G_1G_1G_1G_1G_1$, $G_1G_2G_2G_2G_2$ and $G_2G_2G_2G_2G_2$ ($v = 60$, $p = 5$, rate 0.8) with L_{max} 290, 292, 296 bits respectively. The last concatenation obtained the best yield (98%). For code C7(4200, 3600), we obtained the following concatenations: $G_1G_1G_2G_2G_3G_3G_3$ and $G_1G_2G_3G_4G_5G_5G_4$ ($v = 24$, $p = 7$, rate 0.857) with L_{max} 582, 580 bits respectively. The last concatenation obtained the best yield (96%). Fig. 4 shows the efficiency of the proposed code in Algorithm 2 when correcting burst erasure.

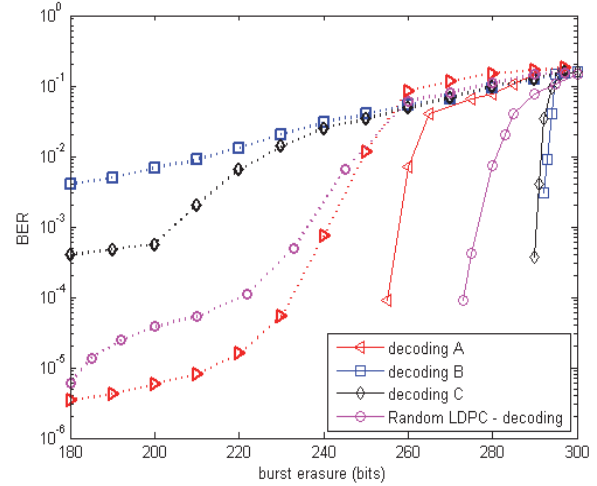


Figure 3. Performance of LDPC codes A= H_1H_2 , B= H_1H_3 , C= H_1H_4 , length 500, rate 0.5 in a random erasure channel; probability $p=0$ (solid curve) and $p=0.01$ (dotted curve).

Source: The authors

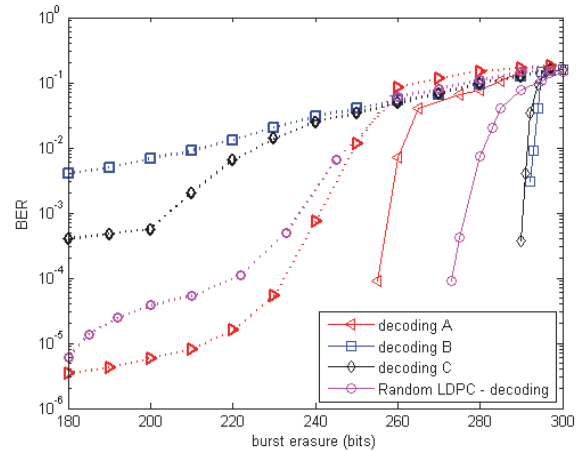


Figure 4. Performance of an LDPC with code length 1500, rate 0.8, in random erasure channel and probability $p=0$ (solid curve) and $p=0.01$ (dotted curve).

Source: The authors

In any of the constructed codes the one that presents the best performance is the code which has the client matrix with a number that indicates the larger movement: $I^{(6)}$ in C6(1500,1200) and $I^{(5)}$ and $I^{(6)}$ in C7(4200, 3600). This indicates that the larger the movement of the client matrix is the better the decoding performance.

For large bursts in BuEC, LDPC codes have been an efficient corrector for errors and erasures [20-23]. Table 3 compares the LDPC codes' correction properties for burst erasure that were constructed in this article with previous results. Fig. 5 shows the performance of an LDPC code with a length of 4170 and rate of 0.833 when correcting burst errors via algorithms 1 and 2 ($N = 5$, $v = 139$, $p = 6$) in the classic bursty channel. Note that the codes randomly constructed and simulated here are not truly random as they have been optimized to avoid repeated columns and cycles of length 4, when possible.

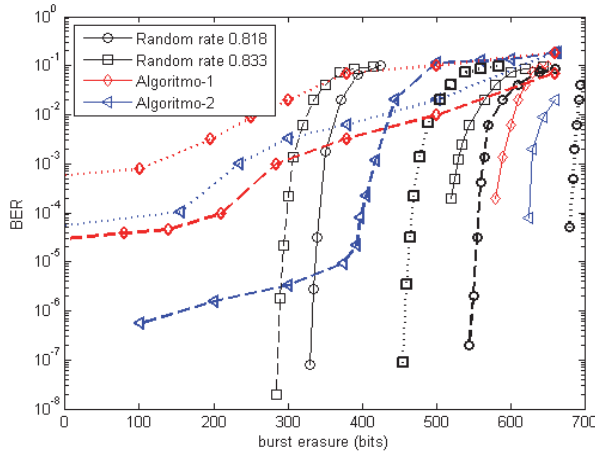


Figure 5. The performance of LDPC codes with a 4170 length on a classic bursty channel with guard band erasure probabilities of $p = 0$ – solid curves, $p = 0.01$ – dashed curves, and $p = 0.05$ – dotted curves
Source: The authors

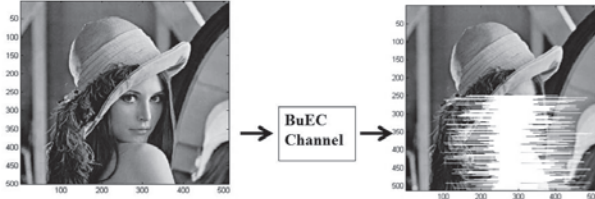


Figure 6(a): Left, a 262.144 pixel image. Figure 6(b): Right, there is a random simulation of the image after it passed through a channel with erasure. The deleted range represents the erasure probability varying from 0.01 to 0.99 (amounting to 58.623 lost pixels) in random points of the image
Source: The authors

4.2. Erasure recovery in a Noisy Channel using the proposed code - Simulation

We will now present the transmission simulations results through a noisy channel. In the simulation, we used the Lena.gif which has a size of 512X512 (262.144 pixels) and an LDPC code with parity-check matrices developed by algorithms 1 and 2. This channel sent a codeword with a length of 1512 and a rate of 0.8, which was adapted to the encoding software. The noisy channel introduces losses of bits in a random point in the codeword and the erased bits are recovered in the receiver.

The Lena.gif (512x512) original image is displayed in Fig. 6(a). The probability of channel erasure varied from 0.01 to 0.99 in increments of 0.05. In this range, the number of lost or erased bits is approximately 20% to 50% of the total amount of bits of the codeword. Finally, with a 0.99 erasure probability (extremely hostile channel) we obtain a lost bit in a range from 90% to 98% of the codeword at the output. Fig. (6) shows Lena's image in the output of the channel with the information erased by the channel

The image obtained in the channel output goes through a recovery process by means of LDPC codes that use algorithms 1 and 2 in the development of parity-check matrices to decode by sum-product. Fig. 7(a) shows the image that was retrieved

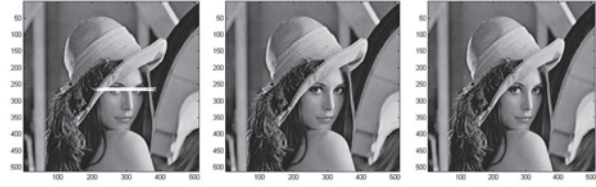


Figure 7(a). Left, Image recovery via LDPC decoding using algorithm 1 with free element $F=I$ (56864 recovered pixels). Figure 7(b). In the central image, we have algorithm 1 and $F=I^{(5)}$ (57450 recovered pixels). Figure 7(c). Right, algorithm 2 with client matrices alternating in $I^{(5)}$ and $I^{(6)}$ (58037 recovered pixels).
Source: The authors

using the concatenated matrices with free element $F=I$. It can be seen that the decoder does not recover 5% of the initial erasure track, and in the most hostile track the decoder does not recover 12% of the bits transmitted by codewords. Fig. 7(b) shows the image retrieved when $F=I^{(5)}$. In this case, regardless of the erasure track, the decoder recovers 98% of the lost bits. When client matrices $I^{(5)}$ and $I^{(6)}$ were used in algorithm 2, an efficiency of 98% recovery of lost bits is achieved, as Fig. 7(c) shows. Thus, we obtained the following result: we can see that the best performance for decoding comes when comparing the performance of the codes for which the values $F=I$ and $F=I^{(5)}$ are used for the free element F in the concatenated matrices, produced by the algorithm 1, $F=I^{(5)}$.

5. Conclusion

In this paper, two algorithms were presented to develop parity-check matrices to generate LDPC codes based on the concatenation of base matrices superposed by circulant client matrices weight 2. The simulations showed that the codes generated by parity-check matrices that were obtained from the proposed base matrices and by the movements of client matrices showed a good efficiency in terms of the correction of burst erasures. Comparing H_1H_2 , H_1H_3 and H_1H_4 , it can be seen that the movement of the free element $I^{(5)}$ of H_4 was responsible for the improvement and efficiency of the code. We can also arrive at this conclusion by comparing the concatenations proposed in example 2 by using the matrix $I^{(6)}$ as a client. Table 3 shows that Algorithm-2 ($p = 6$, $v = 550$), as a client, achieves the same efficiency as the construction proposed by Sara [11]. Although algorithms 1 and 2 are designed with mod 5 dimension matrices, it was possible to evaluate simulations with different dimensions because the vector of circulant matrices enables the combination of numbers that express non-multiple dimensions of 5. The proposed algorithms were used to generate LDPC codes to simulate erasure correction in Lena's image, they were transmitted through a noisy channel, and their results were satisfactory.

Acknowledgements

We would like to thank the support of Federal University of Pará and Federal University of Western Pará for their help with this work.

References

- [1] Riascos-Erazo, S.C., A model for assessing information technology effectiveness in the business environment, *Ingeniería e Investigación*. [Online]. 28(2), pp 158-166, 2010. Available at: <http://www.revistas.unal.edu.co/index.php/ingenv/article/view/14905>
- [2] Martínez-Roso, R., El papel de la comunicación en los sistemas generales. *Ingeniería e Investigación*. [Online]. 22, pp. 68-71, 2012. Available at: <http://www.revistas.unal.edu.co/index.php/ingenv/article/view/20668>
- [3] Correa-Espinal, A., Gómez-Montoya, R.A., Tecnologías de la información en la cadena de suministro. *DYNA*, [Online], 76(157), pp. 37-48, 2009. Available at: <http://www.revistas.unal.edu.co/index.php/dyna/article/view/9551/1475>
- [4] Manrique-Losada, B., Echeverri-Arias, J. and Peláz-Rodríguez, M., Hydroinformatics' contribution to Amazonian water resources and ecosystem management. *Ingeniería e Investigación*, [Online], 31(1), pp. 108-116, 2011. Available at: <http://www.revistas.unal.edu.co/index.php/ingenv/article/view/20532/35684>
- [5] Moreno-López, G.A. and Jiménez-Builes, J.A., Cycle of PDCA T-learning model and its application on interactive digital TV. *DYNA*, [Online], 9(173), pp. 61-70, 2012. Available at: http://dyna.medellin.unal.edu.co/en/verPDF.php?id_articulo=a07v79n173-II&tipo=articulo&id=173-II
- [6] Zhihua, D. and Zhen, J., LDPC codes by circulant decomposition based on difference family, Proceedings of the International MultiConference of Engineers and Computer Scientists, [Online]. Vol I, IMECS, March 18-20, Hong Kong, 2009. Available at: http://www.iaeng.org/publication/IMECS2009/IMECS2009_pp384-386.pdf
- [7] Forney, G.D., Burst-correcting codes for the classic bursty channel, *IEEE Trans. Commun.*, 19(5), pp. 772-781, 1971. DOI: 10.1109/TCOM.1971.1090719
- [8] Yang, M., Ryan, W. and Li, Y., Design of efficiently encodable moderate length high-rate irregular LDPC codes, *IEEE Trans. Communications*, 52(4), pp. 564-571, 2004. DOI: 10.1109/TCOMM.2004.826367
- [9] M. Yang, W. Ryan, Performance of efficiently encodable low-density parity-check codes in noise bursts on the EPR4 channel, *IEEE Trans. Magn.*, 40(2), pp. 507-512, 2004. DOI: 10.1109/TMAG.2004.824125
- [10] MacKay, D.J.C., Good error-correcting codes based on very sparse matrices, *IEEE Trans. Inform. Theory*, 45(2), pp. 399-431, 1999. DOI: 10.1109/18.748992
- [11] Johnson, S.J., Burst erasure correcting LDPC codes, *IEEE Trans. Commun.*, 57(3), pp. 641-652, 2009. DOI: 10.1109/TCOMM.2009.03.060468
- [12] Hosoya, G., Yagi, H., Matsushima, T. and Hirasawa, S., A modification method for constructing low-density parity-check codes for burst erasures, *ICICE Trans. Fundamentals*, E89-A(10), pp. 2501-2509, 2006. DOI: 10.1093/ietfec/e89-a.10.2501
- [13] Tai, Y.Y., Lan, L., Zeng, L., Lin, S. and Abdel-Ghaffar, K.A.S., Algebraic construction of quasi-cyclic LDPC codes for the AWGN and erasure channels, *IEEE Trans. Commun.*, 54(10), pp. 1765-1774, 2006. DOI: 10.1109/TCOMM.2006.881361
- [14] Paolini, E., Chiani, M. and Calzolari, G.P., Construction of near-optimum burst erasure correcting low-density parity-check codes *IEEE Trans. Inform. Theory*, 57(5), pp. 1320-1328, 2009. DOI: 10.1109/TCOMM.2009.05.070047
- [15] Xu J. and Lin, S., A combinatoric superposition method for constructing LDPC codes, in Proc. International Symposium on Information Theory (ISIT'2003), Yokohama, Japan, June 2003, 30P. DOI: 10.1109/ISIT.2003.1228044
- [16] Gallager, R.G., Low-density parity-check codes, *IRE Trans. Inform. Theory*, IT-8(1), pp. 21-28, 1962. DOI: 10.1109/TIT.1962.1057683
- [17] Kan L., Aleksandra, K. and Erden, M., Fatih, construction of burst-erasure efficient LDPC codes for use with belief propagation decoding, *Proc. IEEE* 2010. DOI: 10.1109/ICC.2010.5502511
- [18] Peng, F., Yang, M. and Ryan, W.E., Simplified eIRA code design and performance analysis for correlated Rayleigh fading channels, *IEEE Trans. Wireless Commun.*, 5(4), pp. 720-725, 2006. DOI: 10.1109/TWC.2006.1618918
- [19] Xu, J., Chen, L., Zeng, L.-Q., Lan, L. and Lin, S., Construction of low density parity-check codes by superposition, *IEEE Trans. Commun.*, 53(2), pp. 243-251, 2005. DOI: 10.1109/TCOMM.2004.841966
- [20] Loeliger, H.A., Signal sets matched to groups, *IEEE Trans. Inform. Theory*, IT-37, pp. 1675-1682, 1991. DOI: 10.1109/18.104333
- [21] Davis, P.J., *Circulant Matrices*, Chelsea Publishing, New York, Press, 1994.
- [22] Forney, Jr., G.D., Geometrically uniform codes, *IEEE Trans. Inform. Theory*, 37, pp. 1241-1260, 1991. DOI: 10.1109/18.133243
- [23] Slepian, D., Group codes for gaussian channel, *Bell Syst. Tech. J.*, 47, pp 575-602, 1969. DOI: 10.1002/j.1538-7305.1968.tb02486.x
- [24] Fossorier, M., Quasi-cyclic low density parity check codes from circulant permutation matrices, *IEEE Trans. Inf. Theory*, 8, pp. 1788-1793, 2004. DOI: 10.1109/TIT.2004.831841

E.G. Pelaes, received his BSc. in Physics in 1977 and BSc. in Electrical Engineering in 1977, both from the Federal University of Pará, Brazil, his MSc. in Electrical Engineering in 1982 from the Pontific Catholic University of Rio de Janeiro, Brazil, and his PhD in Electrical Engineering in 1998 from UNICAMP University. He is currently an associated professor at the Federal University of Pará, Brazil. He has experience in Electric Engineering, with emphasis in Processing of Signals. His current research includes the following subjects: processing of signals, transformed wavelet and codification of images, models propagation, communications and quality of energy.
ORCID: 0000-0001-8572-6131

C.A.S. Silva, received his BSc. in Mathematics in 1992 from the State University of Pará, Brazil, and his MSc in Pure Mathematics in 2003 from the Federal University of Paraíba, Brazil. He is an adjunct professor at the Federal University of Western Pará where he teaches Modern Algebra, Theory of Numbers, and History of Mathematics.
ORCID: 0000-0002-9126-9560.



UNIVERSIDAD NACIONAL DE COLOMBIA

SEDE MEDELLÍN
FACULTAD DE MINAS

Área Curricular de Ingeniería
Eléctrica e Ingeniería de Control

Oferta de Posgrados

Maestría en Ingeniería - Ingeniería Eléctrica

Mayor información:

E-mail: ingelcontro_med@unal.edu.co
Teléfono: (57-4) 425 52 64

Heat flow assessment in an underground mine: An approach to improve the environmental conditions

Marc Bascompta ^a, Ana María Castañón ^b, Lluís Sanmiquel ^a & Josep Oliva ^a

^a Department of Mining and Natural Resources, Polytechnic University of Catalonia (UPC), Barcelona, Spain. marc.bascompta@upc.edu, sanmi@emrn.upc.edu, josep@emrn.upc.edu

^b Department of Mining, Topography and Structures, University of León (ESTIM), León, Spain. amcasg@unileon.es

Received: July 29th, 2015. Received in revised form: February 5th, 2016. Accepted: February 24th, 2016.

Abstract

The generation of heat in underground spaces due to working activities is a factor that influences production and productivity rates. This paper analyses the heat generation in an underground mine and provides a number of approaches to enhance the ventilation conditions using electrical, instead of diesel machines. This assessment has been carried out using theoretical equations and modelling software. Investigations prove that sensible and latent heat would be reduced by around 50% and 84% respectively if the change were applied in the case study. This reduction on heat input to the ventilation system would improve the workplace environment because of lower effective temperatures and gas concentrations, which would result in better safety conditions and higher employee efficiency.

Keywords: mine ventilation, health and safety, efficiency, heat generation, mining equipment.

Evaluación de los flujos de calor en una mina subterránea y enfoque para mejorar sus condiciones ambientales

Resumen

La actividad minera en espacios subterráneos genera un aporte de calor al sistema de ventilación que tiene influencia en los niveles de producción y productividad. Este artículo analiza y cuantifica las fuentes de calor en una mina subterránea y propone una alternativa de mejora de las condiciones ambientales mediante un cambio de los equipos diésel por maquinaria eléctrica. Este análisis se apoya en varias expresiones teóricas y programas para modelizar la ventilación. Los resultados muestran una reducción del calor efectivo y aparente del 50% y 84%, respectivamente, una vez aplicados los cambios de equipos en el caso estudiado. La reducción del calor en el sistema de ventilación permitiría una mejora de las condiciones en el lugar de trabajo debido a una menor temperatura efectiva y del nivel de contaminantes, incrementado el nivel de eficiencia de los trabajadores y mejorando el nivel de seguridad.

Palabras clave: ventilación subterránea; seguridad y salud; eficiencia; maquinaria minera; generación de calor.

1. Introduction

Heat flow is an important aspect associated to underground mine ventilation, on which mining equipment has a significant impact. As the work goes deeper and the mine evolves, factors such as temperature and humidity become crucial to keep acceptable environmental conditions and fulfil the legal requirements. Besides, efficiency rates and safety levels are also influenced by this factor. Many

studies have been carried out regarding gases generated by diesel engines [1,2] and the incidence of temperature in underground mines [3-5].

The reduction of heat flow in these cases is usually focused on optimising the efficiency of the refrigeration system and cutting down its operating costs through an improvement of the current systems [6-10], but this important issue has not been approached when trying to change the mining equipment. Diesel equipment has an overall

How to cite: Bascompta, M., Castañón, A.M., Sanmiquel, L. and Oliva, J., Heat flow assessment in an underground mine: An approach to improve the environmental conditions. DYNA 83(197), pp. 175-180, 2016.

efficiency of about one third of the electrical units. Hence, the usage of fuel will produce approximately three times as much heat as electrical machines for the same mechanical work output [11]. Moreover, the combustion process generates harmful pollutants that have to be controlled, especially in the mining sector where the conditions are quite adverse in terms of health and safety [12].

Apart from the type of energy source, there are other important factors that affect the underground air temperature; for instance, the outer climate, the area's geological factors or the method used for mineral extraction [13].

This paper determines these different heat inputs in an underground potash mine by means of empirical equations and modelling software. After this, the heat flow contribution of electrical and diesel equipment is compared in order to expose an alternative to improve the environmental conditions in an underground infrastructure. The procedure followed is:

- Determination of the heat contribution of each source in the case study.
- A comparison of the situation using electrical energy instead of diesel trucks and loaders.

2. Heat input measurement methodology

The data used in theoretical equations and modelling software have been provided by mine staff, and measured in situ between 2008 and 2014 or extracted from bibliography in the case of the initial iterations with the software. The equipment features have been obtained from the manufacturer's data.

First, the airflow behaviour has been determined using Vnet. Fig. 1 is a scheme of the model achieved by means of the software. These initial results will be used to know the climatic conditions of the airways and the heat sources (strata heat, equipment and fragmented rock).

2.1. Strata heat

Heat emission from the strata depends on the type of rock, the exploitation method and depth and length of the airways. However, the amount of heat transmitted decreases over time, the working faces being where the greatest transmission takes place. Sometimes, strata heat can be obtained using empirical methods based on other similar mines [11]. Unfortunately, there is no such information in this case.

Whillier [14] exposed an equation method that defines two expressions depending on the time since the tunnel was opened. If it has been open for more than 30 days, eq. (1) is used to determine the radial heat flow.

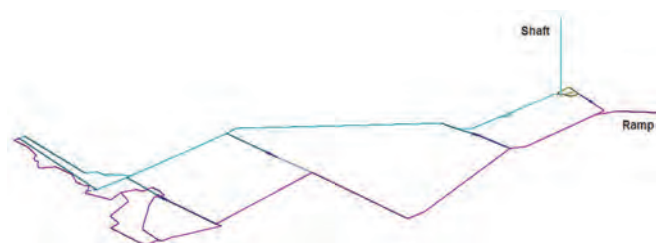


Figure 1 Modelling scheme of the mine using Vnet.
Source: Authors.

$$q = 3,35 \cdot L \cdot k^{0.854} \cdot (VRT - \theta_d) \quad (1)$$

Where q is heat flow from the strata (W); L is length of the tunnel (m); k is thermal conductivity of the rock ($W/m \cdot ^\circ C$); VRT is virgin rock temperature ($^\circ C$); θ_d is mean dry bulb temperature ($^\circ C$). Meanwhile eq. (2) is applied if the advance has taken place within the last 30 days.

$$q = 6 \cdot k \cdot (L + (4 \cdot DFA)) \cdot (VRT - \theta_d) \quad (2)$$

Where L is the length of the drift dug over the last 30 days (m), which cannot be greater than the length advanced in the last month; DFA is daily face advance (m). The main problem from eq. (1) is to find out the period that heat is transferred from the strata to the air until thermal equilibrium is achieved. This setback has been solved by modelling the strata behaviour using ClimSim. The software takes into account the heat flow transferred to the air by radiation and convection methods, determining the heat flow of a circular tunnel for a certain homogenous rock. Heat flow determination is based on the radial heat conduction from Fourier's equations, expressed in polar cylindrical coordinates (W/m^2).

$$k \left[\frac{r}{\partial r^2} + \left(\frac{\partial \theta}{\partial r} \right) + \frac{1}{r} \left(\frac{\partial^2 \theta}{\partial \phi^2} \right) + r \frac{\partial^2 \theta}{\partial z^2} \right] = r \rho C \frac{\partial \theta}{\partial t} \quad (3)$$

Heat transfer can be either from the strata to the air or from the air to the strata depending on where temperature is higher, this continues until there is thermal equilibrium. When airways have been open for a long time, a phenomenon called "thermal flywheel" could arise, transferring heat from the air to the strata during the day and the opposite at night [15].

Fig. 2 explains the ClimSim functioning. First, the climatic variables have to be calculated or measured. Once the initial model is built, it has to be compared with real measures to validate it, applying iterations as many times as necessary to achieve a proper model.

After several iterations, rock conductivity and diffusivity were obtained, $6 W/m \cdot ^\circ C$ and $5.55 m^2/s \cdot 10^{-6}$ respectively. According to the manual, values are considered acceptable when there is a difference of around $\pm 1 ^\circ C$ between modelled and measured mean values, between 2008 and 2014 in this case. Moreover, the iterations have been carried out in two different zones and four periods of the year in order to achieve more reliable results. Table 1 displays the temperature difference once the modelling is correctly adjusted.

On the other hand, Fig. 3 details the effective temperatures, calculated according to Spanish law (RGNBSM, itc 04.7.02), $t_e = 0.9 \cdot t_w + 0.1 \cdot t_d$, where t_e is effective temperature, t_w wet temperature, and t_d dry temperature.

Subsequently, base modelling has been used to calculate the length of the tunnels giving heat to the airways, changing the variables within the software called "age in" and "age out", which take into account the time since the tunnel was opened, until the sensible heat reaches a value of zero. In this case, the contribution of sensible heat to the airways is near zero after approximately one year.

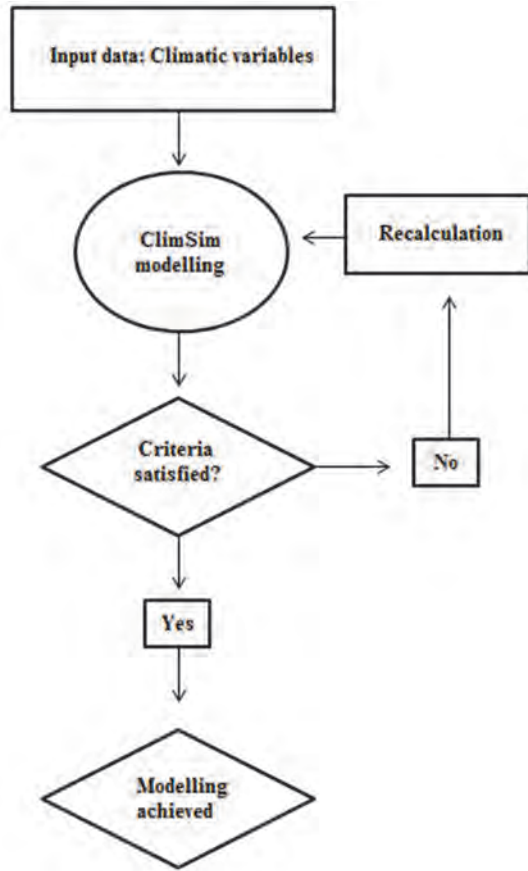


Figure 2 Scheme of the ClimSim functioning based on the user's manual explanation.
Source: Authors.

Table 1.
Temperature comparison of two points from the ventilation layout.

Point 1						
Period	Dry bulb temperature °C			Wet bulb temperature °C		
	Measured	ClimSim	Difference	Measured	ClimSim	Difference
Overall	24	21.62	2.38	17	15.96	1.04
January	19	17.60	1.40	11	12.07	-1.07
April	25	23.07	1.93	18	15.87	2.13
June	30	27.80	2.20	22	20.65	1.35
October	27	24.95	2.05	19	17.77	1.23
Point 2						
Overall	26	25.45	0.55	17	17.93	-0.93
January	26	20.90	5.10	14	12.40	1.60
April	28	26.20	1.80	18	18.81	-0.81
June	31	30.70	0.30	23	22.62	0.38
October	31	27.95	3.05	21	19.75	1.25

Source: Authors.

After that, it has been calculated theoretically to corroborate the modelled values, giving an average variation of only 9.97% both ways. Table 2 and Fig. 4 detail the behaviour of the strata heat using ClimSim.

2.2. Mechanized equipment

The exploitation method determines the heat contribution from the equipment to the ventilation system, there being a huge difference in terms of heat generation between the usage

of diesel and electrical energy. Fig. 5 describes the steps to determine the heat input generated by electrical machines.

On the other hand, the efficiency of diesel machines is, approximately, 1/3 that of the electrical equipment and produces either sensible or latent heat, whereas the electrical

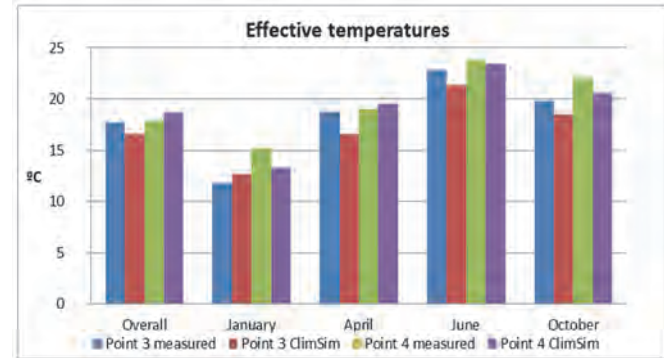


Figure 3 Comparison of the effective temperatures modelled and measured in situ.
Source: Authors.

Table 2
Behaviour of the strata modelled by means of ClimSim.

Age in (days)	Age out (days)	Sensible heat (kW)	Latent heat (kW)	Months the tunnel has been open
30	0	17.93	33.72	1
90	60	6.00	27.28	3
182	152	2.64	25.97	6
365	335	0.13	25.08	12
730	700	-1.85	24.43	24

Source: Authors.

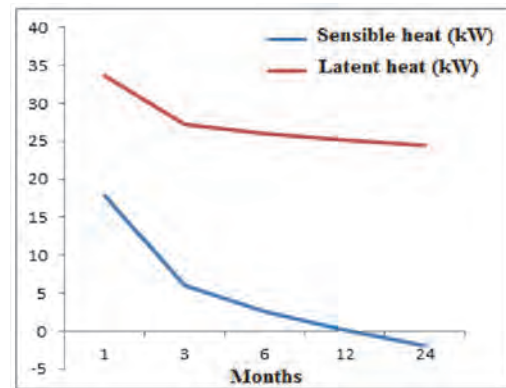


Figure 4 Behaviour of the strata heat, either sensible or latent.
Source: Authors.

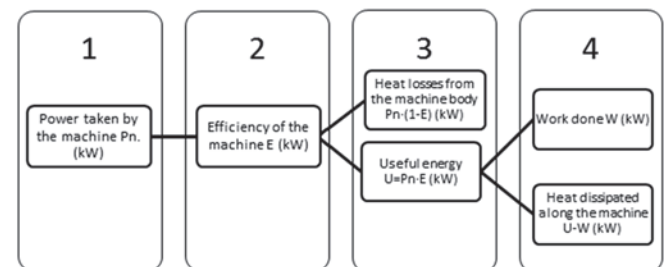


Figure 5 Scheme of the heat generated by electrical machines.
Source: Authors.

equipment only produces sensible heat. The three main heat sources are: 1) radiator and body of the machine, 2) combustion gases, and 3) movement and friction due to the usage of the machine [11]. Its quantification can be achieved considering a ratio of 0.3 litres of diesel per 1 kW per hour, with a calorific value of 34000 kJ/litre, and producing a heat generation of 2.83 kW for each kilowatt of mechanical output.

Each litre of fuel consumed produces around 1.1 litres of water due to combustion gases [16]. However, this value could be several times higher because of the refrigeration system. Some in situ analyses have pointed out that this ratio could vary from 3 to 10 litres per litre of fuel consumed, depending on the power and maintenance [17]. The following equation determines the total heat generated, which comprises latent and sensible heat.

$$q_c = c \cdot \frac{E_c}{100} \cdot PC \quad (4)$$

Where q_c is heat emitted by the combustion (kW); c is combustible (l/s); E_c is combustion efficiency (%); and PC is combustible calorific value (kJ/l). McPherson [11] gives some references for combustion efficiency (95%), and the rate of liquid equivalent per litre of fuel (5). This last parameter is necessary to calculate the quantity of water generated by the combustion.

$$W = c \cdot \frac{E_c}{100} \cdot r \quad (5)$$

Where W is water generated (l), and r is rate of liquid equivalent. After determining the water generated, the latent heat is obtained by taking into account a standard value of the water latent vaporization heat, 2450 kJ/kg, and an equivalency of 1:1 litre-kilogram of water.

$$q_l = \lambda_w \cdot W \quad (6)$$

Where q_l is latent heat (kJ), and λ_w is water latent vaporization heat (kJ/kg). Finally, sensible heat can be obtained by deducting latent heat from the result in eq. 4. Later on, the figures were transformed to kW in order to compare the values for 8 hours of work.

2.3. Fragmented rock

When fragmented rock is exposed to the ventilation airstream and there is a difference between rock and air temperature, heat transference is generated following the expression below [11].

$$q_{fr} = m \cdot C \cdot (\theta_1 - \theta_2) \quad (7)$$

Where q_{fr} is heat load due to rock fragmentation (kW); m is mass flow of the mineral exploited (Kg/s); C is the specific heat of the rock (kJ/kg°C); θ_1 is the temperature of the rock immediately after fragmentation (°C); and θ_2 is temperature of the fragmented rock at the exit of the ventilation system (°C). Temperature θ_1 can be considered equivalent to the virgin rock temperature with enough accuracy according to McPherson [11].

3. Mining equipment

Tables 3 and 4 detail the current mining equipment in the case study and the features needed to determine the heat input.

Table 3.
Diesel equipment characteristics.

Type	Quantity	Nominal power (CV)	Nominal power (kW)	Consumption (l/h)
Truck	22	400	294	67
Loader	12	300	221	58
Car	64	100	74	14
Jumbo	3	90	66	14
Auxiliary equipment	6	88	65	14

Source: Authors.

Table 4.
Electrical equipment characteristics.

Type	Nominal power (kW)	Quantity
Continuous haulage machine	165	1
	180	1
	220	5
Conveyors	56	2
	110	1
	180	1
	200	2
	400	3
	600	6
Continuous miner	529	10

Source: Authors.

The electrical trucks and loaders chosen have very similar sizes and capacities to the diesel ones. The models used were the Scooptram ST1030 and Scooptram EST1030 for the diesel and electrical loader respectively and the MT436B and EMT35 for the trucks.

4. Results and discussion

The heat inputs described above are shown in Tables 5 and 6, taking into account the latent and sensible heat contribution of each source. Table 5 exposes the results for the current situation (diesel loaders and trucks), whereas Table 6 shows the results after the change proposed; electrical loaders and trucks instead of diesel ones.

Table 5.
Heat input using diesel trucks and loaders.

Source of heat	Sensible heat (kW)	Latent heat (kW)	Contribution (%)
Machines	11093	6248	73.8
Conveyors	1072		4.6
Continuous haulage machine	145		0.6
Miners	1455		6.2
Fragmented rock	297		1.3
Strata	1102	2072	13.5
Total	15163	8320	100.0

Source: Authors.

Table 6.

Heat input using electrical trucks and loaders.

Source of heat	Sensible heat (kW)	Latent heat (kW)	Contribution (%)
Machines	5609	987	51.8
Conveyors	1072		8.4
Continuous haulage machine	145		1.1
Miners	1455		11.4
Fragmented rock	297		2.3
Strata	1102	2072	24.9
Total	9679	3059	100.0

Source: Authors.

As can be deduced from tables above and Fig. 6, the main source of sensible or latent heat is the machinery itself. Overall, the change of the loaders and trucks would reduce the contribution of heat from the mining equipment by 23%, and sensible and latent heat would decrease by about 36% and 63% respectively.

Furthermore, Table 7 exposes the current fleet of vehicles using diesel and the proposal, together with their heat generation and the percentage variation of both options.

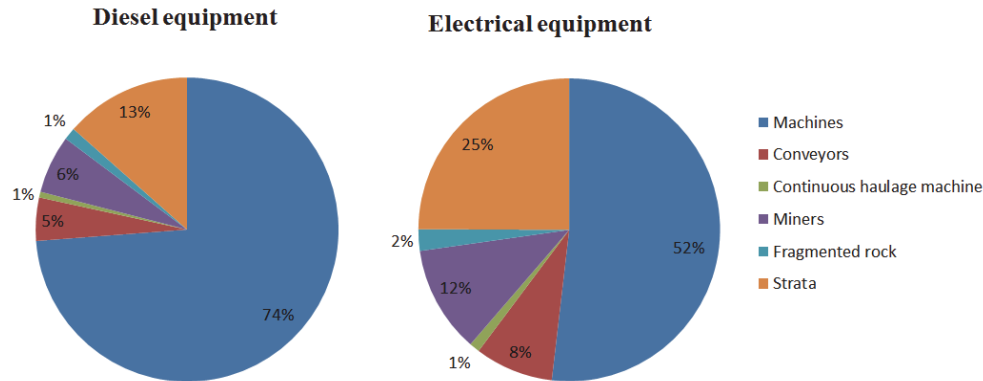


Figure 6 Percentage variation of the different heat inputs using electrical or diesel equipment.

Source: Authors.

Table 7.

Summary of the heat generated by the vehicles.

Type	Unit heat (kW)		Total sensible heat (kW)		Difference of sensible heat (%)	Total latent heat (kW)		Difference of latent heat (%)
	Diesel	Electrical	Diesel	Electrical		Diesel	Electrical	
Truck	450.9	90	6345	1984	68.7	3574	0	-
Loader	390.3	156	2996	1873	37.5	1687	0	-
Car	39.3	-	1607	¹	-	905	²	-
Jumbo	25.1	-	48	¹	-	27	²	-
Auxiliary equipment	25.1	-	96	¹	-	54	²	-
Total			11093	5608	49.4	6248	986	84.2

¹Only loaders and trucks are changed. Thus, other diesel using equipment is added to the total sensible heat input after applying the proposal.

Source: Authors.

The unit heat per machine is considerably reduced using electrical equipment, and the results above show a huge difference in terms of heat generation. Besides, as fuel consumption would be cut down, the generation of pollutants such as NO_x, CO or CO₂ would also decrease. Taking a ratio of 1:1 quantity of pollutants-litres of diesel burned, the generation would be minimized by 88% based on the data used.

Despite the considerable improvements of the hypothetical change, it has to be pointed out that these machines need a trolley or a cable in the majority of the cases to match the power required, reducing the flexibility of the vehicle fleet. Thus, a mixture of both kinds of equipment may be necessary. On the other hand, more research to achieve suitable batteries needs to be undertaken.

5. Conclusions

The usage of electrical loaders and trucks decreases the generation of sensible heat by 49.4% and latent heat by 84.2%. Overall, the contribution of heat from machines dropped from 73.8% to 51.85%. In addition, the modelling by ClimSim allowed us to observe the behaviour of strata heat in a potash mine, finding out the trend of sensible and latent heat in the airways.

Apart from the electrical engines' higher energy efficiency, less consumption of diesel would mean a drop in temperature and pollutants concentration. Therefore, ventilation requirements would be reduced and a better workplace environment could be achieved, leading to higher productivity and production rates. The usage of electrical equipment can also help to reduce the uncertainty in future

mining activity due to oil price variations and more restrictive legal requirements.

Acknowledgements

The authors would like to thank the staff at ICL-Iberia for their willingness and the Iberpotash Chair in Sustainable Mining from the Polytechnic University of Catalonia (UPC).

References

- [1] Kurnia, J.C., Sasmito, A.P., Wong, W.Y. and Mujumdar, A.S., Prediction and innovative control strategies for oxygen and hazardous gases from diesel emission in underground mines. *Science of the Total Environment*, 481, pp. 317-334, 2014. DOI: 10.1016/J.SCITOTENV.2014.02.058
- [2] Mantilla-Gonzalez, J.M., Roncancio, G.J. y Acevedo, H.R., Análisis comparativo del desempeño y emisiones de un motor diésel de gran capacidad operando bajo dos escenarios: Trabajo en ruta activa y trabajo en banco. *Ingeniería e Investigación*, 30, pp. 118-124, 2010.
- [3] Payne, T. and Mitra, R., A review of heat issues in underground metalliferous mines. 12th U.S./North American Mine Ventilation Symposium, pp. 197-202, 2008.
- [4] Garcia-Herrero, S., Mariscal, M.A., García-Rodríguez, J. and Ritzel, D.O., Working conditions psychological physical symptoms and occupational accidents. Bayesian network models. *Safety Science*, 50, pp. 1760-1774, 2012. DOI: 10.1016/J.SSCI.2012.04.005
- [5] Bluhm, S., Moreby, R., Von Glehn, F. and Pascoe, C., Life-of-mine ventilation and refrigeration planning for Resolution Copper Mine. *The Journal of the Southern African Institute of Mining and Metallurgy*, 114, pp. 497-503, 2014.
- [6] Del Castillo, D., Air cycle refrigeration system for cooling deep mines. *International Journal of Refrigeration*, 11(2), pp. 87-91, 1988.
- [7] Swart, C., Optimising the operation of underground mine refrigeration plants and ventilation fans for minimum electricity cost. Dissertation for Doctoral Degree North-West University, USA, 2003.
- [8] Hardcastle, S., Kocsis, C. and Li, G., Analyzing ventilation requirements and the utilization efficiency of the Kidd Creek mine ventilation system. 12th U.S./North American Mine Ventilation Symposium, pp. 27-36, 2008.
- [9] Vosloo, J., Liebenberg, L. and Velleman, D., Case study: Energy savings for a deep-mine water reticulation system. *Applied Energy*, 92, pp. 328-335, 2012.
- [10] Edgar, G., Plessis, D., Liebenberg, L., Mathews, E.H. and Nicolaas, J., A versatile energy management system for large integrated cooling systems. *Energy Conversion and Management*, 66, pp. 312-325, 2013.
- [11] McPherson, M.J., *Subsurface ventilation engineering*. USA: Springer, 1993. ISBN: 978-94-010-4677-0.
- [12] Freijo, M., Sanmiquel, L.I., Edo, J.J. y Vintró, C., Estudio de los accidentes laborales en la industria minera y en la construcción en España en la última década. *DYNA*, 86(6), pp. 726, 2011.
- [13] Xiaojie, Y., Qiaoyun, H., Jiewen, P., Xiaowei, S., Dinggui, H. and Chao, L., Progress of heat-hazard treatment in deep mines. *Mining Science and Technology (China)*, 21(2), pp. 295-299, 2011.
- [14] Whillier, A., Predicting cooling requirements for caving and sublevel stoping in hot rock. *Int. Conf. on Caving and Sublevel Stopping*. AIME Denver, 1981.
- [15] Stroh, R., A note on the downcast shaft as a thermal flywheel. *Journal of the Mine Ventilation Society of South Africa*, 32, pp. 77-80, 1979.
- [16] Kibble, J.D., Some notes on mining diesels. *Mining Technology*, pp. 393-400, 1978.
- [17] McPherson, M.J., The analysis and simulation of heat flow into underground airways. *International Journal of Mining and Geological Engineering*, 4, pp. 165-196, 1986.

M. Bascompta, MSc., is working as a graduate research assistant in the Polytechnic University of Catalonia, in the Department of Mining and Natural Resources.
ORCID: 0000-0003-1519-6133

A.M. Castañón, PhD is working as a professor at the University of Leon in the Department of Mining, Topography and Structures.
ORCID: 0000-0002-3177-5111

L. Sanmiquel, PhD is working as a professor at the Polytechnic University of Catalonia, in the Department of Mining and Natural Resources.
ORCID: 0000-0001-5612-4713

J. Oliva, PhD is working as a professor at the Polytechnic University of Catalonia, in the Department of Mining and Natural Resources.
ORCID: 0000-0001-6214-5713



UNIVERSIDAD NACIONAL DE COLOMBIA

SEDE MEDELLÍN
FACULTAD DE MINAS

Área Curricular de Ingeniería
Geológica e Ingeniería de Minas y Metalurgia

Oferta de Posgrados

Especialización en Materiales y Procesos
Maestría en Ingeniería - Materiales y Procesos
Maestría en Ingeniería - Recursos Minerales
Doctorado en Ingeniería - Ciencia y Tecnología de
Materiales

Mayor información:

E-mail: acgeomin_med@unal.edu.co
Teléfono: (57-4) 425 53 68

Secure point-to-point communication using chaos

Maricela Jiménez-Rodríguez ^a, María Guadalupe González-Novoa ^a, Juan Carlos Estrada-Gutiérrez ^a,
Cuauhtemoc Acosta-Lúa ^a & Octavio Flores-Siordia ^a

^a Departamento de Ciencias Tecnológicas, Centro Universitario de la Ciénega, Universidad de Guadalajara, Ocotlán, Jalisco, México.
m_jimenez_r@yahoo.com, gleznogpe@hotmail.com, jcarlosredes@gmail.com, temo09@gmail.com, o_flores@live.com.mx

Received: October 9th, de 2015. Received in revised form: March 1st, 2016. Accepted: March 15th, 2016

Abstract

This article presents an alternative for resolving the existing vulnerability of systems that implement masking by means of chaotic synchronization. This system avoids detection of the parameters used as the cipher key by an attacker on implementing encryption by means of synchronized chaotic-phase masks of the Rössler oscillator, for encoding and establishing synchronization among transmitter-receiver devices. In addition, it employs two ciphering keys: the first, with a recommended length of 2,048 characters, and the second, which is used as an initial value. Both keys are employed for continual modification of one of the oscillator's parameters. This strengthens the security system and avoiding an attacker from obtaining the oscillator's parametric values by calculating the least average synchronization error. The use of the system developed provides a cipher, which is resistant to statistical attacks. In addition, our system validates the data of the transmitter device (username, password, etc.) in order to authorize transmission.

Keywords: authentication; data encryption; security; integrity; chaos.

Comunicación punto a punto segura usando caos

Resumen

Este artículo presenta una alternativa para resolver la vulnerabilidad existente en los sistemas que implementan el enmascaramiento mediante la sincronización caótica, este sistema evita que los parámetros utilizados como clave de cifrado puedan ser detectados por un atacante al implementar el modelo matemático caótico del oscilador de Rössler para codificar y establecer la sincronización entre los dispositivos transmisor-receptor; además usa dos llaves de cifrado: la primera con una longitud recomendable de 2048 caracteres y la segunda se utiliza como un valor inicial. Ambas llaves se emplean para modificar continuamente uno de los parámetros del oscilador, esto fortalece la seguridad del sistema y evita que un atacante obtenga los valores del parámetro del oscilador calculando el error de sincronización promedio menor. El uso del sistema desarrollado proporciona un cifrado resistente a ataques estadísticos, además valida datos del dispositivo transmisor (nombre de usuario, password, etc.) para autorizar la transmisión hacia el destino.

Palabras clave: autenticación; encriptación de datos; seguridad; integridad; caos.

1. Introduction

The role of chaotic systems is an excellent alternative for information security and privacy. This is due to such systems' great properties, such as high sensitivity to the initial conditions and to the parameters. Another interesting characteristic is that chaos uses frequencies that render it resistant to the customary filtering techniques to separate information superimposed on a signal [1]. In other words, a

message transmitted through a network can be encoded by chaos, using previously cited characteristics. Once transmitted, the information would be very difficult to decode, unless the receptor possesses the inverse-way method, considering the exact values of the parameters, in order to recover and be able to extract the original message with certainty. Some discreet chaotic systems have been used to cipher information, such as the technique implemented by Ranjan and Saumitr, in which the authors compress and

How to cite: Jiménez-Rodríguez, M., González-Novoa, M.G., Estrada-Gutiérrez, J.C., Acosta-Lúa, C. and Flores-Siordia, O., Secure point-to-point communication using chaos. DYNA 83(197), pp. 181-187, 2016

encode text using chaotic Logistic Map system. The authors employ an insecure channel to transmit the encoded data and another, secure channel to send the key [2]. Pareek et al. developed a symmetrical key algorithm for ciphering, in which they used multiple one-dimensional chaotic maps and an external key of 128 bits. Plain text is encoded sequentially using a set of chaotic maps in random fashion [3]. In addition, Pisarchik and Carmona elaborated chaotic map network-based algorithms to encrypt images in color. These authors used the Logistic Map, and as ciphered keys, they used the parameters, the number of iterations, the number of cycles, and the size of the image [4]. Hossam et al. proposed a cryptosystem for encrypting color images or videos that employs a mechanism of iterative encryption, in which each of the image's pixels depends on a secret key, of the logistic map exit and of the previously encrypted pixel [5]. Chaos is also used in the world of medicine. Barbara et al. proposes a transmission method allowing electrocardiogram (ECG) signals obtained from a patient to be combined with algorithms generating chaotic signals, based on the Lorenz equation system [6]. Another very important technique employed in secure communications is the synchronization of chaotic systems [7-13]. Tao and Len used the Chua circuit to modulate a signal in a parameter in a transmitter; subsequently, the authors used an adaptive controller in the receiver to maintain the synchronization and to recover the signal [14]. They developed an Image Encryption Algorithm where logistic map and iterative equation are used, and they switch the position and the pixels values [15]. The authors also employed two different chaotic cryptography techniques in which they employed the logistic map to apply the diffusion technique, and the chaotic synchronization of two, coupled Rössler oscillators to apply the confusion [16]. Zanin and colleagues carried out a systems cryptanalysis that uses synchronization for encryption through message masking, and the authors demonstrated how to detect the parameters used as encrypted keys, calculating the least synchronization error [17]. Thus, in this work, we implemented a method to strengthen the security of systems that use chaotic masking, avoiding the detection of parameters used as the encrypted key. In other studies, the authors have recommended prior encryption of the information before employing synchronization, in order to avoid detection of the parameters, but this approach implies more time to encode and decode [16,17]. A security analysis was carried out on different Wireless Local Area Networks (WLAN), where it was determined that one of the risks comprises the absence of mechanisms of authentication [18]. The method recommended in this work varies the parameter according to the way in which the mathematical system used for synchronizing conducts solving. Therefore, no extra time is required for previous ciphering. The identification of users who transmit information provides greater security. Thus, in this work, in addition to encoding, a technique is performed to validate users who attempt to transmit. The remainder of this article is organized as follows: in Section 2, the methodology for developing the system used is clearly explained, and in Section 3, we show how the communication channels for transmitting encrypted information in a point-to-point network using chaos are implemented. Later, in Section 4, we show the results of using tests to evaluate the

system's functionality; we have also included a statistical analysis and the conclusions obtained.

2. Methodology

2.1. Chaotic synchronization

Pecora and Carroll demonstrated unidirectional coupled chaotic systems [19], an excellent tool used in the area of secure communication. In this investigation, the Rössler oscillator is implemented to transmit encrypted information employing chaotic synchronization where the Master $m(t)$ oscillator is described in an eq. (1) system [20], in which x , y , z , are the system's state variables, and the system's parameters are rendered by: a_1 , a_2 , a_3 .

$$\begin{aligned}\frac{dx}{dt} &= -(y + z) \\ \frac{dy}{dt} &= (x + (a_1)y) \\ \frac{dz}{dt} &= a_3 + z(x + a_2)\end{aligned}\quad (1)$$

The Slave oscillator $s(t)$ is defined by the eq. (2) system, where x' , y' , z' are state variables and the system's parameters are: a_1 , a_2 , a_3 .

$$\begin{aligned}\frac{dx'}{dt} &= -(y' + z') \\ \frac{dy'}{dt} &= (x' + (a_1)y') \\ \frac{dz'}{dt} &= a_3 + z'(x' + a_2)\end{aligned}\quad (2)$$

Complete synchronization is an entity between the trajectories of two Master-Slave systems [19, 21], which can be determined by the synchronization error $e(t)$, that is $\lim_{t \rightarrow \infty} e(t) = 0$, where $e(t) \equiv ||m(t) - s(t)|| = 0$ [21]. In this study, chaotic synchronization is implemented using the Rössler oscillator for point-to-point communication in a secure manner, where the Master $m(t)$ system is used to encrypt in the transmitter and the Slave $s(t)$ in order to decrypt in the receptor; however, prior to this, these should be synchronized using variable y for coupling [19]. That is, the exit y_{master} is coupled with the entry y_{slave} to ensure synchronization. Functioning is exhibited in Fig. 1.

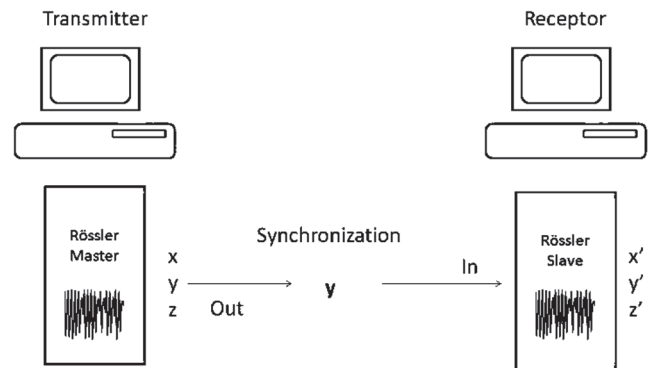


Figure 1. Master-Slave synchronization method. Source: Authors.

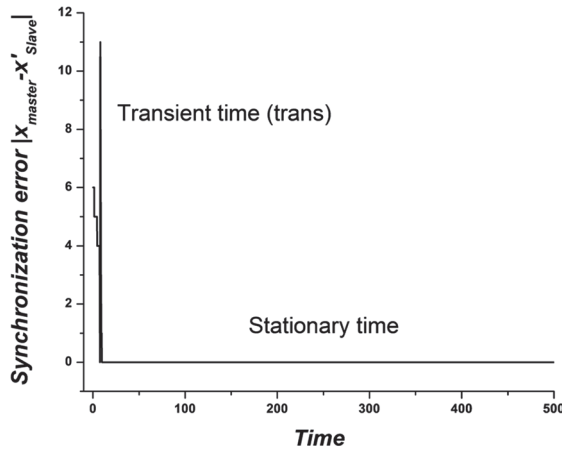


Figure 2. Transitory (trans) and stationary time determined by the synchronization error.
Source: Authors.

Prior to the synchronization, a transitory time $trans$ should pass, during which the Master-Slave systems yield different results until the stationary time, which is when the exit variables x, y, z of the Master system are identical to those of the exit variables x', y', z' of the Slave system. Thus, the synchronization error $e(t) = 0$ and the synchronization itself is complete, as depicted in Fig. 2.

When the encrypted transmitter and the receptor are completely synchronized, the receptor can decipher without any problem because both the transmitter and the receptor generate the same values of the variables. This technique is used for encoding information because it is very interesting that two irregular behaviors are based on a sole behavior, generating a chaotic orbit that is similar to noise that is very difficult to predict. This represents an important application in secure communications and cryptography. It is for this reason that this work was implemented.

3. System functioning

The system allows for point-to-point communication establishing channels to encrypt and transmit information with greater security through the network. In addition, the system is more robust because the transmitter should authenticate it in order to permit it to send information. Next, chaotic synchronization is implemented by means of the Rössler Master oscillator, equation system (1) to encrypt the final decrypted outcome using equation system (2). Below are the steps to follow:

1. The transmitter sends the name and password of the user to the receptor.
2. The receptor authenticates the user's data and the IP and MAC electronic addresses of the transmitter. If the latter are not correct, the transmission is cancelled.
3. If it is an allowed user, the authorization is sent.
4. The transmitter uses the algorithm to encrypt what is explained in Section 3.2.
5. The information is sent.
6. The receptor uses an algorithm of Section 3.2 for deciphering.

This process is illustrated in Fig. 3.

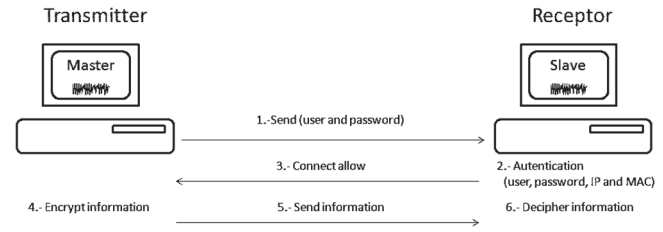


Figure 3. Master-Slave synchronization method.
Source: Authors.

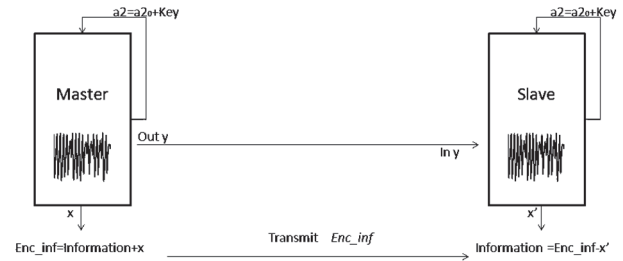


Figure 4. Data encryption and transmission.
Source: Authors.

3.1. Encryption with parameter variation

The transmitter solves equation system (1), modifying the value of parameter a_2 that is used as the encrypting key ($a_2 = a_{2_0} + key$), to avoid detection by an attacker. Once the transitory time ($trans$) has passed and complete synchronization is initiated, the transmitter encrypts the information by adding up $Enc_inf = Information + x_{master}$. After encrypting it together with the coupling variable y_{master} so that the receptor can synchronize itself, solving equation system (2), modifying the parameter. After this, it deciphers, subtracting $Information = Enc_inf - x'_{slave}$. This procedure is shown in Fig. 4.

3.2. Algorithms

In this section, we describe the functioning of the algorithms for encrypting and deciphering an archive of text, but any type of information can be encrypted by dividing it into bytes.

Nomenclature:

- C: Original information, with values between 0 and 255.
- n: Length of original information.
- trans: Number of iterations necessary to achieve synchronization (transitory time).
- l: Length of the key k.
- Encrypting keys:
 - a) k : External key used to modify parameter a_2 and to encrypt the information.
 - b) a_{2_0} : Initial value of parameter a_2 in the chaotic regime.

3.2.1. Algorithm for encryption

Step 1. Convert each character of the text file into its American Standard (ASCII) value.

$$Information = [C_1, C_2, C_3, \dots, C_n]$$

Step 2. Divide each element that contains the Information vector by 255 to obtain values between 0 and 1.

$$Information = [C_1, C_2, C_3, \dots, C_n] / 255$$

Step 3. Convert the key k into its ASCII value and store it in the vector.

$$key = [k_1, k_2, k_3, \dots, k_l]$$

Step 4. Divide each element of the key k by 255 to obtain values between 0 and 1.

$$key = [k_1, k_2, k_3, \dots, k_l] / 255$$

To vary parameter $a2$

Step 5. Calculate the value of $a2$ by adding $a2_0$ plus an element of the vector key and solve equation system (1). In the case of all of the elements of key having been already used, initiate the re-run of the vector key from the first position.

$$a2 = a2_0 + key_{1..l}$$

Step 6. After solving the system trans times, encrypt the information by adding $Information + x_{master}$ and storing the result in vector Enc_inf . Observe the functioning in Fig. 5.

Step 7. Store y_{master} in vector $y_sync = [y_1, y_2, y_3, \dots, y_{trans+n}]$

Step 8. Repeat steps 5 through 7 trans + n times.

3.2.2. Algorithm for deciphering

It is necessary to possess the encrypted information Enc_inf , the values of y_sync for synchronization of the Slave and the encrypted keys key , $a2_0$.

Step 1. Perform Steps 3 and 4 of the algorithm to encrypt.

Step 2. Calculate the value $a2$ by adding $a2_0$ plus one element of the vector key and solve equation system (2). In the case that all of the elements of the key have already been used, initiate the re-run from the first position.

$$a2 = a2_0 + key_{1..l}$$

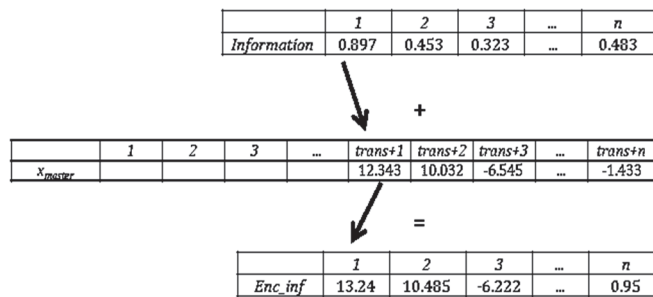


Figure 5. Encryption process using Step 6.
Source: Authors.

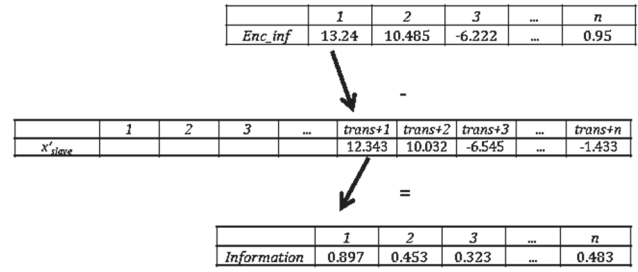


Figure 6. Deciphering information according to Step 3.
Source: Authors.

Step 3. After solving the trans times system, decipher the remaining information $Enc_inf - x'_{slave}$ and store the result in the vector Information. Observe the functioning in Fig. 6.

Step 4. Perform Steps 2 and 3 trans + n times to decipher all of the information.

Step 5. Multiply each element of the vector information by 255 to obtain the original information.

$$Information = [C_1, C_2, C_3, \dots, C_n] * 255$$

4. Results and Conclusions

The results obtained in the statistical tests employed to determine the system's robustness on encrypting and sending a text message are shown below.

4.1. Correlation diagram

These allow us to demonstrate graphically the relationship between two variables and in addition how to obtain the correlation coefficient, which can fall within the range of -1 and 1, indicating that the nearer the two they are to each other, the stronger the linear association will be. In the case of its nearing a 0, this indicates a weak or null association if this is 0.

The correlation diagram (Fig. 7) depicts the values of the deciphered information in the horizontal axis, and the values of the original information in the vertical axis. We can observe that the correlation coefficient is 1. Thus, the original information and the information that was deciphered is identical; that is, the system manages excellent data integrity, therefore, there is no information deformation at the moment of encrypting and deciphering.

In the correlation diagram (Fig. 8), the original vs. encrypted information is shown, and it yields a correlation coefficient of 0.000032329, which indicates that the relationship between encrypted and original information is nearly null. This is very favorable because some attackers study the relationship that exists between these two variables in order to attempt to determine the encryption key that has been employed.

4.2. Histograms

These are used to represent the distribution of the message, where the horizontal axis is shown with vertical bars and the data that are being transmitted. The height of each bar corresponds the number of frequencies of the data.

In Fig. 9, the data that are being transmitted (the original information) are between 0.1 and 1, with the majority of data being between 0.375 and 0.475.

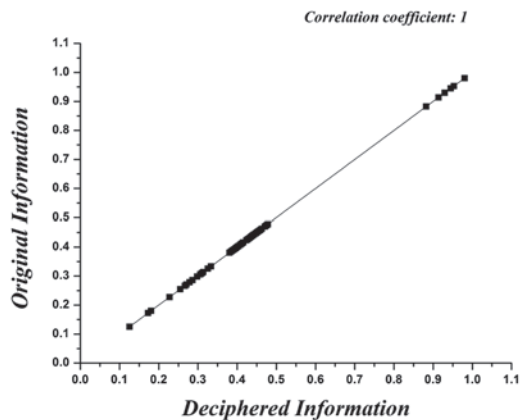


Figure 7. Correlation diagram: original vs. deciphered information.
Source: Authors.

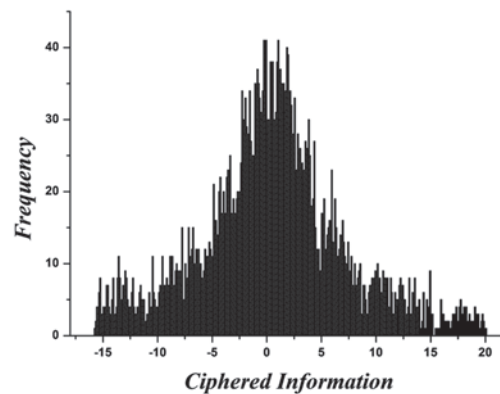


Figure 10. Histogram of ciphered information.
Source: Authors.

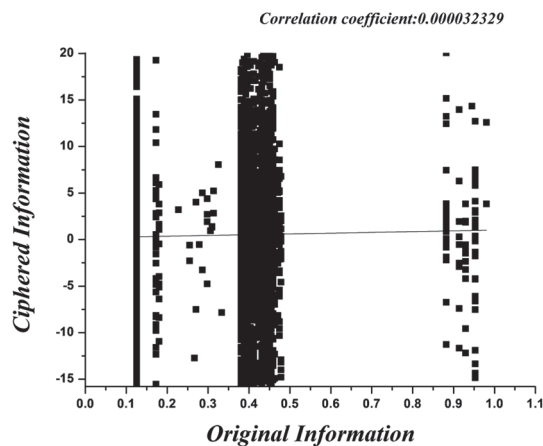


Figure 8. Correlation diagram of encrypted vs. original information.
Source: Authors.

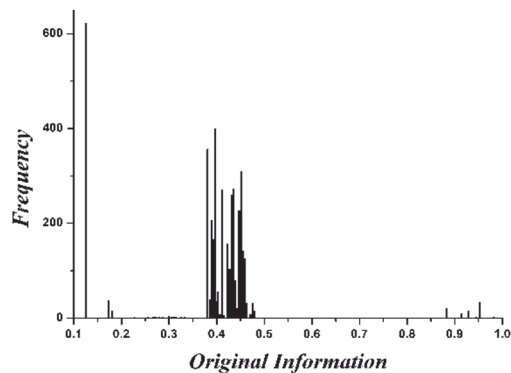


Figure 9. Histogram of original information.
Source: Authors.

Fig. 10 shows the histogram of the encrypted information, where it is clear that the encrypted data are found within the range of -15.25 and 20 . If Figs. 9 and 10 are compared, we can observe that there is no relationship between the number of frequencies and the range of values of the graphs' horizontal

axis. This helps to make it more difficult for an attacker to find a relationship between original and encrypted information.

4.3. Conclusions

The communications system developed truncates the orbit or trajectory by constantly changing one of the chaotic mathematical model's parameters; this speeds up the encoding time in comparison with other cryptographic systems that employ various orbits. Thus, more computer processing is necessary. Our system guarantees confidentiality because only the receiver, which possesses the keys used for encoding, can calculate the parameter's different values, in order to reconstruct the orbit; thus, decoding the information received. The constant change of the parameter employed avoids an attacker from detecting the parameter by using the least average synchronization error.

To verify the robustness of the system developed, the original information (plaintext) is compared to the ciphered text (ciphertext), using the correlation diagram depicted in Fig. 8, in which a coefficient near the value of 0 is exhibited. This indicates that the linear association is nearly null. With the latter, it is proven that the system is resistant to statistical attacks, such as those based on clear text (differential and linear), in which an attempt was made to determine the key on searching for some relationship between the ciphertext and the plaintext. Another of the tests conducted is illustrated in the histograms shown in Figs. 9 and 10, in which there are very different ranges of values and frequencies. This renders the system sufficiently robust for avoiding an attack in which the frequencies or repeat chains are analyzed in order to attempt to find correspondence between ciphered and deciphered information.

To evaluate the integrity of the data transmitted vs. the data received once the encoding and decoding process was accomplished, the correlation diagram in Fig. 7 provides a coefficient with a value of 1. This indicates that the information is not altered on passing through the ciphering and communications process.

The system employed within this investigation protects

the communication point-to-point, combining authentication and encoding techniques, in order to safeguard the information transmitted from one point to another. In this way, it complies with integrity, confidentiality, and security services. Additionally, the technique used for varying the parameter contributes robustness to the communications system.

Acknowledgment

The authors acknowledge the support from PRODEP-SEP (Convocatoria de Fortalecimiento de Cuerpos académicos) (Mexico).

References

- [1] Hilborn, R.C., Chaos and Nonlinear Dynamics. New York: Oxford University Press, 2000.
- [2] Ranjan, B. and Sumitr, P., A novel compression and encryption scheme using variable model arithmetic coding and coupled chaotic system. *IEEE Transactions on Circuits and System-I*, 53(4), pp. 848-857, 2006. DOI: 10.1109/TCSI.2005.859617.
- [3] Pareek, N., Patidar, V. and Sud, K., Cryptography using multiple one-dimensional chaotic maps. *Communications in Nonlinear Science and Numerical Simulation*, 10(7), pp. 715-723, 2005. DOI: 10.1016/j.cnsns.2004.03.006
- [4] Alexander, P. and Flores, N., Computer algorithms for direct encryption and decryption of digital images for secure communication. *Proceedings of the 6th WSEAS International Conference on Applied Computer Science*, (Tenerife, Canary Islands, Spain), pp. 29-34, 2006.
- [5] Hossam, E., Ahmed, H., Haamdy, K. and Osama, F., An Efficient chaos-based feedback stream cipher (ECBFSC) for image encryption and decryption. *Informatica*, 31, pp. 121-129, 2007.
- [6] Barbara, E., Alba, E. and Rodríguez, O., Modulating electrocardiographic signals with chaotic algorithms. *Ingeniería e Investigación*, 32(2), pp. 46-50, 2012.
- [7] Sundar, S. and Minai, A., Synchronization of randomly multiplexed chaotic systems with application to communication. *Physical Review Letters*, 85(25), pp. 5456-5459, 2000. DOI: 10.1103/PhysRevLett.85.5456
- [8] Parlitz, U., Kocarev, L., Stojanovski, T. and Preckel, H., Encoding messages using chaotic synchronization. *Physical Review E*, 53(5), pp. 4351-4361, 1996. DOI: 10.1103/PhysRevE.53.4351
- [9] Parlitz, U., Chua, L., Kocarev, L., Halle, K. and Shang, A., Transmission of digital signals by chaotic synchronization. *International Journal of Bifurcation and Chaos*, 2(4), pp. 973-977, 1992. DOI: 10.1142/S0218127492000562
- [10] Annovazzi, V., Donati, S. and Sciré, A., Synchronization of chaotic lasers by optical feedback for cryptographic applications. *IEEE Journal of Quantum Electronics*, 33(9), pp. 1449-1454, 1997. DOI: 10.1109/3.622622
- [11] Kocarev, L. and Parlitz, U., General approach for chaotic synchronization with applications to communication. *Physical Review Letters*, 74(25), pp. 5028-5031, 1995. DOI: 10.1103/PhysRevLett.74.5028
- [12] Zhao, Y., Experimental demonstration of chaotic synchronization in the modified Chua's oscillators. *International Journal of Bifurcation and Chaos*, 7(6), pp. 1401-1410, 1997. DOI: 10.1142/S0218127497001126
- [13] Cuomo, K., Oppenheim, A. and Strogatz, S., Synchronization of Lorenz-based chaotic circuits with applications to communications, *IEEE Transactions on Circuits and Systems- II: Analog and Digital Design Processing*, 40(10), pp. 626-633, 1993. DOI: 10.1109/82.246163
- [14] Yang, T. and Chua, L.O., Secure communication via chaotic parameter modulation. *IEEE Transactions on Circuits and Systems - I: Fundamental Theory and Applications*, 43(9), pp. 817-819, 1996. DOI: 10.1109/81.536758
- [15] Zhang, J. and Zhang, Y., An image encryption algorithm based on balanced pixel and chaotic map. *Hindawi Publishing Corporation: Mathematical Problems in Engineering*, 2014, pp. 1-7, 2014. DOI: 10.1155/2014/216048
- [16] Jiménez, M., Rider, J. and Alexander, P., Secure communication based on chaotic cipher and chaos synchronization. *Discontinuity, Nonlinearity and Complexity*, 1, pp. 57-68, 2012. DOI: 10.5890/DNC.2012.02.003
- [17] Zanin, M., Sevilla, J., Jaimes, R., García, J., Huerta, G. and Pisarchik, A.N., Synchronization attack to chaotic communication systems. *Discontinuity, Nonlinearity, and Complexity*, 2(4), pp. 333-343, 2013. DOI: 10.5890/DNC.2013.11.003.
- [18] Julián, M., Fredy, A. and Fabián, Ch., Security analysis of a WLAN network sample in Tunja, Boyacá, Colombia. *DYNA*, pp. 226-232, 2015. DOI: 10.15446/dyna.v82n189.43259
- [19] Carroll, T.L. and Pecora, L.M., Synchronizing chaotic circuits. *IEEE Transactions on Circuits and Systems*, pp. 453-456, 1991. DOI: 10.1109/31.75404
- [20] Ramírez, C.A., Masking information through synchronized chaotic systems, MSc. Thesis, Department of Mathematics, Universidad Nacional de Colombia, Bogotá, Colombia, 2011.
- [21] Schuster, H.G., Handbook of chaos control. New York: Wiley-VCH, 1999.

M. Jiménez-Rodríguez, has an BSc. degree in Computation Engineering, from the UdeG in 1999, a MSc. degree in Applied Computation in 2003, Universidad Central Martha Abreu, Cuba. In 2005, she obtained a Cisco Certified Network Associate certification, and in 2007, a Cisco Certified Academy Instructor certification. She was awarded a Dr. degree in Science and Technology at the Centro Universitario de los Lagos (CULagos), UdeG, in 2012. Currently, she is a professor in the Department of Technological Sciences at the Centro Universitario de la Ciénega (CUCiénega) and conducts investigation in the areas of Security Systems and Communications and Systems Elaboration, in addition to Applied Mathematics in Systems Development.
ORCID: orcid.org/0000-0002-4935-2731

M.G. González-Novoa, is a full-time professor, working at the Department of Basic Sciences of the UdeG, Ciénega. She has a MSc. degree in Applied Communications with a specialty in databases, awarded in 2005. Her area of specialization is object-oriented programming and software development, distributed systems, the application of algorithms, and data structure. She Participates in diverse investigation projects, is the author of various international and national publications, books and peer-review articles, all with reference to the line of investigation with which she collaborates: Elaboration of Security Systems and Communications. Currently she is working on the development of applications with technology for networks and security.
ORCID: orcid.org/0000-0002-1170-1238

J.C. Estrada-Gutiérrez, has an BSc. degree in Computational Engineering 2001, a MSc. degree in Applied Computation from the UdeG, in 2005, a PhD in Sciences from the UdeG in 2014, awarded in 2005 by Cisco CCNA (Cisco Certified Network Associate), and in 2007 received a certification as a Cisco Certified Academy Instructor. He is also a Candidate to be a National Instructor in the Mexican National System of National Investigators (SNICONACYT) 2014-2016. He is a professor at the Department of Technological Sciences, CUCiénega, and carries out research in the areas of Telecommunications, Applied Physics, and Biomedical Engineering.
ORCID: orcid.org/0000-0002-6727-3500

C. Acosta-Lúa, obtained his BSc.in Electronic Engineering from the Technological Institute of Morelia in 2001. He completed his MSc.degree in 2003 and Ph.D. in 2007 in Science in Electrical Engineering at CINVESTAV Guadalajara Unit. He participated in stays at INSA Lyon, France and DEWS Research Center in L'Aquila Italy. He carried out his Postdoctoral studies at DEWS Research Center in L' Aquila, Italy and the Centre for Research and Implementation of the Ford Motor Company. Since 2009, he has been an Associate Professor of Automatic Control at the University of Guadalajara. He is currently engaged in the development of nonlinear techniques for vehicle control and observers for nonlinear subsystems thereof.
ORCID: orcid.org/0000-0002-7398-2629.

O. Flores-Siordia, is a full-time professor in the Department of Technological Sciences at CUCiénega, UdeG. He has an BSc. degree in Chemical Engineering, a MSc. degree in Chemical Engineering, and a PhD in the Teaching Methodology. His areas of specialization include applied mathematics. He participates in the following lines of investigation: Elaboration of Security Systems and Communications, and collaborates in diverse research projects, and has various international and national publications, books and peer-reviewed articles, in reference to the development of applications with technology for networks and security.

ORCID: orcid.org/0000-0003-3611-4512



UNIVERSIDAD NACIONAL DE COLOMBIA

SEDE MEDELLÍN
FACULTAD DE MINAS

Área Curricular de Ingeniería
de Sistemas e Informática

Oferta de Posgrados

Especialización en Sistemas
Especialización en Mercados de Energía
Maestría en Ingeniería - Ingeniería de Sistemas
Doctorado en Ingeniería- Sistema e Informática

Mayor información:

E-mail: acsei_med@unal.edu.co
Teléfono: (57-4) 425 5365

Las TIC navegando hacia la sostenibilidad económica y ambiental en el transporte marítimo

Alfonso I. López-Díaz ^a, Miguel Ángel Gutiérrez-García ^a, Diego González-Aguilera ^b,
Alejandro Morales-Sánchez ^b & Jaime Ruiz-Martí ^c

^a Universidad Católica de Ávila, UCAV, Ávila, España. alfonso.lopez@ucavila.es, miguel.gutierrez@ucavila.es

^b Universidad de Salamanca, USAL, Salamanca, España. daguilera@usal.es, alejandro.morales@usal.es

^c Autoridad Portuaria de Santa Cruz de Tenerife, España. jaimer@puertosdetenerife.org

Received: January 27th, 2016. Received in revised form: March 8th, 2016. Accepted: April 4th, 2016.

Resumen

El Sistema Portuario Español pone en marcha, en abril de 2015, la plataforma web DUEPORT (Declaración Única de Escala Portuaria) con el fin de ser un observatorio de tráfico marítimo de buques, mercancías y residuos en tiempo real, erigiéndose como la medida más importante para luchar y prevenir la contaminación marina. En este artículo se realiza un análisis por separado de las necesidades del negocio marítimo actual y los instrumentos legales y económicos existentes en materia de prevención de la contaminación marina, mostrando la solución desarrollada por las Tecnologías de la Información y Comunicación (TIC) para dar respuesta a todas las necesidades de forma razonable. Las conclusiones de la investigación vienen a mostrar que las TIC completan el “puzzle” de los intereses del sector marítimo, capaces de crear un entorno sostenible entre los intereses del negocio marítimo y las necesidades de información ambiental de los buques a través de la aplicación de los instrumentos legales y económicos existentes.

Palabras clave: negocio marítimo, contaminación marina, tecnologías de la información y comunicación, buques, vía electrónica.

ICT sailing to economic and environmental sustainability in the maritime transport

Abstract

The Spanish Port System launched in April 2015, the web platform DUEPORT (Single Declaration of Port Scale) to be an observatory of marine vessel traffic, goods and waste in real time, establishing itself as the most important measure to fight and prevent marine pollution. In this article an economic existing in the prevention of marine pollution separate analysis of the needs of today's maritime business and legal instruments and is done by showing the solution developed by the Information Technology and Communication (ICT) to meet all reasonably needs. The research findings come to show that ICTs complete the "puzzle" of the interests of the maritime industry, able to create a sustainable environment between the interests of maritime business and environmental information needs of ships through the application existing legal and economic instruments.

Keywords: maritime business, marine pollution, information technology and communication, ship, electronically.

1. Introducción

El Sistema Portuario Español [1] pone en marcha, en abril de 2015, la plataforma web DUEPORT (Fig. 1) como punto de encuentro entre los intereses del negocio marítimo y las

formalidades de información derivadas de los requisitos legales, económicos y medioambientales de información de los buques [2].

How to cite: López-Díaz, A.I., Gutiérrez-García, M.A., González-Aguilera, D., Morales-Sánchez, A. y Ruiz-Martí, J., Las TIC navegando hacia la sostenibilidad económica y ambiental en el transporte marítimo. DYNA 83(197), pp. 188-194, 2016.

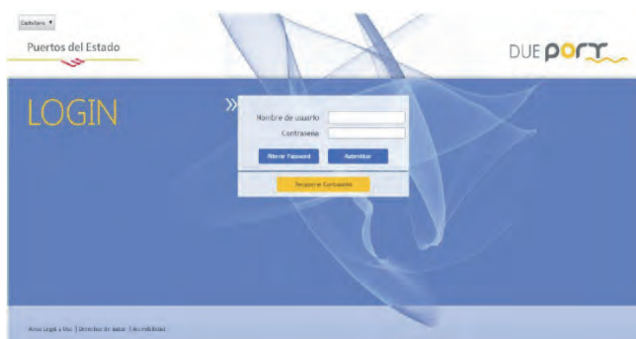


Figura 1. Pantalla de entrada al sistema (en proyecto) DUEPORT.
Fuente: Manual de Puertos del Estado.

A raíz de los desastres ecológicos de los buques petroleros Erika (1999)¹ y Prestige (2002)², se creó la Agencia Europea de Seguridad Marítima en el año 2002 (en adelante EMSA), organismo descentralizado de la UE, para reducir el riesgo de accidentes marítimos y luchar contra la contaminación marina. El proyecto de Ventanilla Única Nacional por parte de la EMSA, basadas en las Tecnologías de la Información y Comunicación (en adelante TIC), viene a establecer un sistema compartido de información, entre todos los estados europeos, de seguimiento de buques y mercancías en aras de eliminar los buques deficientes, aumentar la protección de los pasajeros y las tripulaciones, reducir el riesgo de contaminación del medio ambiente, y garantizar que aquellos buques con buenas prácticas medioambientales no se encuentren en desventaja comercial frente a los buques que toman atajos en materia de seguridad.

DUEPORT se presenta como la Ventanilla Única Española pero además se establece como trámite único compartido para el resto de administraciones españolas [3], evitando así burocracia y duplicidades. En el ámbito ambiental, los documentos en papel dejan paso a los datos digitales, con lo que se permite el tratamiento informático facilitando el control y seguimiento de la gestión de residuos de los buques y por tanto mejoras en la prevención de la contaminación marina por buques.

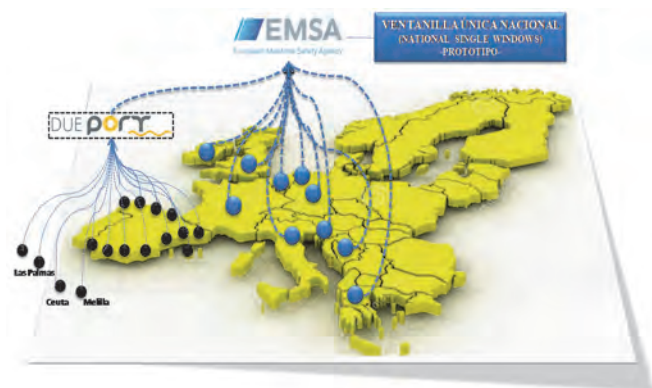


Figura 2. Escenario de los sistemas integrados (DUEPORT y EMSA) entre puertos y estado español con Europa.
Fuente: Elaboración propia

En la Fig. 2, se muestra el escenario de intercambio de comunicaciones electrónicas que se pretende implantar en toda Europa, apreciándose por un lado el proyecto español DUEPORT y por otro lado el proyecto Europeo de Ventanilla Nacional.

La implantación de la plataforma DUEPORT, abre el camino a los siguientes objetivos:

- Identificar las oportunidades de mejora que supondrá el uso de las TIC para el negocio marítimo, en cuanto a la agilidad y eficiencia en los trámites de entrada/estancia/salida de las escalas en los puertos españoles.
- Identificar las oportunidades de mejora que supondrá la plataforma DUEPORT para las administraciones públicas, con competencia en el tráfico marítimo internacional, para ejercer el control y seguimiento de forma más eficaz y eficiente.
- Valorar el grado de mejora que supone el sistema DUEPORT, en cuanto a la trazabilidad de las sustancias contaminantes producidas por las mercancías peligrosas y las basuras que genera un buque durante las escalas en puertos españoles y/o europeos.
- Identificar y describir los instrumentos legales y económicos en materia de prevención de la contaminación producida por buques, desde convenios internacionales hasta las directivas europeas y legislación nacional, mostrando la respuesta que tiene la solución informática para garantizar su cumplimiento.
- Simular el impacto económico coercitivo que tiene los instrumentos económicos de cara a controlar la gestión de los residuos producidos en los buques. Evidenciando el coste económico que supone para los buques un compromiso ambiental frente a buques que no lo tengan.
- Establecer líneas de mejoras por la implantación de TIC, en cuanto a la prevención de la contaminación de los buques, por el uso de datos en vez de documentos en papel [4,5].

2. El negocio marítimo

El transporte marítimo es referente del comercio internacional y de la economía mundial. Alrededor del 80% del comercio mundial, en volumen, y más del 70% en valor, se transporta por mar y transita por puertos de todo el mundo [6].

La constante incorporación de nuevos buques a un mercado con un exceso de oferta, junto con la debilidad de la economía mundial, ha ejercido una fuerte presión a la baja sobre los fletes. Así, los bajos precios del flete registrados en 2012 disminuyeron los ingresos de los transportistas hasta niveles iguales, o incluso inferiores, a los costes operacionales. Ante esta situación, los transportistas han puesto en marcha diferentes estrategias, entre ellas, medidas para reducir el consumo de combustible [7,8]. A pesar de que el tráfico marítimo crece, el sector se caracteriza por unas rentabilidades bajas debido al exceso de oferta, situándose en torno a un 8% de beneficio (ver Fig. 3).

¹ Erika, petrolero con bandera de conveniencia maltesa, que transportaba 30.000 toneladas de fuel pesado, que debido un mar de fuerza 8 a 10 y una edad de 25 años, provocó una marea negra en las costas de Bretaña.

² Prestige, petrolero con armador griego, sociedad en Liberia y bandera de Bahamas que transportaba fuel ruso y se hundió frente a las costas gallegas, ocasionando un vertido de crudo que provocó uno de los mayores desastres ecológicos de la historia de Europa.

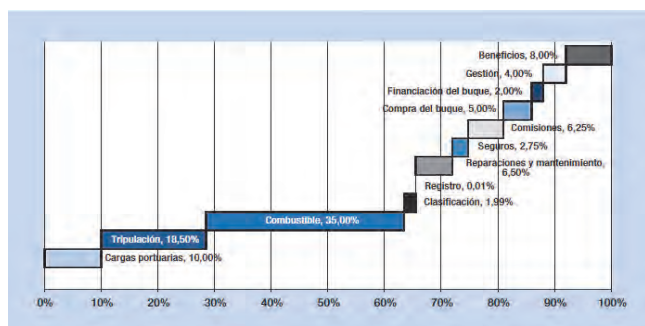


Figura 3. Porcentaje de gastos y beneficio final según los datos recibidos de un operador de buques en febrero de 2012. Datos provenientes de la Conferencia de las Naciones Unidas sobre Comercio y Desarrollo (UNCTAD).

Fuente: El Transporte Marítimo 2012.

3. Instrumentos legales y económicos para prevenir la contaminación por buques

Los instrumentos legales y económicos se presentan como herramientas efectivas y coercitivas. Buques con buena predisposición medioambiental aceptan la legislación e incluso se acogen a los incentivos económicos frente a buques con menor sensibilidad medioambiental que se encuentran con mercados más restringidos por su difícil acceso a los puertos europeos.

3.1. En cuanto a los instrumentos legales a nivel internacional

En la década de los 60, la Organización Marítima Internacional (en adelante OMI), organismo dependiente de la Organización de las Naciones Unidas (ONU), lideró el desarrollo de convenios para la seguridad de los buques, para simplificar la excesiva burocracia de entrada de buques en los puertos, para favorecer el comercio y para incorporar la variable ambiental dentro del negocio marítimo.

- En 1960 se aprobó el convenio para la seguridad de la vida humana en el mar o SOLAS³, denominación inglesa del convenio: "Safety Of Life At Sea", a raíz del accidente del RMS Titanic, que inspiró el desarrollo de normativa relativa a la seguridad de los buques. Unos de sus capítulos, el transporte de mercancías peligrosas, fue objeto del desarrollo del código IMDG⁴ (International Maritime Dangerous Goods) para facilitar la información sobre el tipo de mercancías peligrosas y contaminantes para el mar.
- En 1965 se aprobó el convenio para facilitar el Tráfico Marítimo Internacional o FAL 65, a resultados de un informe preparado por la industria naviera de la costa del Pacífico de los Estados Unidos, en colaboración con la School of World Business, y el San Francisco State College, de California, con fecha 1 de abril de 1959 y titulado "La marina mercante en un mar de papeles".
- Como resultado, se simplificaron y normalizaron los siguientes formularios: FAL 1: Declaración General,

FAL 2: Declaración de carga, FAL 3: Declaración de provisiones del buque, FAL 4: Declaración de efectos de la tripulación, FAL 5: Lista de la tripulación, FAL 6: Lista de pasajeros y FAL 7: Mercancías Peligrosas

- En 1973 y posterior modificación en 1978 se aprobó el convenio internacional para prevenir la contaminación por los buques o MARPOL 73/78 (abreviación de polución marina) [9], que estableció un conjunto de normativas con el objetivo de prevenir la contaminación por los buques. La estructura del convenio consta de una introducción, del texto y protocolo y una serie de anexos que contienen reglas que abarcan las diversas fuentes de contaminación por los buques.

3.2. En cuanto a los instrumentos legales a nivel de la Comunidad Europea

- La Directiva 2014/100/UE, de 28 de noviembre de 2014, por la que modifica la Directiva 2002/59/CE relativa al establecimiento de un sistema comunitario de seguimiento y de información sobre tráfico marítimo "SafeSeaNet", encomienda a la EMSA aprovechar la experiencia y seguir facilitando más la información e incluye la incorporación de técnicas geomaticas o de información geoespacial SIG (sistemas de información geográfica, en inglés GIS).
- La Directiva 2010/65/UE, de 20 de octubre, sobre las formalidades de información exigibles a los buques a su llegada o salida de los puertos de los estados miembros y por la que se deroga la Directiva 2002/6/CE, que tiene como objetivo simplificar y armonizar los procedimientos administrativos aplicados al transporte marítimo, mediante la generalización de la transmisión electrónica normalizada de datos y la racionalización de las formalidades informativas. Los estados miembros transmitirán a través de una ventanilla única, que será el punto en el que se comunique toda la información una sola vez y se ponga a disposición de las distintas autoridades competentes y los países de la UE. Los países de la UE deberán cerciorarse de que la información recibida de conformidad con las formalidades informativas esté disponible en su sistema SafeSeaNet nacional y pondrán a disposición de los demás países de la UE las partes pertinentes de esa información a través del sistema SafeSeaNet. Los países de la UE aceptarán los formularios normalizados desarrollados por el Convenio FAL.
- El desarrollo del proyecto europeo se está realizando mediante la creación de un prototipo de Ventanilla Única Nacional (National Single Windows), aprovechando la experiencia de SafeSeaNet, que aumenta las prestaciones y que reemplazará en un futuro a SafeSeaNet.
- La Directiva 2002/84/CE, de 5 de noviembre de 2002, modifica las Directivas relativas a la seguridad marítima y la prevención de la contaminación por los buques. El objetivo es mejorar la aplicación de la

³

https://es.wikipedia.org/wiki/Convenio_Internacional_para_la_Seguridad_de_la_Vida_Humana_en_el_Mar

⁴ <https://es.wikipedia.org/wiki/IMDG>

legislación comunitaria en el ámbito de la seguridad marítima, la protección del entorno marino y las condiciones de vida y trabajo a bordo de los buques.

- La Directiva 2000/59/CE, de 27 de noviembre de 2000 y su posterior modificación Directiva 2007/71/CE sobre instalaciones portuarias receptoras de desechos generados por buques y residuos de carga. La Comunidad Europea preocupada por la contaminación de los mares y las costas de los estados miembros, producida por las descargas de desechos y residuos de carga procedentes de los buques, establece la aplicación del convenio MARPOL 73/78.

3.3. En cuanto a los instrumentos legales a nivel de España

El 19 de noviembre de 2002 se hundió frente a las costas de Galicia el buque petrolero “Prestige”, ocasionando uno de los mayores desastres ecológicos de la historia de Europa y poniendo en evidencia a España por el desconocimiento del tráfico marítimo que navega por sus costas. En ese momento, la vía analógica en formato papel era el procedimiento establecido para facilitar la información en cada puerto.

- La orden FOM/3056/2002, en vigor el 5 de enero de 2003, que establece el procedimiento integrado de escalas de buques en los puertos de interés general, exige la implantación de las ventanillas únicas de escala en cada puerto que a su vez debían estar informando a la ventanilla única nacional, potenciándose la vía electrónica con una estructura básica de datos del buque.
- La orden FOM/1498/2014 [10], de 1 de agosto de 2014, que modifica el procedimiento integrado de escalas de buques en los puertos de interés general, establece la obligatoriedad electrónica de todos los datos de la declaración única de escalas (DUE), incorporando como novedad la transmisión electrónica de datos medioambientales en base al convenio MARPOL 73/78.

A su vez, la orden FOM/1498/2014 recoge la exigencia informativa del Real Decreto 1334/2012 de formalidades exigibles a los buques en los puertos de interés general (Trasposición de la Directiva 2010/65/CE); la exigencia informativa del Real Decreto 201/2012 del Sistema de Información de Tráfico Marítimo (Trasposición de la Directiva 2011/15/UE); y la exigencia informativa del Real Decreto 1084/2009 para instalaciones portuarias de desechos de los buques.

Por tanto la orden FOM/1498/2014 es la ley española que recoge las sinergias del resto de las normativas en cuanto a facilitar información de tráfico marítimo y es la fuente de inspiración para el proyecto DUEPORT.

3.5. En cuanto a los instrumentos económicos a nivel de España

- El Real Decreto Legislativo 2/2011 del texto refundido de la Ley de Puertos y de la Marina Mercante [11] establece una tarifa de recogida de desechos para los anexos MARPOL I y V, en sintonía con el Real Decreto 1084/2009 sobre instalaciones portuarias de desechos de buques.

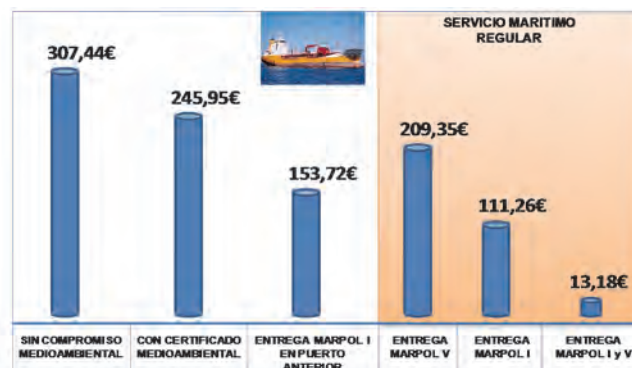


Figura 4. Cálculo de la tarifa de desechos para una escala de un buque petrolero de 121 m de eslora en un puerto español.

Fuente propia.

En la Fig. 4 puede apreciarse el impacto económico coercitivo que supone para un buque tipo (petrolero de 121 metros de eslora) que realice una escala en puerto español sin acreditar ningún compromiso medioambiental, se le aplica la tasa por escala más alta (en este caso de 307,44 €), si el buque puede acreditar un sistema de gestión ambiental certificado o entrega de residuos MARPOL I (residuos procedentes de hidrocarburos) en puerto anterior, puede beneficiarse de una bonificación del 20 al 50% de la tarifa. Si además, el buque realiza una ruta regular con varios puertos de forma regular, puede declarar a las administraciones marítimas esa ruta como servicio marítimo regular y presentar un plan de gestión de residuos para el MARPOL I y V (residuos procedentes de las basuras), obteniendo bonificaciones en la tarifa de desechos. En la Fig. 4, se muestra que se apuesta porque los buques de servicio regular tengan compromisos de gestión para los residuos MARPOL I y V conjuntamente, obteniendo con ello un precio simbólico de 13,18 € por escala.

4. El papel de las TIC

La Directiva 2010/65/UE puso rumbo a la simplificación y armonización de los procedimientos administrativos en el transporte marítimo europeo, mediante el uso de las TIC, estableciéndose la creación de Ventanillas Únicas Nacionales (National Single Windows)⁵ electrónicas bajo el concepto de la “e-Maritime”. Las TIC se erigen como el medio más sostenible entre los intereses del negocio marítimo y el cumplimiento de los distintos instrumentos legales y económicos, en particular los del ámbito ambiental.

El uso de la información electrónica e internet están cambiando el mundo. El transporte marítimo, no ajeno a estos cambios, utiliza para las notificaciones, declaraciones, certificados, solicitudes y órdenes de servicio, la gestión y almacenamiento en forma electrónica en lugar del formato papel [12-14]

España con la plataforma web DUEPORT, no solo dará cumplimiento a la comúnmente conocida Directiva sobre las formalidades informativas (Reporting Formalities Directive), sino que además se convierte en punto único de información para el resto de administraciones nacionales con competencia en el ámbito del transporte marítimo⁶.

⁵ <http://emsa.europa.eu/nsw.html>

⁶ Dirección General de la Marina Mercante, Salvamento Marítimo, Aduanas, Sanidad Exterior, Policía Nacional, Guardia Civil y la Armada Española.

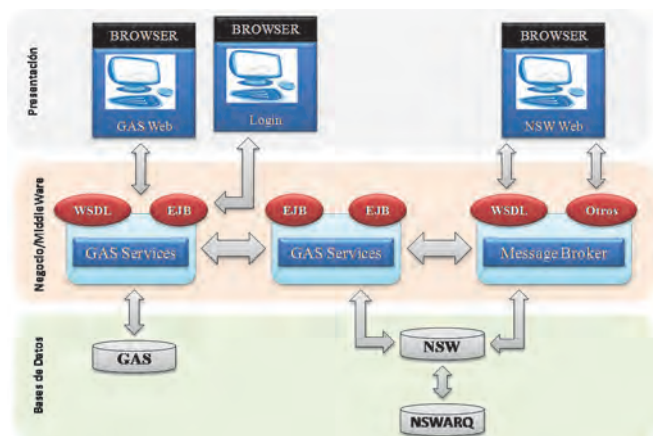


Figura 5. Estructura en capas del Sistema DUEPORT.

Fuente propia.

El nuevo sistema DUEPORT (NSW-ES) está compuesto por un conjunto de aplicaciones de negocio (responsables de la prestación de los servicios que pueden ser consumidos por aplicaciones externas o por las propias aplicaciones clientes del propio sistema DUEPORT), un conjunto de aplicaciones de presentación o GUI, cuya funcionalidad es proporcionar al usuario una forma fácil e intuitiva de cumplir con sus compromisos y obligaciones declarativos y consumir servicios de las aplicaciones de negocio. Por último, existe una base de datos, con la lógica separada en tres esquemas distintos: NSW (negocio), GAS (Gestión de Accesos y Seguridad) y NSWARQ (Archivo de datos de negocio).

La Fig. 5 muestra las aplicaciones existentes, estructuradas en un modelo de 3 capas: capa de presentación, capa de aplicación o tratamiento de datos y finalmente una capa de persistencia de datos.

La arquitectura del sistema DUEPORT, está basado en un modelo de 3 capas: capa de presentación, capa de aplicación o tratamiento de datos y finalmente una capa de persistencia de datos.

La capa de presentación tiene dos partes. Una interfaz de usuario tipo web (cliente ligero) cuya funcionalidad es la de proveer una capa de presentación y validación de datos básica. Esta interfaz es interactiva de manera que los usuarios (personas) pueden trabajar sobre ella directamente.

Adicionalmente tiene una segunda interfaz con otros sistemas externos a DUEPORT mediante ficheros o comunicación de datos estructurada. Por un lado permite el envío/recepción de ficheros EDI/EDIFACT⁷ o XML mediante un servidor FTP. En un siguiente nivel estaría la capa de aplicación (middleware), encargada del tratamiento de datos y donde se implementa toda la lógica de negocio del sistema. Aquí hay varios módulos:

- Un módulo encargado del tratamiento, validación y lógica necesaria para la comunicación con la interfaz web de usuario.
- Un módulo encargado de la interpretación y traducción de los ficheros en formato EDI/EDIFACT.

⁷ UN/EDIFACT United Nations/Electronic Data Interchange For Administration, Commerce and Transport (Intercambio electrónico de datos para la Administración, Comercio y Transporte) es un estándar de la Organización de las Naciones Unidas para el Intercambio electrónico de datos en el ámbito mundial. Existiendo subestándares para cada entorno de

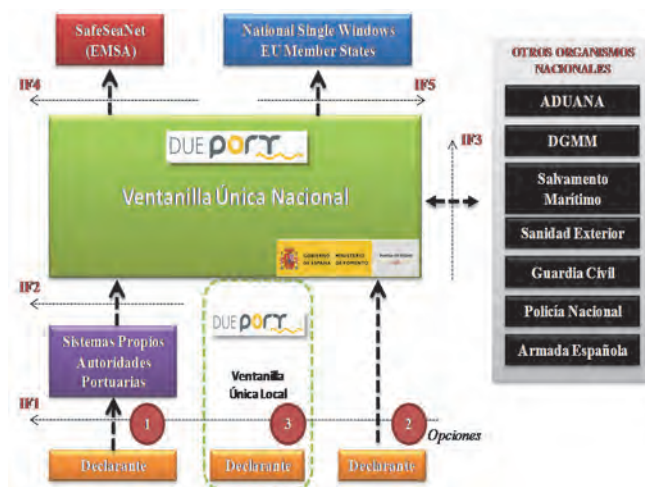


Figura 6. Mapa del Sistema DUEPORT.

Fuente Propia.

- Un módulo encargado de la recepción de los datos provenientes de ficheros EDI/EDIFACT y realizar las acciones solicitadas correspondientes a las funciones de cada fichero. Así mismo, este módulo se encarga de la entrega de datos al módulo de interpretación y traducción de ficheros EDI/EDIFACT para la generación de los ficheros de salida en el mismo formato, para los sistemas/usuarios externos.

Finalmente está la capa de persistencia de datos que consta de un sistema de base de datos relacional, donde se almacenan los datos correspondientes y necesarios para el funcionamiento del sistema.

En la Fig. 6 queda representado el mapa de sistemas que se pretende establecer. Los consignatarios de buques (declarantes) contarán con tres opciones: (i) operar con plataformas propias establecidas en algunos puertos españoles siendo los puertos responsables del envío posterior a DUEPORT en formato EDI; (ii) operar directamente en la plataforma DUEPORT Nacional; (iii) operar en la plataforma DUEPORT instalada en los servidores de cada puerto comunicándose por servicios web y facilitar la carga de trabajo a los sistemas nacionales.

La complejidad del mapa de sistemas viene por la necesidad de la interoperabilidad con los distintos sistemas existente. Las siglas IF del Mapa de Sistemas de la Fig. 6 muestran donde deben incorporarse herramientas de dinámica de sistemas para garantizar la correcta transmisión de la información.

5. Conclusiones

Por el carácter internacional del tráfico marítimo, la Comunidad Internacional tomó conciencia de la necesidad de aumentar la seguridad, prevenir la contaminación y facilitar la información de los buques que realizan escalas en los puertos para permitir el desarrollo del negocio marítimo en armonía con el respeto al medioambiente.

negocio (distribución, automoción, transporte, aduanero, etc) o para cada país. EDIFACT define las normas de reglas de sintaxis (ISO 9735), directorio de elementos de datos, directorio de elementos de datos compuestos, directorio de segmentos, directorio de mensajes y lista de códigos EDIFACT.

Así, en la década de los 60, la Organización de las Naciones Unidas creó la OMI para desarrollar y aprobar innumerables convenios internacionales, destacando los que son objeto de este trabajo: el convenio SOLAS (desarrollando el código IMDG para Mercancías Peligrosas) para la seguridad de los buques y la mercancía; el convenio FAL 65 para facilitar el tráfico marítimo normalizando la información; y el convenio MARPOL 73/78 para prevenir la contaminación por los buques.

Europa, tras las catástrofes ambientales de los buques petroleros ERIKA y PRESTIGE, creó en el año 2002 la Agencia Europea de Seguridad Marítima (EMSA) para garantizar un nivel elevado, uniforme y eficaz de seguridad marítima. Apoyado en los convenios internacionales y el uso de las TIC, la EMSA está impulsando un sistema comunitario de información de tráfico marítimo que favorezca el conocimiento y aumente el control de los buques que navegan por aguas europeas, y a su vez sea fuente de inspiración técnica de desarrollos legislativos.

En la actualidad, el proyecto piloto liderado por la EMSA para la implantación de una Ventanilla Única Nacional para el intercambio de información entre los estados Europeos y la plataforma web DUEPORT por parte de Puertos del Estado, son los referentes portuarios para entender el papel tan importante que están adquiriendo las TIC en el transporte y la seguridad marítimas. El desarrollo de la normativa presente y futura en materia de seguridad y contaminación de buques tendrá una referencia al sistema de información y comunicación establecido.

Desde una dimensión medioambiental, las TIC incorporan como dato la información relativa al convenio MARPOL 73/78 y código IMDG de mercancías peligrosas, obteniendo como resultado práctico:

- Posibilidad de aplicar las herramientas económicas de la tasa de desechos de una forma planificada y más eficiente.
- Posibilidad de autorizar y controlar el paso de aquellas mercancías consideradas contaminantes para el mar, de manera que las administraciones puedan planificar y realizar las inspecciones de los buques y sus mercancías.
- Posibilidad de planificar los servicios de recogida de desechos en tierra por parte de las empresas prestadoras del servicio.
- Posibilidad de simular, de forma teórica, como es la gestión de los residuos de un buque por su trazabilidad.
- Controlar y cierre del espacio marítimo europeo, permitiendo tomar decisiones derivadas del conocimiento.

5.1. Futuras líneas de investigación

1. Tomando como punto de partida la nueva plataforma DUEPORT, la ampliación de la funcionalidad debe pasar por incorporar herramientas de tratamiento de datos que facilitan la inteligencia de negocio (del inglés business intelligence), capaces de generar estrategias y aspectos relevantes, enfocados a generar conocimiento. Por ejemplo, realizar el seguimiento y vigilancia de todo el tráfico europeo en su conjunto.
2. Incorporación de los anexos VI “Prevención de la contaminación atmosférica” y VII “Manejo de lastres” a DUEPORT para el análisis de datos.

3. Los datos declarados del MARPOL I y V en cada puerto y en el sistema portuario español, permitirán realizar estudios económicos de viabilidad de las empresas que se dedican al servicio de recogida de MARPOL en tierra.

Referencias

- [1] Rodríguez, G., Gobierno electrónico: Hacia la modernización y transparencia de la gestión pública. Revista de Derecho, Universidad del Norte, [En línea]. 21, pp. 1-23, 2004. Disponible en: http://ciruelo.uninorte.edu.co/pdf/derecho/21/1_GOBIERNO%20EL%20ELECTRONICO_DERECCHO_No%2021.pdf
- [2] Pardo, L., Aplicación de las nuevas tecnologías en la administración pública. Revista de Contabilidad y Dirección, [En línea]. 13, pp. 105-126, 2011. Disponible en: http://www.accid.org/revista/documents/Aplicacion_de_las_nuevas_tecnologias_en_la_Administracion_Publica.pdf
- [3] Gascó, M., Los retos de la colaboración: ¿A qué, si no a eso, pretendemos hacer frente con la interoperabilidad? XV Congreso Internacional del CLAD sobre la Reforma del Estado y de las Administraciones Públicas. Santo Domingo, Rep. Dominicana, [En línea]. pp. 9-12 2010. Disponible en: <http://siare.clad.org/fulltext/0065701.pdf>
- [4] Barrero, D., Criado, I. y Ramilo, C., Política y Web 2.0. III Congreso Online del Observatorio para la Cibernética. [En línea]. Noviembre-Diciembre 2006. Disponible en: https://www.uam.es/personal_pdi/derecho/icriado/PDF/BarreroCriadoRamilo_Cibersociedad2006_Web2_0_Politica.pdf
- [5] Barrero, D. y Criado, I., Integrando la gestión del conocimiento en las administraciones públicas. Una solución desde la web semántica y los servicios web. VIII Jornadas sobre Tecnologías de la Información para la Modernización de las Administraciones Públicas (TECNIMAP 2004). Murcia, [En línea]. septiembre-octubre 2004. Disponible en: https://www.uam.es/personal_pdi/derecho/icriado/PDF/BarreroCriado_VIII TECNIMAP_2004.pdf
- [6] San Simón & Duch. Las tendencias del transporte marítimo. [En línea]. [Consultado en diciembre 24 de 2013]. Disponible en: <http://www.lsansimon.com/las-tendencias-del-transporte-maritimo/>
- [7] López-del Corral, O., López, A., Ortega, A. y Blanco, F., La importancia del transporte marítimo en la industria de la automoción. Dyna Ingeniería e Industria. [En línea]. 86(1), pp. 26-27. Disponible en: <http://www.revistadyna.com/busqueda/la-importancia-del-transporte-maritimo>
- [8] Ortega, A., López-del Corral, O., López, A., Puente, C., Gutiérrez, M. A. and Blanco, F., Maritime transport as a key element in the automotive industry. Journal of Maritime Research. [Online]. X(3), pp. 79-82, 2013. Available at: <http://www.jmr.unican.es/index.php/jmr/article/view/278/246>
- [9] Convenios Marítimos Internacionales. [En línea]. MARPOL 73/74. [Consultado en julio 8 de 2015]. Disponible en: <http://conveniosmaritimos.blogspot.com.es/2015/02/marpol-7378.html>
- [10] España. Orden FOM/1498/2014, de 1 de agosto, por la que se modifica la Orden FOM/1194/2011, de 29 de abril, por la que se regula el procedimiento integrado de escala de buques en los puertos de interés general. [En línea]. BOE, 11 de agosto de 2014 (núm. 194). Disponible en: <https://www.boe.es/boe/dias/2014/08/11/pdfs/BOE-A-2014-8624.pdf>
- [11] España. Real Decreto Legislativo 2/2011, de 5 de septiembre, por el que se aprueba el Texto Refundido de la Ley de Puertos del Estado y de la Marina Mercante. [En línea]. BOE 20 de octubre de 2011 (núm.253). Disponible en: <https://www.boe.es/boe/dias/2011/10/20/pdfs/BOE-A-2011-16467.pdf>
- [12] Gutierrez, M.A., López, A.I., Herráez, F. y Zubelzu, S., Aplicación para el estudio de implantación de un sistema no satelitario de identificación de buques. Revista Iberoamericana de Automática e Informática Industrial. 9, pp. 244-248, 2012. DOI: 10.1016/j.riai.2012.05.003
- [13] Fonseca, S.L., Castellanos, O.F. and Jiménez, C.N., Consideration for generating and implementing technological strategies. Ingeniería e

Investigación. [Online]. 32(2), pp. 83-88, 2012. Available at: <http://www.revistas.unal.edu.co/index.php/ingenv/article/view/31947/34566>

- [14] Cravero-Leal, A., Manzón J.N. and Trujillo, J., A business orientated approach to data warehouse development. Ingeniería e Investigación. [Online]. 33(1), pp. 59-65, 2013. Available at: <http://www.revistas.unal.edu.co/index.php/ingenv/article/view/37668>

A.I. López-Díaz, es Dr. por el Departamento de Ciencias y Técnicas de la Navegación y de la Construcción Naval de la Universidad de Cantabria, España, en 2007. Lic. en Marina Civil (Sección: Radioelectrónica Naval) por la Universidad de Cádiz, España, en 1992. Ing. Técnico de Telecomunicación por el Instituto de Empresa Universidad (IE-Universidad) en 2009. Oficial de la Marina Mercante Española en 2000. Autor de más de treinta y cinco artículos científicos en revistas nacionales e internacionales, varios libros de investigación y manuales universitarios. Profesor en la Facultad de Ciencias y Artes de la Universidad Católica de Ávila, España. ORCID: 0000-0001-5088-5802.

M.Á. Gutiérrez -García, es Ing. en Informática por la Universidad de Salamanca, España y Dr. por la Universidad Católica de Ávila, España. Profesor de Lenguajes y Sistemas Informáticos. Coordinador del Grado en Ingeniería de Sistemas de Información. Director del Máster en Ciberseguridad UCAV-Deloitte. ORCID: 0000-0001-8387-881X.

D. González-Aguilera, es graduado por la Universidad de Salamanca, España, en Ing. Topográfica, en 1999 y Geodesia y Cartografía en 2001, comenzó su labor investigadora como asistente del Instituto de Visión Computacional y Robótica (INRIA) de Grenoble, Francia, donde también elaboró su Tesis Doctoral, titulada "Reconstrucción 3D a partir de una sola vista" en 2005. Fruto de los resultados investigativos obtenidos en sus Tesis Doctoral, obtuvo un total de cuatro premios internacionales por la Sociedad Internacional de Fotogrametría y Sensor Remoto (ISPRS/ASPRS) y el Premio de Investigación 2015 de la Real Academia de Ingeniería. Autor de más de 100 artículos en revistas científicas de impacto y conferencias, actualmente es Profesor Titular de la Universidad de Salamanca, España y Director del Departamento de Ingeniería Cartográfica y del Terreno. ORCID: 0000-0002-8949-4216.

A. Morales-Sánchez, es graduado por la Universidad de Salamanca, España en Ing. Informática en 2000, Premio de Grado de Salamanca (Universidad de Salamanca) de Ingeniería Informática 2003 - 2004. Profesor asociado y miembro del Grupo de Investigación Robótica del Departamento de Informática y Automática de la Universidad de Salamanca desde el año 2005. Autor de varios artículos en revistas científicas de impacto. ORCID: 0000-0001-8331-0212

J. Ruiz-Martí, es Ing. Agrónomo en 1994, por la Universidad Politécnica de Valencia, España, recientemente he realizado el MSc. en Medio Ambiente y Desarrollo Sostenible por la Universidad de Ávila, España en 2016. Empezó a trabajar en una empresa de Biotecnología en Canarias con relaciones internacionales con Israel. Posteriormente entró a trabajar en un grupo Naviero como Adjunto a la Gerencia/Coordinador Operativo. Realizó en 1998 un MSc. de Gestión de Empresa (MBA) por la Universidad de la Laguna, España. Actualmente trabaja en la Autoridad Portuaria de Santa Cruz de Tenerife, en calidad de Jefe de División de Sostenibilidad realizando el MSc. en Negocio Marítimo y Gestión Portuaria en 2010, por la Universidad Internacional Menéndez Pelayo, España. ORCID: 0000-0003-1186-8537.



UNIVERSIDAD NACIONAL DE COLOMBIA

SEDE MEDELLÍN
FACULTAD DE MINAS

Área Curricular de Medio Ambiente

Oferta de Posgrados

Especialización en Aprovechamiento de
Recursos Hidráulicos
Especialización en Gestión Ambiental
Maestría en Ingeniería Recursos Hidráulicos
Maestría en Medio Ambiente y Desarrollo
Doctorado en Ingeniería - Recursos Hidráulicos
Doctorado Interinstitucional en Ciencias del Mar

Mayor información:

E-mail: acia_med@unal.edu.co
Teléfono: (57-4) 425 5105

Prioritization of TQM practices in Brazilian medical device SMEs using Analytical Hierarchy Process (AHP)

Glauco Henrique de Sousa-Mendes ^a, Eduardo Gomes-Salgado ^b & Bruno Enrico Moro-Ferrari ^a

^a Department of Production Engineering, Federal University of São Carlos, São Carlos, Brazil. glauco@dep.ufscar.br

^b Exact Sciences Institute, Federal University of Alfenas, Alfenas, Brazil. eduardosalgado@unifal-mg.edu.br

^a Department of Production Engineering, Federal University of São Carlos, São Carlos, Brazil. bemferrari@gmail.com

Received: July 30th, 2015. Received in revised form: April 8th, 2016. Accepted: April 24th, 2016.

Abstract

The purpose of this article is to prioritize Total Quality Management (TQM) practices and examine their relative importance in TQM implementation in medical device small and medium enterprises (SME) using the Analytic Hierarchy Process (AHP) method. Mathematical modeling was the research method adopted in this article. The choice of AHP was based on the need to acquire knowledge from different companies and prioritize the most important TQM practices. Some implications are highlighted. Two strategic factors (Quality Management System and Quality Planning) and one tactical factor (Process Management) are strongly correlated and indicated by experts as having the major importance. Finally, The AHP method developed in the present study was shown to be useful for prioritizing practices relating to the implementation of TQM in medical device companies. The results can provide useful information to medical device managers to work upon the main practices in order to improve TQM performance.

Keywords: Total Quality Management, Analytical Hierarchy Process (AHP), medical device companies and small and medium enterprises.

Priorización de las prácticas de GCT en empresas de dispositivos médicos brasileños utilizando el Analytical Hierarchy Process (AHP)

Resumen

El propósito de este artículo es dar prioridad a Gestión de Calidad Total (GCT) prácticas y examinar su importancia relativa en la implementación de GCT en pequeñas y medianas empresas (PYME) de la industria de dispositivos médicos, utilizando el método de Analytical Hierarchy Process (AHP). Modelación matemática fue el método de investigación adoptado. La elección de AHP también se basó en la necesidad de adquirir conocimientos de diferentes empresas. Algunas implicaciones se pueden destacar. Dos factores estratégicos (Sistema de Gestión de Calidad y Planificación de la Calidad) y un factor táctico (Gestión de procesos) están fuertemente correlacionados y indicados por los expertos como tener la mayor importancia. Por último, el método AHP desarrollado en el presente estudio ha demostrado ser útil para dar prioridad a las prácticas relativas a la aplicación de la GCT en las empresas de dispositivos médicos. Los resultados pueden ser información útil para los administradores de dispositivos médicos que trabajan en las principales prácticas con el fin de mejorar el rendimiento de la GCT.

Palabras clave: Gestión de la Calidad Total, Analytical Hierarchy Process (AHP), las empresas de dispositivos médicos y las pequeñas y medianas empresas.

1. Introduction

Total Quality Management (TQM) is a management approach, aimed at incorporating quality thinking in organizations, that became very popular in the 1980s [1]. TQM

encompasses a set of principles, methods and practices that emphasize customer satisfaction, leadership, total employee involvement, continuous improvement, strategic and systematic approach, process management, factual approach to decision making, performance measurement and win-win supplier

How to cite: de Sousa-Mendes, G.H., Gomes-Salgado, E. and Moro-Ferrari, B.E., Prioritization of TQM practices in Brazilian medical device SMEs using Analytical Hierarchy Process (AHP). DYNA 83(197), pp. 195-203, 2016

relationships [2]. TQM is a managerial model for improving an organization's competitive capacity and maximizing its participation in the market [3,4]. Several authors [5-7] have indicated that implementation of TQM improves organizational performance.

The literature includes many practices associated with TQM, but little attention has been paid to prioritizing such practices [8, 9]. Thus, it becomes critical to determine, among all such practices, those that most contribute to successful TQM implementation. Prioritization of these practices would ensure a better fit between the allocation of resources and the results achieved. Moreover, according to Bayazita and Karpak [10], although some companies successfully implement these practices, a considerable number fail in attempting to do so, which therefore justifies this research.

The Analytical Hierarchy Process (AHP) is a multi-criteria decision-making method for decisions that involve complexity and subjectivity [11]. AHP helps to solve problems that require prioritization, since it allows the complexity of unstructured decisions to be broken down into simpler components, thereby arranging a hierarchical order for each component of the analysis [11].

This paper presents the results of applying the Analytic Hierarchy Process (AHP) to evaluate TQM practices that are most important in TQM implementation in small and medium-sized Brazilian medical device manufacturers. A set of 19 TQM practices that most influence TQM implementation were identified from a literature review. Then, ten TQM experts were interviewed in order to establish the relative importance of these practices.

Brazilian companies in the medical device industry have made notable efforts towards implementing quality management in order to meet Brazilian standards and the requirements of international markets. In addition, Brazil is the largest producer and market for medical devices in South America and the country is both a major producer and a major importer in the region. This industry still only represents a small proportion of Brazil's economy, but has many opportunities for growth in the coming years.

A number of similarities can be drawn between Brazilian and Colombian medical device industries; for example, their dependence on imports, growth perspective, predominance of small and medium enterprises (SMEs) in this industry, and the need to improve the quality of medical devices. Therefore, the contribution made by this paper is to systematize practices adopted in implementing TQM and classifying them in terms of priority. Furthermore, companies in the medical device industry would be able to improve their TQM implementation processes, further emphasizing practices that most contribute to the success of TQM implementation.

After this introduction, Section two presents a theoretical review considering practices involved in TQM literature and AHP method. The research method is described in Section 3, after which the results are presented, followed by a discussion of the key findings. Finally, conclusions are drawn.

2. Literature Review

2.1. Critical factors in implementing TQM

As indicated by several researchers [12,13], there are a number of benefits to implementing TQM: the improvement

of a competitive advantage, higher profitability, fewer losses and less waste, higher productivity, improvement of product quality, employee skill development, greater customer satisfaction, higher motivation, operational improvements, continuous improvement of processes, and shorter waiting times.

Rockart [15] argued that the critical factors are good practices that should receive managerial attention given that they influence the performance of the organization positively. In addition, studies focusing on TQM practices have increased in recent years [14]. In this paper, practice can be understood as a principle, a method, a process or a tool that has a positive impact on the success of TQM implementation. From the TQM literature, 19 critical practices (critical factors) were identified and grouped into three categories: strategic factors, tactical factors and operational factors.

2.1.1. Strategic factors

Strategic factors represent the basis of TQM [8], and TQM principles are values that should be incorporated into organizational culture. The importance of these values has been confirmed by quality award prizes, such as the Malcolm Baldrige National Award, European Quality Award and even ISO 9001 standards [16]. Five TQM practices were identified as strategic factors: top management commitment (leadership), quality management system, continuous improvement, quality culture and quality planning.

The first strategic factor is the top management commitment (leadership), which is fundamental for creating an organizational environment in which people can become involved in quality objectives and customer satisfaction. Pheng and Jasmine [17] argue that if top management is not committed to TQM, it is impossible to achieve success in TQM implementation. Thus, several authors confirm that top management commitment is a strategic factor in TQM implementation [18-21].

Another strategic factor is the quality management system (QMS) that consists of a set of elements adopted by the organization in order to guarantee that its product and services meet customer requirements [22,24]. A QMS integrates organizational structure, processes, procedures, practices, roles, responsibilities, documents and resources within the organization and intends to provide a framework to direct and control an organization in terms of quality [23].

The third strategic factor is continuous improvement, which is a set of recurring activities that are carried out in order to enhance performance of processes, products, and services. According to Bessant and Francis [25] continuous improvement is an organizational process that guides an organization toward incremental improvements and requires involvement by everyone in the organization.

Creating a culture of quality is the fourth strategic factor. Organizational culture encompasses values, principles, attitudes and standards that influence organization members, defining their behavior. Consequently, quality culture encompasses an organizational behavior that aims to achieve customer satisfaction and quality improvements in whole organization [6,22]. According to Deming [26], a quality culture guarantees that an organization is involved in quality.

At the very least, the fifth strategic factor is quality planning,

which pertains to the Juran Trilogy [24] and focuses on setting quality objectives and specifying necessary operational processes and related resources to fulfil the quality objectives [23]. In this study, quality planning is stressed as it provides direction for TQM implementation.

2.1.2. Tactical factors

Tactical factors support the strategic factors and they sustain TQM implementation [8]. The following tactical factors were considered: employee involvement, reward systems and recognition, training and education, quality information systems, supplier relationships and process management.

First, employee involvement has been cited as one of the most important factors in TQM implementation [9,22,26]. It is essential to satisfy different stakeholders in an organization, and in order to achieve employee involvement, empowerment and delegation should be supported. Consequently, employees will develop knowledge, pursue quality goals and take on their responsibilities in ensuring quality [6,22,27].

The second factor is the reward system and recognition. Crosby [28] considers recognition as one of the most important factors in TQM. Moreover, since rewards and recognition practices increase employee motivation (which will influence customer satisfaction), it is necessary to create mechanisms for rewarding employees for their good contributions to the quality improvements [29].

Training and education are the third tactical factor. They provide qualifications and skills for employees in principles, methods and tools applied to quality management. Thus, training and education contribute to increasing employee skills and organizational competence [21,27,30].

The quality information system (QIS) is essential for implementing TQM because it generates data and information that are needed to make good decisions. Decision-making based on facts and data is one the TQM principles, but this is achieved when the information is correct and available. Therefore, the quality information system is another critical factor in TQM [22,27]. In this study, QIS is the fourth tactical factor.

Mutually beneficial supplier relations (fifth factor) refer to the relationship between an organization and its suppliers. In TQM, a good and strong relationship can increase productivity and stimulate cooperation and collaboration in solving quality problems [23]. Thus, a win-win relationship needs to be established between the organization and its suppliers [30].

Process management is the last tactical factor considered in this study. The process concept induces a new way of thinking about organization and introduces customer satisfaction in all organization process [24,26]. Process management is a systematic approach to identifying key processes, ensuring a flow of activities and resources that adds value for the customer [6,22].

2.1.3. Operational factors

Operational factors are mainly methods and techniques. They are usually more visible and their results should appear

in the short term [6,8]. The considered operational factors were techniques and tools for analyzing and problem-solving, teamwork, process mapping, visual management tools, benchmarking, statistical process control, suggestion plans, and customer surveys.

Problem-solving techniques are the first operational factor. They are simple tools used for solving quality-related operational problems [31]. The PDCA cycle, the seven basic quality tools, the seven managerial tools, and other techniques applied in continuous improvement activities are examples of problem-solving techniques [9].

The second factor is teamwork. TQM stimulates teamwork in order to overcome individual limitations and provide synergy in a group seeking quality improvements. Teamwork is considered essential for implementing TQM [7] and its promotion in an organization will demand cooperation, communication and a focus on the common goals [22,27].

The third factor comprises process mapping, which is the technique used to identify, improve and document the activities and resources that make up a business process. Several process modeling techniques can be applied for this purpose [32]. Process mapping is one of the techniques employed in the process management approach.

Visual management tools are the fourth operational factor. Quality information needs to be promptly available and visible to everyone, thereby facilitating decision-making [33].

The fifth factor is benchmarking and it is a method where an organization can compare its performance with other references (competitors, for example). It is also important for establishing higher standards and achievement improvements through learning from external experiences [6,22,27]. In this article, benchmarking is understood as an operational process developed by an organization in order to measure its performance and make improvements.

The Statistical Process Control (SPC) comprises a set of statistical techniques used for monitoring and improving production processes. It is essential in quality control, evaluating and promoting stability and the capacity of processes [4]. Thus, SPC is the fifth operational factor.

Suggestion plan is a structured strategy that seeks to stimulate the employee involvement. Through this practice, new ideas are creating and they can be implemented, which can generate improvement opportunities [31]. Suggestion plans constitute the seventh operational factor considered in this study.

The last factor is market research, which is any organized effort to gather information about customer needs. Capturing the voice of the customer and translating it into products specifications and process standards may represent a competitive advantage [7].

2.2. Analytical Hierarchy Process (AHP)

Shimizu [34] argues that the Analytical Hierarchy Process (AHP) is a structured technique for organizing and analyzing complex decisions that has received a lot of attention from academics and managers. This is confirmed by Wallenius et al. [35].

AHP aims to solve problems that arise when several criteria

are used in a single assessment. It breaks down the complexity of unstructured problems into simpler components. Therefore, it can be used to classify the importance of the factors considered in a decision-making process [36].

According to Saaty [36], AHP is carried out through pairwise comparisons and relies on the judgments of experts to derive priority scales. Moreover, expert judgments are taken into consideration by means of binary comparisons of factors in order to form a priority scale.

A judgment or comparison is a numerical representation of a relationship between two elements. Using comparisons in pairs in each judgment, the experts indicate which element of the pair is more important and then attribute weights in order to differentiate the relative importance of each element [11]. The judgments, inserted in the comparisons matrices, are often based on the fundamental scale of absolute numbers that is specific to AHP, as shown in Table 1.

According to Salgado et al. [38], the weights of the criteria and the priorities for the alternatives are obtained by means of judgments provided by experts. These judgments are inserted into a matrix (A) so as to pair the data. In this method, the importance values for the criteria are obtained using the eigenvector w of the comparison matrix A, as shown in Equation 1, where λ_{max} is the maximum eigenvalue.

$$A w = \lambda_{max} . W \quad (1)$$

The eigenvalue is a measurement of the consistency of the comparison matrix. In a matrix that is 100% consistent, a relationship of transitivity, $a_{ij} = a_{ik} a_{kj}$, is supported for all comparisons. In this case, $\lambda = n$, where n is the order of the matrix. However, in the AHP method, to ensure the consistency of the pairwise comparison matrix, the consistency judgment must be checked for the appropriate value μ , as showed in Equation 2:

$$\mu = (\lambda - n)/(n - 1) \quad (2)$$

For a comparison matrix that is 100% consistent, $m = 0$ because $\hat{1} \gg n$. Furthermore, Saaty [36] proposed the Consistency Ratio, which is a comparison between the Consistency Index (CI) and Random Consistency Index (RI). The Random Consistency Index depends on the matrix size (n), as shown in the Table 2. Hence, the Consistency Ratio (CR) formula is shown in Equation 3:

$$CR = CI / RI \quad (3)$$

Table 1.
Criterion comparison scale

Intensity of importance	Explanation
1	Both elements are of equal importance
3	One element is moderately more important than the other.
5	One element is strongly more important than the other
7	One element is very strongly more important than the other
9	One element is extremely more important than the other.
2, 4, 6, 8	Intermediate values between the adjacent opinions.

Source: [37].

Table 2.

Random Consistency Index (RI)										
n	1	2	3	4	5	6	7	8	9	10
RI	0,00	0,00	0,58	0,90	1,12	1,24	1,32	1,41	1,45	1,49

Source: [36].

For a comparison matrix that is 100% consistent, $\mu = 0$ because $\lambda = n$. Saaty [11] recommends that for values of μ greater than 0.20, the comparisons should be reviewed. Review of the comparisons is a systematic procedure for improving the decision-making process. The largest eigenvalue λ_{max} is estimated by means of the arithmetic mean of the elements of the consistency vector.

The number of comparisons is a quantitative attribute and is the inverse of the performance of the application of the multi-criteria method. The larger the number of comparisons, the greater the effort needed to achieve decision-making. More comparisons may consume more resources [38]. Thus, x , the number of comparisons needed for completing a comparison matrix, can be obtained using Equation 4.

$$x = n(n - 1)/2 \quad (4)$$

To conclude, AHP has been employed successfully in different contexts. Sarathy [9] applied it in assessing quality management practices in the construction industry. In the same way, Salgado et al. [38] has used AHP to identify critical activities in new product development. In the case of this article, AHP will be used to prioritize TQM practices in the medical device industry.

3. Method

The research method was based on the works of Talib et al. [8], Lewis et al. [12], Kumar et al. [13], Khanna et al. [39] and Souza Junior et al. [40]. It follows the stages described below:

- **Review of the literature:** The literature on the main topics covered in this study was reviewed in order to identify and systematize the most critical factors in implementing TQM. Concerning AHP, the aim was to describe the procedures for applying AHP in complex problems.
- **Hierarchy of factors:** After the identification of TQM practices, they were divided into three categories (strategic factors, tactical factors, and operational factors). These categories and practices are shown in Fig. 1. There are three decision-making levels considered in this study. The first level represents the objective of the study; the second level of prioritization represents the categories of practices (criteria); and lastly, the third level represents the prioritization of TQM practices (sub criteria).
- **Construction of binary comparison matrices:** Four matrices were designed: a general comparison between the factors in the second level of periodization (strategic, tactical and operational) and one matrix per category. Simplified spreadsheets were used to make it easier for the experts to fill them in.
- **Selection of experts:** In the selection of the experts, criteria such as the field of activity, academic and

professional experience, and availability to attend interviews were taken into consideration. Following these criteria, ten experts took part in this study. Four were academics in the field of quality management and have experience in developing projects in medical device companies. Six were TQM managers in medical device companies and they held the responsibility for implementing TQM.

- Application: Interviews were conducted in which pairwise comparisons were made based on the matrices that had been made for this purpose. The results from the judgments were aggregated using the principle of Aggregation of Individual Priorities (AIP) [41]. AIP is used when a group of individuals have no association or common goals, and, as such, tend to act according to their preferences, values and goals. It is noteworthy that experts are from different companies, and that the choice for the AIP was made because it is necessary to keep this individual analysis. Furthermore, the arithmetic mean was used to aggregate the judgments.
- Analysis of the results: Finally, the results of prioritization process were analyzed.

4. Findings and Discussion

At the second level of prioritization, the interviewees were asked to make a pairwise comparison of the three categories of factors (criteria). Table 3 shows the values after normalization (sum of the elements in each column of the judgment matrices with subsequent division of each element of these matrices by the sum of the values of the respective column), along with the local priorities among these criteria. The consistency ratio (CR) index was calculated, and this indicated that there was consistency between the pairs of criteria. The upper limit for the consistency index is 20% and the maximum result obtained in this study was 4.0%, which means that the data can be considered consistent [11].

Table 3 shows the priorities among the three categories of factors: strategic factors (0.6370), tactical factors (0.2583) and operational factors (0.1047). These results are in accordance with the studies developed by Talib et al. [8] and

Table 3.

Relative importance among the three factor categories

Factors	Strategic	Tactical	Operational	Local Priorities
Strategic	1	3	5	63.70%
Tactical		1	3	25.83%
Operational			1	10.47%

CR = 4.0% and largest eigenvector $\lambda = 3.039$

Source: The authors

Table 4.

Relative importance among the strategic factors

	TMC	QMS	CI	CQ	QP	Local priorities
TMC	1	3/2	2	8/7	5/3	0.2698
QMS		1	5/3	2/3	7/5	0.1929
CI			1	1/2	7/6	0.1350
CQ				1	2	0.2663
QP					1	0.1360

CR = 0.4% and largest eigenvalue $\lambda = 5.020$

Top Management Commitment (TMC), Quality Management System (QMS), Continuous Improvement (CI), Culture of Quality (CQ) and Quality Planning (QP)

Source: The authors

Souza Junior et al. [40]. Therefore, strategic factors are considered more important than other factors, which mean that the success of TQM implementation depends on the emphasis made on such factors.

In the following, the experts made judgments relating to TQM practices within each category (sub criteria). The results are presented in Tables 3 to 5. As mentioned before, the local priorities and consistency ratios were also calculated for each category of factors.

Table 4 shows TQM practices classified as strategic factors. The experts considered “top management commitment” (0.2698) as the most important strategic factor, and the development of a “culture of quality” is considered the second most important practice in this category (0.2663) with a score that is very close to the first factor.

These two practices emphasize that successful TQC implementation depends on the ability to align individual and collective attitudes in seeking customer satisfaction. Moreover, this alignment should be supported by the top management commitment with the quality. The importance, the examples, and resources provided by top management and driven to improving quality will contribute to creating a quality culture, putting quality at the center of organizational culture.

This is truly important for SMEs, where the leadership has a great influence on management and performance. Company owners participated directly in activities and had the final say in quality issues. Although the findings are similar to those found by Talib et al. [8], they differ from the result found by Lewis, Pun and Lalla [19], for whom “top management commitment” and “culture of quality” occupies the 10th and 13th places, respectively, in terms of importance.

The quality management system (QMS) is the third most important strategic factor (0.1929). Particularly in Brazilian medical device companies, three main norms and standards have been used as references for structuring QMS, namely:

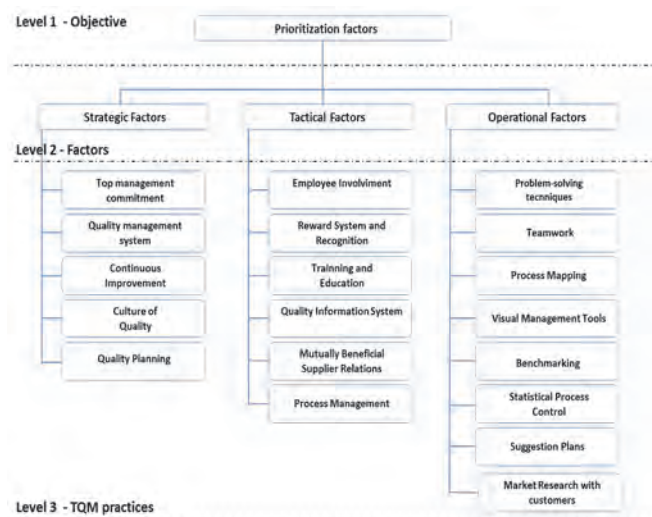


Figure 1. Structure of the hierarchical model proposed.

Source: The authors

ISO 9001:2008, ISO 13485:2004, and Good Manufacturing Practices (GMP). Certifying their QMS according these standards has been encouraged in the medical device industry due to customer demands (some international clients for example) and Brazilian legislation (all medical device companies should at least have GMPs implemented). Although QMS is only the third factor, its implementation influenced the first two factors, given that the role of leadership is one of the main requirements of these standards.

Concerning the tactical factors, the experts highlight three factors as being the most important in implementing the TQM: “employee involvement” (0.2862), “process management” (0.2163), and “training and education” (0.1833). Together, they account for 68.58% of the relative importance of this category (Table 5). These findings coincide with the eight principles of quality, and agree with the results presented by Sarathy [9]. However, in terms of prioritization, they are not similar to studies by Talib et al. [8] and Souza Junior et al. [40].

Getting people involved in quality and providing them with training and education are fundamental in TQM. Although these findings are widely recognized, they represent challenges for SMEs due to their lack of resources and skills. The lack of resources in SMEs compared to larger enterprises may contribute to favor approaches that are more informal in SMEs. As such, this situation reinforces the need for investment in motivating and training employees. While these two practices affirm the human side of quality highlighting the need for employee engagement, process management is a method that seeks to map and standardize work processes. Quality standards (ISO 9001 and ISO 13485), which are common in medical device companies, have process management as a main element.

Table 6 shows the priorities given to TQM operational practices. Observing the local priority values, it can be concluded that there is no single method or tool that really stands out in relation to the others. “Statistical Process control” (0.1905) is considered the most important practice, but it is closely followed by the following factors: “Problem-solving Techniques” (0.1585), “Teamwork” (0.1523), and “Market Research with customer” (0.1497). As expected, there is a relationship between the most important operational factors with the most important tactical factors. Teamwork and Problem-solving Techniques provide the basis for training and involving employees. The values for other practices are shown in Table 6.

Table 5.
Relative importance among the tactical factors

Factors	EI	RS	TE	QIS	SR	PM	Local priorities
EI	1	8/3	2	9/5	5/2	3/2	0.2862
RS		1	½	1	9/7	4/9	0.1026
TE			1	2	2	7/9	0.1833
QIS				1	6/5	5/7	0.1234
SR					1	1/3	0.0882
PM						1	0.2163

CR = 1.2% and largest eigenvalue $\lambda = 6.077$

Employee Involvement (EI), Reward System and Recognition (RS), Training and Education (TE), Quality information System (QIS), Mutually Beneficial Supplier Relations (SR) and Process management (PM).

Source: The authors

Table 6.
Relative importance among the operational factors

Factors	PST	TW	PM	VMT	BM	SPC	SP	MR	Local priorities
PST	1	5/4	1	2	3	2/3	5/2	1	0.1585
TW		1	3/2	5/3	3	1	2	3/4	0.1523
PM			1	9/5	7/3	2/3	9/5	5/6	0.1310
VMT				1	3/2	3/7	1	4/7	0.0821
BM					1	¼	3/5	3/8	0.0533
SPC						1	2	4/3	0.1905
SP							1	5/7	0.0825
MR								1	0.1497

CR = 1.2% and largest eigenvalue $\lambda = 6.077$

Problem-Solving Techniques (PST), Teamwork (TW), Process Mapping (PM), Visual management Tools (VMT), Benchmarking (BM), Statistical Process Control (SPC), Suggestion Plans (SP) and Market research with customers (MR)

Source: The authors

The findings shed light on the role of technical capabilities in supporting the TQM implementation. In addition, they stress the need to improve employee skills in using quality tools, especially those applied in solving problems and controlling processes. Empirical research developed by Oprime et al [31] has also established a positive relationship between technical capabilities and success in TQM implementation.

In a further analysis, the overall priority of the criteria and sub criteria (i.e., considering the relative importance of each practice) was calculated considering the 19 investigated TQM practices. Table 7 presents these results. The column “Geral priority” was calculated using the values presented in the “Local Priority Column,” however for each specific practice, it was necessary to multiply for its related categories. For example, the value of local priority related with the practice “TMC” (0.2698) was multiplied for the value of local priority category “Strategic factors” (0.630).

Table 7 highlights that “top management commitment” (0.1719), “culture of quality” (0.1606) and “quality management system” (0.1229) are the three main practices indicated by the experts. These should therefore be emphasized in implementing TQM in SMEs in the medical device sector. On the other hand, the practices with the lowest relative importance belong to the operational factors category: “visual management tool” (0.0086), “suggestion plans” (0.0086), and “benchmarking” (0.0056).

The results shown in Table 7 are strongly influenced by the weights given to the categories placed in level 2. For this reason, all of the strategic factors are more important than the tactical and operational factors, and the tactical factors are more important than the operational factors. This reinforces that a successful implementation of TQM depends on a managerial approach that involves leadership, commitment, and integration of quality improvement into the management system.

Table 7
Overall priorities for the 19 TQM implementation practices

Level	Factor categories and TQM practices	Local priorities		Overall priorities		
		Local Priority	Local Ranking	Geral Priority	%	Geral Ranking
Level 2	Strategic factors	0.6370		0.6370	63.70	
Level 3	Top Management Commitment (TMC)	0.2698	1	0.1719	17.19	1
	Quality Management System (QMS)	0.1929	3	0.1229	12.29	3
	Continuous Improvement (CI)	0.1350	5	0.0860	8.60	5
	Culture of Quality (CQ)	0.2663	2	0.1696	16.96	2
	Quality Planning (QP)	0.1360	4	0.0866	8.66	4
Level 2	Tactical factors	0.2583		0.2583	25.83	
Level 3	Employee Involvement (EI)	0.2862	1	0.0739	7.39	6
	Reward System and Recognition (RS)	0.1026	5	0.0265	2.65	10
	Training and Education (TE)	0.1833	3	0.0473	4.73	8
	Quality information System (QIS)	0.1234	4	0.0319	3.19	9
	Mutually Beneficial Supplier Relations (SR)	0.0882	6	0.0228	2.28	11
	Process management (PM)	0.2163	2	0.0559	5.59	7
Level 2	Operational factors	0.1047		0.1047	10.47	
Level 3	Problem-Solving Techniques (PST)	0.1585	2	0.0166	1.66	13
	Teamwork (TW)	0.1523	3	0.0159	1.59	14
	Process Mapping (PM)	0.1310	5	0.0137	1.37	16
	Visual management Tools (VMT)	0.0821	7	0.0086	0.86	17
	Benchmarking (BM)	0.0533	8	0.0056	0.56	19
	Statistical Process Control (SPC)	0.1905	1	0.0199	1.99	12
	Suggestion Plans (SP)	0.0825	6	0.0086	0.86	18
	Market research with customers (MR)	0.1497	4	0.0157	1.57	15
Level 2	Sum of criteria (factor categories)			1	100.00	
Level 3	Sum of sub criteria (TQM practices)			1	100.00	

Source: The author

5. Conclusion

The aim of this study was to investigate and categorize the TQM practices and prioritize them on the basis of their relative importance for TQM implementation in the Brazilian medical device SMEs.

As SMEs suffer with the lack of resources and cannot allocate equal amounts of efforts or resources to each TQM practice, the findings of this study are very important in providing insights for medical device SMEs as they can evaluate their current practices and re-allocate reasonable resources to improve their TQM performance. Besides medical device SMEs, managers from other health organizations (i.e., hospital or health public systems) can also use these findings for selecting and evaluating their own suppliers.

From the TQM literature, 19 TQM practices were identified and they were grouped into three categories of factors (strategic, tactical, and operational). Subsequently, ten experts were consulted and, with the aid of the AHP (Analytical Hierarchy Process) method, it was possible to prioritize these practices. The TQM practices identified cover

a wide range of activities from strategic, to tactical, to operational factors. Therefore, medical device SMEs should focus on the factors and practices that these experts considered of highest priority for TQM implementation.

Considering the results of this study, it is possible to present some managerial implications. One first managerial implication of this study is that the results indicate that TQM implementation is complex. Its success depends on strong commitment by top management, on the capacity to incorporate quality values in the organizational culture, and on the need to create involvement among collaborators (the most important practice at tactical level) in planning, controlling, and quality improvement activities. Although the three factor categories are important, TQM will not be successful unless human and political dimensions are also considered.

Another important managerial implication is the importance given to structuring a Quality Management System. Two strategic factors (Quality Management System and Quality Planning) and one tactical factor (Process Management) are strongly correlated and they were pointed out as having a major importance. QMS implementation

needs to be supported by an organizational structure, technical procedures, and commitment.

In conclusion, the leadership of top management and the creation of strong organizational values emphasizing quality are the main drivers for the organizational transformation promoted by TQM. Specifically, in the case of the medical device companies, the QMS is not the ultimate purpose or the main factor ensuring success in implementing quality. For the experts, the factors (top management commitment and culture of quality) are requisites for implementing the QMS as well as the TQM.

Finally, the AHP method developed in the present study was shown to be useful for prioritizing practices relating to TQM implementation in medical device companies. It can be useful to decision-makers as a guideline for implementing TQM and in evaluating the effectiveness of their current TQM practices.

Even though this study provides implications for academics and for medical device companies, it also presents some limitations. One of them is that its results are limited in scope because it relies on the judgments of only 10 experts. The second limitation is that research on critical factors in implementing TQM is not new and that many of the results found in the present study have been pointed out previously. However, it is important because the discussion on TQM practices is contextualized for the small and medium medical device companies, which have their own organizational and structural characteristics. Due to the similarity between Brazilian and Colombian medical device companies, the results can help TQM managers from both countries.

Future research could expand the number of experts including more TQM managers from medical device companies. Another proposal is to validate the criteria and sub-criteria proposed in this research using different research approaches, such as case studies or surveys. This would help to provide complementary information to develop appropriate strategies for improving the success of TQM in medical device companies.

References


- [1] Savolainen, T., Towards a new workplace culture: Development strategies for employer-employee relations. *Journal of Workplace Learning*, 12(8), pp. 318-326, 2000. DOI: 10.1108/13665620010355566.
- [2] Ross, J.E., *Total quality management: Text and cases readings*. Kogan Page, London. 1994.
- [3] Escanciano, C. and Rodríguez, F.J.I., Quality management and integrated total quality in spanish mining: Results of an empirical study. *DYNA*, 79(171), pp. 167-174, 2012.
- [4] Chen, J.-K. and Chen, I.-S., TQM measurement model for the biotechnology industry in Taiwan. *Expert Systems with Applications*, 36(5), pp. 8789-8798, 2009. DOI: 10.1016/j.eswa.2008.11.013.
- [5] Brah, S., Wong J.L. and Rao, M., TQM and business performance in the service sector: A Singapore study. *International Journal of Operations and Production Management*, 20(11), pp. 1293-1312, 2000. DOI: 10.1108/01443570010348262
- [6] Salaheldin, S.I., Critical success factors for TQM implementation and their impact on performance of SMEs. *International Journal of Productivity and Performance Management*, 58(3), pp. 215-237, 2009. DOI: 10.1108/17410400910938832
- [7] Yusuf, Y., Gunasekaran, A. and Guo, D., Implementation of TQM in China and organization performance: an empirical investigation. *Total Quality Management*, 18(5), pp. 509-530, 2007. DOI: 10.1080/14783360701239982
- [8] Talib, F., Rahman, Z. and Qureshi, M.N., Prioritizing the practices of total quality management: An analytic hierarchy process analysis for the service industries. *Total Quality Management & Business Excellence*, 22(12), pp. 1331-1351, 2011. DOI: 10.1080/14783363.2011.625192
- [9] Sarathy, P.S., TQM practice in real-estate industry using AHP. *Quality & Quantity*, 47, pp. 2049-2063, 2013. DOI: 10.1007/s11135-011-9641-8
- [10] Bayazita, O. and Karpak, B., An analytical network process-based framework for successful total quality management (TQM): An assessment of Turkish manufacturing industry readiness. *International Journal of Production Economics*, 105(1), pp. 79-96, 2007. DOI: 10.1016/j.ijpe.2005.12.009
- [11] Saaty, T.L., *The analytic hierarchy process*. New York, NY: Pergamon Press. 1988.
- [12] Lewis, W.G., Pun, K.F. and Lalla, T.R.M., An AHP-based study of TQM benefits in ISO 9001 certified SMEs in Trinidad and Tobago. *The TQM Magazine*, 17(6), pp. 558-572, 2005. DOI: 10.1108/09544780510627651
- [13] Kumar, R., Garg, D. and Garg, T.K., TQM success factors in North Indian manufacturing and service industries. *The TQM Journal*, 23(1), pp. 36-46, 2011. DOI: 10.1108/17542731111097470
- [14] Harrington, H.J., Voehl, F. and Wiggin, H., Applying TQM to the construction industry. *The TQM Journal*, 24(4), pp. 352-362, 2012. DOI: 10.1108/17542731211247373
- [15] Rockart, J., Chief executives define their own information needs. *Harvard Business Review*, 57(2), pp.81-93. 1979.
- [16] Martínez-Lorente, A.R. and Martínez-Costa, M., ISO9000 and TQM: Substitutes or complementaries? An empirical study in industrial companies. *International Journal of Quality and Reliability Management*, 21(3), pp. 260-276, 2004. DOI: 10.1108/02656710410522711
- [17] Pheng, L.S. and Jasmine, A.T., Implementing total quality management in construction firms. *Journal of management in Engineering*, 20(1), pp. 1-9, 2004. DOI: 10.1061/(ASCE)0742-597X
- [18] Oropesa-Vento, M., García-Alcaraz, J.L., Rivera, L. and Manotas, D.F., Effects of management commitment and organization of work teams on the benefits of Kaizen: Planning stage. *DYNA* 82(191), pp. 76-84, 2015. DOI: 10.15446/dyna.v82n191.51157
- [19] Lewis, W.G., Pun, K.F. and Lalla, T.R.M., Exploring soft versus hard factors for TQM implementation in small and medium-sized enterprises. *International Journal of Productivity and Performance Management*, 55(7), pp. 539-554, 2006. DOI: 10.1108/17410400610702142
- [20] Awan, M.U., Raouf, A., Ahmad, N. and Sparks, L., Total quality management in developing countries: A case of pharmaceutical wholesale distribution in Pakistan. *International Journal of Pharmaceutical and Healthcare Marketing*, 3(4), pp. 363-380, 2009. DOI: 10.1108/17506120911006056
- [21] Mallur, S.B. and Hiregoudar, N., A survey of TQM practices in North Karnataka manufacturing SMEs: An empirical evaluation. *Proceedings of the World Congress on Engineering*, London. 2010.
- [22] Sila, I. and Ebrahimpour, M., An investigation of the total quality management survey based research published between 1989 and 2000: A literature review. *International Journal of Quality and Reliability Management*, 19(7), pp. 902-970, 2003. DOI: 10.1108/02656710210434801
- [23] ISO 9000:2008, *Quality Management Systems: Requirements*, 2nd ed., Technical Committee/subcommittee TC 176/SC 1; ISO Standards. 2008.
- [24] Juran, J. and Gryna, F., *Quality Control Handbook*. McGraw-Hill. New York. 1999.
- [25] Bessant, J. and Francis, D., Developing strategic improvement capability. *International Journal of Operations & Production Management*, 19(11), pp. 1106-1119, 1999. DOI: 10.1108/01443579910291032
- [26] Deming, W.E., *Out of the crisis*, MIT press. 1986.
- [27] Kim-Soon, N. and Jantan, M., Quality management practices in Malaysia: Perceived advancement in Quality Management System and Business Performance, *IEEE ICMIT Conference*, Singapore. 2010. DOI: 10.1109/ICMIT.2010.5492699
- [28] Crosby, P., *Quality is Free*, New York: McGraw-Hill. 1979.
- [29] Mehra, S., Hoffman, J.M. and Sirias, D., TQM as a management strategy for the next millennia. *International Journal of Operations &*

- Production Management, 21(5/6), pp. 855-876, 2001. DOI: 10.1108/01443570110390534
- [30] Motwani, J., Critical factors and performance measures of TQM. The TQM Magazine, 13(4), pp. 292-300, 2001. DOI: 10.1108/13683040010362300
- [31] Oprime, P.C., Mendes, G.H.S. and Pimenta, M.C., Continuous improvement: Critical factors in Brazilian industrial companies, International Journal of Productivity and Performance Management, 61(1), pp. 69-92, 2012. DOI: 10.1108/17410401211187516
- [32] Ould, M.A., Business process: Modeling and analysis for re-engineering and improvement. New York: Wiley. 1995.
- [33] Martínez-Lorente, A.R., Sánchez-Rodríguez, C. and Dewhurst, F.W., The effect of information technologies on TQM: An initial analysis. International Journal of Production Economics, 89(1), pp.77-79, 2004. DOI: 10.1016/j.ijpe.2003.06.001
- [34] Shimizu, T., Decisão nas organizações. São Paulo: Editora Atlas. 2010
- [35] Wallenius, J., Dyer, J.S., Fishburn, P.C., Steuer, R.E., Zionts, S. and Déb, K., Multiple criteria decision making, multi attribute utility theory: recent accomplishments and what lies ahead. Management Science, 54(7), pp. 1336-1349, 2008.
- [36] Saaty T.L., Decision making with the analytic hierarchy process. International Journal of Services Sciences, 1(1), pp. 83-98, 2008. DOI: 10.1504/IJSSCI.2008.017590
- [37] Saaty, T.L., How to make a decision the analytic hierarchy process. European Journal of Operational Research, 48, pp. 9-26, 2009. DOI: 10.1016/0377-2217(90)90057-1
- [38] Salgado, E.G., Salomon, V.A.P. and Mello, C.H.P., Analytic hierarchy prioritization of new product development activities for electronics manufacturing. International Journal of Production Research, 50(17), pp. 4860-4866, 2012. DOI: 10.1080/00207543.2012.65797
- [39] Khanna, H.K., Sharma, D.D. and Laroia, S.C., Identifying and ranking critical success factors for implementation of total quality management in the Indian manufacturing industry using TOPSIS. Asian Journal on Quality, 12(1), pp. 124-138, 2011. DOI: 10.1108/15982681111140598
- [40] Junior, C.A., Salgado, E.G., Neves, F.O. and Alvarenga, A.D., Priorização das práticas de TQM na indústria farmacêutica. Revista Espacios, 35(4), pp. 7, 2014.
- [41] Forman, E. and Peniwati, K., Aggregating individual judgments and priorities with the analytic hierarchy process. European Journal of Operational Research, 108(1), pp. 165-169, 1998. DOI: 10.1016/S0377-2217(97)00244-0

G.H. de Sousa-Mendes is an Adjunct Professor in the Department of Production Engineering at the Federal University of São Carlos, Brazil. He completed his MSc. degree and PhD in Production Engineering at the Federal University of São Carlos, Brazil. He has carried out postdoctoral studies at the University of São Paulo, Brazil, focusing on Product-Service System development. He has written a number of books in Quality Management in Brazil, and his research interests center on quality management, product-service system, new product development and innovation.
ORCID: 0000-0001- 8086-3321

E. Gomes-Salgado has a BSc. degree in Production Engineering with emphasis on Mechanical Engineering in 2005, from the Federal University of Itajubá, Brazil; a MSc. Degree in Production Engineering in 2008 at the Federal University of Itajubá; and a PhD. in Mechanical Engineering in 2011, from São Paulo State University (UNESP), Brazil. He has carried out postdoctoral studies at the University of Glasgow (UK), focusing on New Product Development, and technical studies at the University of Minho, Portugal focusing on Quality Management. Since 2009, he has been an Adjunct Professor at the Federal University of Alfenas (UNIFAL-MG). During ICPR22, he received the Early Career Mentoring Award from IFPR. He has experience in Industrial Engineering, mainly in the following areas: New Product Development, innovation management, MCDM, Total Quality Management and ISO 9001.
ORCID: 0000- 0002-8940-4014

B.E. Moro-Ferrari has a BSc. degree in Production Engineering from the Federal University of São Carlos, Brazil and is a MSc. Degree student in Engineering at São Paulo University, Brazil. The author's research interests center on quality management and innovation.
ORCID: 0000-0001--7190-1744.



UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS

**Área Curricular de Ingeniería Administrativa e
Ingeniería Industrial**

Oferta de Posgrados

**Especialización en Gestión Empresarial
Especialización en Ingeniería Financiera
Maestría en Ingeniería Administrativa
Maestría en Ingeniería Industrial
Doctorado en Ingeniería - Industria y Organizaciones**

Mayor información:

E-mail: acia_med@unal.edu.co
Teléfono: (57-4) 425 52 02

Strategic guidelines for supply chain coordination in healthcare and a mathematical model as a proposed mechanism for the measurement of coordination effects

Delio Alexander Balcázar-Camacho ^a, Cesar Amílcar López-Bello ^b & Wilson Adarme-Jaimes ^c

^aFacultad de Ingeniería, Universidad Nacional de Colombia, Bogotá, Colombia. dabalcazar@unal.edu.co

^bFacultad de Ingeniería, Universidad Distrital Francisco José de Caldas y Universidad de la Sabana, Bogotá, Colombia. clopezb@udistrital.edu.co

^cFacultad de Ingeniería, Universidad Nacional de Colombia, Bogotá, Colombia. wadarme@unal.edu.co

Received: February 4th, 2016. Received in revised form: April 1st, 2016. Accepted: April 20th, 2016

Abstract

Supply chain (SC) management implies a large number of interactions and a great deal of cooperation between the organizations that make up the supply chain in order to fulfill the purposes of logistics processes such as supply, distribution and warehousing. Much research has been carried out to illustrate the advantages of coordinating these processes at different stages of supply chain planning. However, these studies appear to be separate efforts. This study proposes methodological guidelines for the coordinated planning of the supply chain in the context of the healthcare services, considering the global benefits to its members by using primary information collected at several health service institutions. The proposal also includes a mixed-integer programming model to show how organizational coordination contributes to a reduction of the overall costs throughout the chain.

Keywords: supply chain, coordination, medical supply chain, medical operations management, healthcare supply chain, mixed-integer programming.

Lineamientos estratégicos para coordinación en la cadena de suministro de medicamentos y propuesta de un modelo matemático para medir los efectos de la coordinación

Resumen

La gestión de la cadena de abastecimiento implica la articulación intra e inter organizacional de las áreas funcionales participantes en los procesos de aprovisionamiento, almacenamiento y distribución a través de las organizaciones que la componen. Diversas investigaciones se han realizado mostrando las ventajas de la coordinación de estos procesos en diferentes etapas de la planeación de la cadena de abastecimiento, sin embargo, parecen ser enfoques separados. Este artículo propone pautas metodológicas para la planeación coordinada de la cadena de suministro en el contexto de la prestación de servicios de salud considerando beneficios globales para sus integrantes a través de información primaria recolectada en varias instituciones prestadoras de servicios de salud, proponiendo también un modelo de programación lineal para mostrar los efectos de la coordinación en la disminución de los costos globales de la cadena.

Palabras clave: coordinación; logística; cadena de abastecimiento; medicamentos; gestión de operaciones; programación lineal.

1. Introduction

In line with the progress made in economic and productive systems, operations management has turned into

a set of increasingly complex tasks, which imply the articulation of organizational processes according to the goals and expectations of the supply chain directors and stakeholders.

How to cite: Balcázar-Camacho, D.A., López-Bello, C.A. and Adarme-Jaimes, W., Strategic guidelines for supply chain coordination in healthcare and a mathematical model as a proposed mechanism for the measurement of coordination effects. DYNA 83(197), pp. 204-212, 2016

Industries went from focusing on mass production to focusing on customization. Operations strategies were developed such as just-in-time systems, total-quality management, flexible manufacturing, lean Systems, process outsourcing, and supply-chain management systems, among others [1]. Each of these strategies has been addressed from a planning and practice perspective, yielding operations management models and control schemes that make use of mathematical techniques.

Operations management has evolved from a set of separate activities into a single integrated application within the management of supply chains (SC). SC can be defined as the set of suppliers, producers, and distributors that interact in order to deliver a product or to provide customer service, including the customer as a part of the chain, and also contemplating the stakeholders that are affected or involved in the operations of the members of the chain [2]. SC has been addressed using various approaches, such as risk management, agile SC, lean SC, Flexible SC, global SC and regional SC, among others [3]. Different models have been proposed considering isolated activities or different functions of the SC; however, such models appear to be fragmented efforts [4].

Researchers have mainly focused on generating models and solutions aimed at providing strategies and models for the goods industry. More recently, their attention has shifted towards the service sector; however, there is a lack of research results in activities derived from the healthcare service in countries like Colombia.

In Latin America and the Caribbean, the service sector accounts for about 60% of both GDP (gross domestic product) and the total employment pool in the region. Although the service sector's share has increased, its growth remains low when compared to other sectors of the economy [5]. In Colombia, the service sector presented the largest reported increase in revenue in 2013, with a growth of 38.4% over the previous year [6]. With the development of the sector, new opportunities arise to improve the perception of the services provided and therefore improve productivity.

The service sector and its importance in the economy lies in its potential for regional development and job-vacancy creation. This sector generates around 50% of GDP in low-income countries and, usually, its growth is linked to the country's development [7]. Within the service sector, health, as one of its main components, exhibits "special features from the normative and practical context that directly affect the quality of the service provided to patients, namely it is not enough to guarantee the procedures for the detection and diagnosis of particular disease but it is also necessary to ensure the timely availability and good condition of drug stocks, which may represent a critical factor for the proper treatment of patients" [8].

In today's competitive environment, several medical organizations have taken the necessary actions to ensure they are providing the best care at the lowest possible cost [9]. Traditionally, the pharmaceutical industry has focused on the development of new medicines, on promoting alternative sales strategies, and on modern marketing activities. More recently, the industry's attention has turned to SC optimization as a means of delivering added value [10]. SC management within the health-care sector has become ever

more complex due to the integration of several regulation entities (i.e. governmental agencies) as well as the monopolization of markets and the requirements of patients.

This article shows the results of a study conducted by the "Society, Economy and Productivity" (SEPRO) research group at Universidad Nacional de Colombia. The study was funded by Colciencias within the scope of a project entitled "A methodological proposal for the definition of policies, negotiation rules, and coordination activities applied to oncological medicine supply management in Colombia". The present paper provides a set of guidelines to achieve the coordination of agents in the SC as well as a mathematical model that serves as a mechanism to verify the impact of agents' coordination. The project's primary information, obtained from various institutions providing healthcare services in Bogotá, was collected and analyzed to identify potential structures or strategic guidelines for SC coordination. Subsequently, a mixed-integer programming model is proposed to illustrate the potential benefits of coordination using SC global cost as a metric to quantify improvements.

Section 2 contains a review of relevant research literature. Section 3 gathers a set of guidelines for the coordination of agents in SC. The proposed model to verify the potential effects of coordination is presented in section 5. The results that measure the impact of coordination and its associated effects appear in Section 6. Finally, Section 7 presents various points for further discussion.

2. Literature review

Commonly, two types of structures for SC management are suggested depending on the degree of control that every SC agent has over the others, namely a centralized structure and a non-centralized structure. When each member of the chain carries out its planning individually, regardless of other actors, local optima do not always contribute to achieving the overall objectives of the SC. Instead, coordinated SC planning has shown to have a positive impact on some performance metrics [11] such as delivery times, production costs, customer service perception [12]. Additionally, coordinated SC planning reduces the overall SC costs [13]. Such coordination involves not only the implementation of enterprise resource planning systems (ERP) and its associated links [14], but also the joint planning processes and the development of synchronized forecasts [15]. Coordination requires breaking down communication barriers and overcoming the reluctance to share information, it also requires aligning the different interests and management styles. All these requirements allow coordination to move towards an integrated management of chains that ultimately replaces any other functional process management approach [16]. Coordination is reflected when the members of a SC jointly minimize operating costs and share profits after production planning policies and programming operations have been deployed [17].

Typically, the literature distinguishes several strategies to achieve coordination of operations in an SC (e.g., purchasing policies, bulk discounts, inventory management, and lot sizes). However, few of these approaches have been applied to the sector of pharmaceutical SC (Medicines and medical

supplies) in Latin American countries.

The structure of pharmaceutical SC includes one or more of the following actors: primary manufacturing industries, secondary manufacturing industries, warehouses and distribution centers, wholesalers and retailers (i.e., hospitals) [10]. Pharmaceutical SC management is highly complex and sensitive since any change in the level of customer service is unacceptable and so demands great responsibility to ensure that the right drug reaches the right patient at the right time in the right condition [18]. Pharmaceutical SC management has to deal not only with the upstream information flow, such as market demand, but also with the downstream flow of technical information that supports demand creation [19]. Additionally, there is the challenge of reconciling the freedom of health professionals in writing their prescriptions and favoring the preferences of the hospital's purchase department when offered bulk discounts [20]. Another challenge lies in the high levels of uncertainty associated to pharmaceutical demand along with the short life cycle of products due to their physical-chemical characteristics [21].

From the perspective of logistics activities, it is common to find proposals in the literature dealing with 1) overall planning of the SC in the pharmaceutical/medical industry focused on facility location; 2) chains of two or more levels presenting the case of inventory management between wholesalers, distributors and healthcare service provider institutions (IPS); and 3) optimal order quantity focused on the application of volume discounts, determination of the number of orders, and reorder point policies.

In terms of the number of studies in the current literature concerning SC in the health sector, researchers have focused on the purchasing process and supply management, distribution and production of pharmaceutical products, health logistics, and research and development of new products [22]. Most of the work conducted in science, within pharmaceutical and medical SC management, refers to the overall planning and focuses on developing better solutions for large-scale challenges; however, it is necessary to develop practical models that can quickly solve the problems associated to the pharmaceutical SC. Such models should be easy to implement and should not omit the real aspects of the problem [23]. From the perspective of SC operations and management, research in the health sector focuses on generating strategies and objectives of operations, planning, scheduling, and control of services [24].

Regarding recent studies about SC in health care, Narayana, Kumar Pati & Vrat [22] conducted an extensive review of the recent literature that addresses the issue of managing pharmaceutical supply chains (PSC), showing the growing interest of researchers in using information technology, analysis and process optimization. Susarla & Karimipresent presented a model of mixed integer linear programming for global planning multi-period, multi-level supply chains within the pharmaceutical industry, also contemplating purchasing, production and distribution processes [23]. Uthayakumar & Priyan developed a model that includes the integrated planning of two levels of the SC to optimize inventory, also considering the customer service level and showing the possibility to reduce inventory costs while maintaining service levels near 100% [18]. Sousa, Shah & Papageorgiou proposed a model for the Global

Optimization of SC in the pharmaceutical industry, considering production factors and investment capacity for facility location [25]. Kelle et al. proposed policies for reorder points and security levels with space restrictions in the pharmaceutical industry SC from the perspective of a hospital [20]. Masoumi, Yu & Nagurney developed a multiproduct model for optimal allocation of resources for the production of vaccines and medicines in the design of the SC [26].

In Colombia, it is only recently that have researchers become interested in the SC of the healthcare sector. By consulting specialized databases, it is possible to corroborate that there are only a few research and scientific publications dealing with the mechanisms that allow a visualization of the effects of supply chain coordination in the context of healthcare service provision. Castrellón, Torres & Adarme proposed a model for reducing the costs of logistics operations in the distribution of medicines in Colombia [27]. Zamora presented a model for risk management in the procurement process of oncologic drugs [28]. Balcázar presented a methodology for the coordination of agents in oncology medicines SC [8].

Based on the content of the studies above, it is possible to identify a couple of gaps in the development of the research on SC coordination in healthcare, namely a lack of guidelines intended to improve coordination, and the corresponding mathematical models to measure the impact of such coordination. The following section presents various guidelines that can be used to coordinate processes within an SC.

3. Guidelines for coordination

SC coordination implies the articulated and harmonized actions of the organizations involved in the particular SC. These actions should promote value creation and the reduction of losses. To achieve coordination of logistic processes, high levels of standardization and communication are necessary to facilitate trust between the chain members (Fig.1).

It is possible to identify several strategic guidelines that can be used to achieve coordination in SC planning. Some of the main strategic guidelines, identified through our fieldwork with multiple healthcare institutions, are included in the following paragraphs.

The coordination process depends on the implementation and the proper use of information technology as well as the use of suitable systems for warehouse management (WMS). Real-time information on inventory levels and availability of resources (for the handling and transporting medicines) facilitates tactical and operational planning.

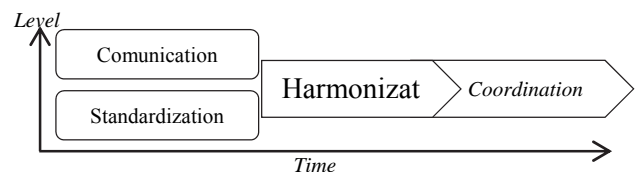


Figure 1. Road to coordination
Source: The Authors.

The deployment strategy for coordination begins with a diagnostic phase. This initial phase is intended to identify the information needs of the processes that must be integrated, the characteristics of the tangible and intangible elements of the participants and also to determine the team members that generate joint operation plans. These members are responsible for conducting the necessary logistical coordination during the planning horizon (Fig.2). Thus, it is necessary to determine various additional aspects, such as the structure of the communication channels, the main generators of information, the power levels for decision-making, the command channels, the response times, and the responsible members in charge of each process together with their substituting members (when required). Once the instructions and manuals have been settled, this information should mitigate the effects of plausible contingencies as well as of any deviation from the plans. These actions contribute to having minor adverse effects and delays in operations.

The equipment for SC planning must be composed of delegates from all the organizations (that participate in the process of harmonization and coordination) and professionals of all organizational levels, who know the details of each process in the logistic cycle (Fig. 3).

Planning for the coordination of activities in the SC must help identify redundant and concordant issues both in the operation and in the planning horizons of all organizations. Such planning should also match every activity according to the business strategy of each business unit. The redundant activities must be eliminated by considering which particular activity has the lesser impact on value creation, i.e., duplicate efforts can cause delays and become non-value adding activities. Therefore, in each process of the logistic cycle, activities should be checked in detail, e.g., inspections, checks, temporary storage, use of packaging and secondary packaging, signage and marking of products, among others.

In an ideal scenario, a high degree of coordination planning (for the supply, storage and distribution of all organizations of the chain) allows us to identify the impact of coordination on the logistics network and to determine the best location of inventories based on aspects such as demand, availability of resources and capacity of production and storage. It is possible to obtain advantages in the overall SC planning by using a midpoint in the chain as a reference (either manufacturing or distribution) for the exchange of information. Such a midpoint can then be used as a centralized decision-making spot or as an inventory centralized management spot, allowing better decision making and achieving global cost reduction [12,28].

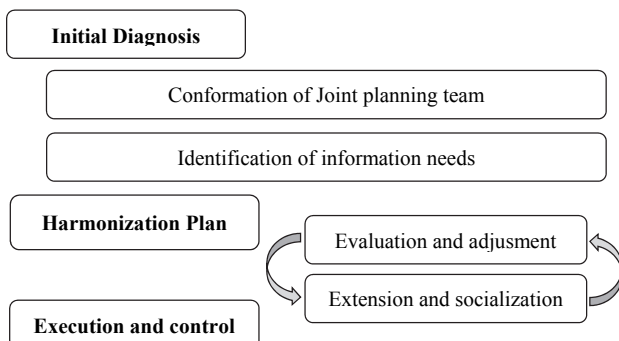


Figure 2. Deployment strategy for coordination.
Source: The Authors.

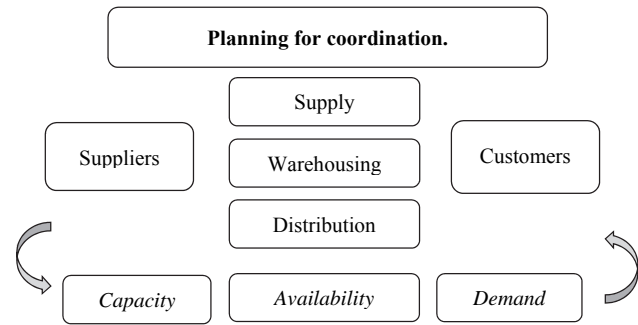


Figure 3. Planning for coordination.
Source: The Authors.

It is necessary to establish control mechanisms and improvement measures for the activities and integrated processes in order to evaluate the degree of coordination achieved so far, as well as the advantages or benefits in terms of organizational goals. These performance indicators may include:

- * Number of processed orders/number of received orders.
- * Delays in execution of orders or average waiting times.
- * Joint use of means of labor/Means of labor available.
- * Savings on the use of primary or secondary packaging and wrapping.
- * Overbooking savings (over-execution of budget to meet unplanned orders).

Achieving economies of scale, incentives for volume and quantity discounts may represent motivating factors to initiate processes of inter-organizational coordination. In logistics, coordination strategies regarding the size of lots, common or joint replenishment times, purchasing pools, joint negotiations for contracting support services (e.g., security services, locative maintenance, transportation, marketing, etc.) or logistics operators, are procedures available to all organizations without requiring major structural or background changes in their operations.

To achieve inter-organizational coordination, it is first necessary to have a high degree of internal coordination within each of the organizations in the SC. This is to say that the areas within each organization (namely distribution, supply and warehouse) must be aligned with its own particular business strategy and also simultaneously with the whole-chain strategy [2]. Both the joint goals that motivate the articulation and coordination of operations and the organizational goals must be aligned with the overall goals of the chain. This, in turn, generating a management cycle that requires constant updating and revision by the joint planning management team.

4. Context of the medicines supply chain

In the pharmaceutical industry, it is common to find supply chains of multiple levels depending on the degree of intermediation required for the acquisition and adequacy of medicines. This is particularly true in the case of oncology drugs, which must be conditioned prior to application to the patients. In Colombia, not all institutions providing health services (IPS) have the facilities and the necessary certifications

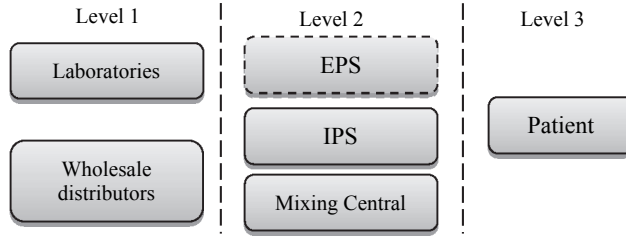


Figure 4. Oncology drugs SC Structure
Source: The Authors.

for the proper conditioning of drugs. Therefore, IPS must rely on outsourced services provided by specialized mixing centers, which are responsible for collecting, reconditioning and redistributing medicines to their customers (Fig. 4).

The SC of medicines and medical supplies has special characteristics that regulate its management and operation. First, given the high cost of some medicines and their physical-chemical characteristics that determine specific storage conditions, it is necessary to ensure the existence of a cold chain for many of these medicines. In addition, at every IPS, the people responsible for logistics processes may not have the required experience in areas such as inventory management, procurement of stock, or the use of demand prediction techniques. Most importantly, the majority of IPS lack WMS systems to facilitate their work. The risk of shortages, or medicine stock-outs, leads to having specialists prescribing medicines subject to availability and cost criteria. Such restrictions promote the use of substituting, less-effective medicines that match the procurement plans of particular institutions (e.g. IPS).

In the case of oncological medicines in Bogotá, and based on fieldwork carried out through surveys and interviews in 45 IPS that provide oncological services, it was possible to determine the following relevant aspects regarding the management and logistic activities in the health-care SC: 1) the schedule for the application and in-hospital drug distribution is made mostly with daily frequency at specific times, 2) not all IPS have information systems that show the real-time status of medicines inventories, 3) very few IPS have indicators to allow monitoring of logistics management, 4) Only 22% of respondents perform audits to the process of distribution and administration of oncology drugs, and 18% have improvement plans for the distribution process and management of their inventories.

Additionally, it was determined that there are multiple ways to purchase: according to historical demand, according to budget, and according to epidemiological profiles or care plans. Regarding the integration of logistics activities, it was possible to identify that some IPS have begun to work in conjunction with wholesaler distributors by implementing pharmacies that operate inside the corresponding IPS.

5. Mathematical model for measuring the effects of coordination

We developed a mixed integer linear programming model to show the potential benefits of coordination in the health-care SC. The model considers global cost of inventory and physical flows in the chain. The main idea is to choose the

position of inventory levels as a function of three factors, namely demand, capacity and cost. The model assumes known cost structures as well as known demands and capacities. In order to choose the best mix of providers to deliver medication to the patients while minimizing the overall costs of the chain, the model provides a structure in which i ($i=1, \dots, m$.) suppliers provide type- p drugs ($p=1, \dots, q$.) to k ($k=1, \dots, K$.) IPS over t ($t=1, \dots, T$.) periods to be applied to j ($j=1, \dots, n$.) patients (Fig. 5).

The model parameters are:

$CF_{i,k}$ Fixed transport cost from the supplier i to IPS k .

$cv_{i,k}^p$ Variable cost of transporting one unit of single type drug reference p from supplier i to IPS k .

ch_i^p Inventory cost associated to holding one unit of drug p at supplier i at period t .

$CO_{i,t}^p$ Supply capacity of supplier i to deliver drug p at period t .

$NV_{i,k,t}$ Maximal trips available at period t from supplier i to IPS k .

$ED_{i,t}$ Warehouse space of Supplier i at period t .

ch_k^p Inventory cost associate to holding one unit of drug p at IPS k .

$ED_{k,t}$ Warehouse space of IPS k at period t .

$D_{j,t}^p$ Demand of drug p necessary for patient j at period t .

pc^p Purchase price of drug p .

$PD_{k,t}$ Budget of IPS k at period t to buy drugs.

h^p Space required to store one unit of product p .

a^p Space required to transport one unit of product p .

CCC Cargo capacity of the transport medium.

The model's variables are:

$x_{i,k,t}^p$ Units of drug p sent from supplier i to IPS k at period t .

$y_{k,j,t}^p$ Units of drug p delivered to the patient j at IPS k at period t .

$I_{i,t}^p$ Inventory level of drug p held by supplier i at period t .

$I_{k,t}^p$ Inventory level of drug p held by IPS k at period t .

$U_{i,k,t}$ Number of trips from supplier i to IPS k at period t .

The objective function is:

$$\begin{aligned} \text{Min } Z = & \sum_{t=1}^T \sum_{i=1}^m \sum_{k=1}^K CF_{i,k,t} * U_{i,k,t} \\ & + \sum_{t=1}^T \sum_{i=1}^m \sum_{k=1}^K \sum_{p=1}^q cv_{i,k,t}^p * x_{i,k,t}^p \\ & + \sum_{t=1}^T \sum_{i=1}^m \sum_{p=1}^q ch_{i,t}^p * I_{i,t}^p + \sum_{t=1}^T \sum_{k=1}^K \sum_{p=1}^q ch_{k,t}^p \\ & * I_{k,t}^p \quad (0) \end{aligned}$$

The constraints are:

$$I_{i,t}^p = CO_{i,t}^p + I_{i,t-1}^p - \sum_{k=1}^K x_{i,k,t}^p \quad \forall i, p, t \quad (1)$$

$$I_{k,t}^p = \sum_{i=1}^m x_{i,k,t}^p + I_{k,t-1}^p - \sum_{j=1}^n y_{k,j,t}^p \quad \forall k, p, t \quad (2)$$

$$D_{j,t}^p = \sum_{k=1}^K y_{k,j,t}^p \quad \forall j, p, t \quad (3)$$

$$\sum_{k=1}^K x_{i,k,t}^p \leq CO_{i,t}^p \quad \forall i, p, t \quad (4)$$

$$\sum_{p=1}^q h^p * I_{k,t}^p \leq ED_{k,t} \quad \forall k, t \quad (5)$$

$$\sum_{p=1}^q h^p * I_{i,t}^p \leq ED_{i,t} \quad \forall i, t \quad (6)$$

$$\sum_{p=1}^q a^p * x_{i,k,t}^p \leq U_{i,k,t} * CC \quad \forall i, k, t \quad (7)$$

$$U_{i,k,t} \leq NV_{i,k,t} \quad \forall i, k, t \quad (8)$$

$$\sum_{p=1}^q \sum_{j=1}^n pc^p * y_{k,j,t}^p \leq PD_{k,t} \quad \forall k, t \quad (9)$$

$$x_{i,k,t}^p, y_{k,j,t}^p, I_{k,t}^p, I_{i,t}^p, U_{i,k,t} \geq 0, \quad \text{Integers} \quad (10)$$

The Objective function (0) includes the fixed cost of transport according to various factors, namely the number of trips made, the variable cost of transportation and the cost of inventory from suppliers and IPS.

Constraint (1) represents the balance between the flows of medicines of supplier i . This constraint also establishes that the amount of inventory of medication p in supplier i at period t equals the amount of type- p drug available plus the inventory of type- p drug at period $t-1$, minus the amount of drug p sent to IPS k .

Constraint (2) represents the balance between the flows of drugs at IPS k . This constraint states that the amount of p -type drug inventory at IPS k , in time period t , is equal to the amount of p -type medication received from suppliers in the same period, plus the type- p drug inventory from period $t-1$, minus the amount of drug p given to patient j during the t period.

Constraint (3) shows the balance of type- p drugs required for patient j and the amount delivered from IPS k at period t .

Constraint (4) guarantees that the amount of type- p drug to be sent to IPS k cannot exceed the quantities of type- p drug available at supplier i at period t .

Constraint (5) guarantees that the amount of all drugs stored in IPS k at period t cannot exceed the available space.

Constraint (6) guarantees that the amount of all drugs stored at supplier i at period t cannot exceed the available

space.

Constraint (7) ensures that the total volume occupied by the drugs to be transported from supplier i to IPS k at period t is less than or equal to the number of trips that can be handled by the volume capacity of the transport medium.

Constraint (8) ensures that the number of scheduled trips from supplier i to IPS k at period t is less than the amount of allowable trips between these two locations for the same period.

Constraint (9) ensures that the total purchase cost assumed by IPS k is less than or equal to the total budget in period t .

Constraint (10) ensures that all decision variables are non-negative and integer.

Likewise, the initial conditions are established:

$$I_{i,0}^p = A1 \quad \forall i = 1, \dots, m \quad p = 1, \dots, q$$

$$I_{k,0}^p = A2 \quad \forall k = 1, \dots, m \quad p = 1, \dots, q$$

Where A1, A2 represent the initial inventory levels.

And final conditions:

$$I_{k,T}^p = A3 \quad \forall k = 1, \dots, m \quad p = 1, \dots, q;$$

A3 represents the desired level of inventory at the end of the planning horizon T .

6. Results

To show that coordinated planning can improve the performance of the supply chain, the mathematical model of mixed integer linear programming, presented in section 5, was used to compare two scenarios with varying demand distributions in the same planning horizon. Patient care is performed under the assumption of coordination involving a free flow of information and cooperation to satisfy demand in the first scenario; and the assumption of non-cooperation in the second scenario, where demand exceeds the capacity and should be deferred.

Emulated data resembling real data was used in order not to compromise the organization's private data. The main objective of the proposed model is to show the potential benefits in terms of global SC cost reduction resulting from information sharing in the attention planning process.

To run the proposed model, a sample instance with 3 suppliers, 2 IPS, and 7 patients (requiring 3 types of drugs), in a planning horizon of 5 periods of time was solved. To this end, the mathematical modeling package GAMS was used; specifically, the CPLEX solver was run on a machine with a Core i5 processor of 2.67 GHz and 4 GB RAM. The demand distribution data is presented in Table 1.

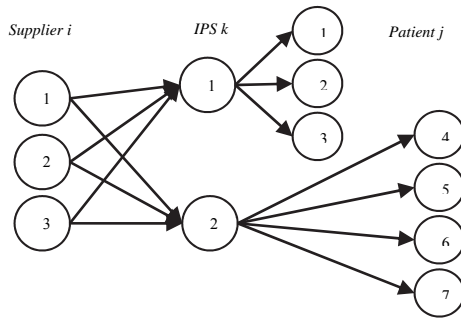


Figure 5. Model Structure.
Source: The Authors.

To run the model, the following assumptions hold: 1) The fixed and variable costs of supplier 1 are higher than those of provider 2 and, in turn, the costs of supplier 3 are higher than those of provider 2. 2) The costs of keeping type- p drugs in inventory are 4, 2 and 1 percent the price of the corresponding drug for suppliers 1, 2 and 3, respectively. 3) The cost of maintaining a type- p drug in the IPS inventory is 80% of the product's price, since a drug suitable for one patient is rarely appropriate for another. 4) The space to store drugs in the oncology department at the IPS is limited. 5) The means of transportation from the supplier to the IPS has limited capacity. 6) A one-to-one linear relationship is assumed between drugs and patients.

Two scenarios were proposed to show the effects of coordination in terms of global SC cost. The baseline scenario assumes uncoordinated patient attention ignoring capacity restrictions of IPS. The modified scenario assumes coordination for patient attention. The results of the two scenarios show a reduction in the overall costs of the chain of 16.33% due to a variation in the behavior of inventories.

Demand leveling across various planning periods as a result of the joint planning process contributes to avoiding bottlenecks and shortages of capacity in order to better cope with patient attention in the proposed scenarios. The model shows how, by varying only the demand in the different time periods of the planning horizon (i.e. rescheduling patient appointments), it is possible to achieve lower costs as a result of the changes in inventory levels (see Table 3 and Table 4). In real life, members of the SC would be able to achieve savings only by sharing information at different levels of the chain.

Table 1.
Patients (p) attended by period (t) in baseline and modified scenarios and drug required.

p\t	Baseline					Modified					Drug		
	1	2	3	4	5	1	2	3	4	5	1	2	3
1	0	0	0	1	0	1	0	0	0	0	1	0	0
2	0	0	0	1	0	0	1	0	0	0	1	0	0
3	0	0	0	1	0	0	1	0	0	0	0	1	0
4	0	1	0	0	0	0	0	1	0	0	0	1	0
5	0	0	0	1	0	0	0	0	1	0	0	0	1
6	0	1	0	0	0	0	0	0	0	1	0	0	1
7	0	0	1	0	0	0	0	0	0	1	0	0	1
Total	0	2	1	4	0	1	2	1	1	2	2	2	3

Source: Authors.

Table 2.

Minimum overall cost of the chain in each scenario.

Baseline	Modified	Savings
23,231,000	19,437,000	3,794,000

Source: Authors.

Table 3.

Behavior of inventories in the base scenario.

Drug		t1	t2	t3	t4	t5
Supplier1	P1	1	1	4	5	5
	P2	0	0	4	4	4
	P3	1	3	3	4	6
Supplier2	P1	0	0	0	1	1
	P2	1	1	2	2	3
	P3	0	1	1	2	2
Supplier3	P1	0	2	2	2	2
	P2	0	0	4	4	4
	P3	1	2	2	2	3
Total		4	10	22	26	30

Source: Authors.

Table 4.

Behavior of inventories in the Modified scenario.

Drug		t1	t2	t3	t4	t5
Supplier1	P1	0	0	3	4	4
	P2	0	0	4	4	4
	P3	1	5	5	6	6
Supplier2	P1	0	0	0	2	2
	P2	1	1	2	2	3
	P3	0	1	1	2	2
Supplier3	P1	0	1	1	2	2
	P2	0	0	4	4	4
	P3	1	2	2	2	3
Total		3	10	22	28	30

Source: Authors.

7. Discussion and recommendations

By comparing the proposed scenarios, it is possible to show that coordination in the SC can have significant effects in reducing overall chain costs. However, it is not enough to implement strategies and mechanisms that contemplate only an isolated logistic chain process. It is, therefore, necessary to conduct an integrated planning that allows the articulation of multiple processes based on global achievements for the chain, such as economies of scale, eliminating duplicate efforts and waste minimization. Our example shows how, by simply sharing information, it is possible to achieve cost reduction. Future research is necessary to find out how to share the benefits fairly among all the SC members.

To achieve coordination through the SC, standardization of processes and procedures as well as the use of similar information systems that are easily bondable may be a contribute to facilitating the exchange of information for the joint planning of the logistics activities and also to controlling the inventory levels across the chain.

In this paper, we have presented a set of methodological guidelines that can be applied to achieve coordination of activities in the SC. Furthermore, to show that such coordination can have a positive effect on reducing the

overall costs of the chain, we used a model of mixed integer linear programming that applies to similar contexts and serves to illustrate the potential effects of coordinated operations.

To achieve a practical approach to implementing the guidelines presented, there must be a high degree of commitment and trust between members of the SC to share information about their processes and engage in joint planning processes. For Future research, we suggest simulation scenarios that consider agreements and revenue-sharing contracts, showing the benefits that can be accomplished when sharing the savings that result from the coordination of activities in the logistics context.

In Colombia, the extension and appropriation of recent developments in the theory of SC management can have a great impact on the planning of logistics activities within medical institutions, leading to their improvement and thus contributing to improving the services delivered to the patient. The development of programs or partnerships that enable and facilitate the interaction of professionals and researchers in logistics areas, within health institutions and control agencies, is needed to generate logistics management methodologies, mechanisms and policies that suit the needs of the healthcare sector, thus facilitating coordination for the whole supply chain.

References

- [1] Gunasekaran, A. and Ngai, E.W.T., The future of operations management: An outlook and analysis, *Int. J. Prod. Econ.*, 135(2), pp. 687-701, 2011. DOI: 10.1016/j.ijpe.2011.11.002
- [2] Chopra, S. and Meindl, P., *Supply chain management: Strategy, planning and operation*. Pearson Prentice Hall, 2007, pp. 22-25.
- [3] Gattorna, J., *Gower Handbook of Supply Chain Management*. Gower, 2003, 297 P.
- [4] Kanda, A. and Deshmukh, S.G., Supply chain coordination: Perspectives, empirical studies and research directions, *Int. J. Prod. Econ.*, 115(2), pp. 316-335, 2008. DOI: 10.1016/j.ijpe.2008.05.011
- [5] Arias-Ortiz, E., Crespi, G., Rasteletti, A. and Vargas, F., *Productivity in Services in Latin America and the Caribbean*, 2014.
- [6] Superintendencia de Sociedades, *Comportamiento de las 1000 empresas más grandes del sector real*, Informe, [En línea]. 2014. Disponible en: <http://www.supersociedades.gov.co/noticias/Documents/INFORME SECTOR REAL 1000 empresas SSyotras Mayo 7 2014.pdf>. [Accessed: 08-Oct-2014].
- [7] Cali, M., Ellis, K. and te Velde, D.W., The contribution of services to development and the role of trade liberalisation and regulation, 2008.
- [8] Balcázar, D.A., *Metodología para la coordinación de actores en la cadena de abastecimiento de medicamentos de las instituciones prestadoras de servicios de salud del sector oncológico*, Tesis, Universidad Distrital Francisco José de Caldas, Bogotá, Colombia, 2014.
- [9] Bandyopadhyay, J.K., Six sigma approach to healthcare quality and productivity management, *Int. J. Qual. Product. Manag.*, 5(1), 2005.
- [10] Shah, N., Pharmaceutical supply chains: Key issues and strategies for optimisation, *Comput. Chem. Eng.*, 28(6-7), pp. 929-941, 2004. DOI: 10.1016/j.compchemeng.2003.09.022
- [11] Adarme, W., *Desarrollo metodológico para la optimización de la cadena de suministro esbelta con m proveedores y n demandantes bajo condiciones de incertidumbre: Caso aplicado a empresas navieras colombianas*, Universidad Nacional de Colombia, 2011.
- [12] Vickery, S.K., Jayaram, J., Droge, C. and Calantone, R., The effects of an integrative supply chain strategy on customer service and financial performance: An analysis of direct versus indirect relationships, *J. Oper. Manag.*, 21(5), pp. 523-539, 2003. DOI: 10.1016/j.jom.2003.02.002
- [13] Jaimes, W.A., Serna, M.D.A. y Balcázar, D.A., Modelo para la coordinación de agentes en un sistema logístico de la industria astillera colombiana, *Ing. e Investig.*, 31(2), pp. 102-111, 2011.
- [14] Awad, H.A. and Nassar, M.O., Supply chain integration: Definition and challenges, *Proc. Int. MultiConference Eng. Comput. Sci.*, 2010.
- [15] Skjoett-Larsen, T., Thernøe, C. and Andresen, C., Supply chain collaboration: Theoretical perspectives and empirical evidence, *Int. J. Phys. Distrib. Logist. Manag.*, 33(6), pp. 531-549, 2003. DOI: 10.1108/09600030310492788
- [16] Power, D., Supply chain management integration and implementation: A literature review, *Supply Chain Manag. An Int. J.*, 10(4), pp. 252-263, 2005. DOI: 10.1108/13598540510612721
- [17] Hill, R.M. and Omar, M., Another look at the single-vendor single-buyer integrated production-inventory problem, *Int. J. Prod. Res.*, 44(4), pp. 791-800, 2006. DOI: 10.1080/00207540500334285
- [18] Uthayakumar, R. and Priyan, S., Pharmaceutical supply chain and inventory management strategies: Optimization for a pharmaceutical company and a hospital, *Oper. Res. Heal. Care*, 2(3), pp. 52-64, 2013. DOI: 10.1016/j.orhc.2013.08.001
- [19] Pedroso, M.C. and Nakano, D., Knowledge and information flows in supply chains: A study on pharmaceutical companies, *Int. J. Prod. Econ.*, 122(1), pp. 376-384, 2009. DOI: 10.1016/j.ijpe.2009.06.012
- [20] Kelle, P., Woosley, J. and Schneider, H., Pharmaceutical supply chain specifics and inventory solutions for a hospital case, *Oper. Res. Heal. Care*, 1(2-3), pp. 54-63, 2012. DOI: 10.1016/j.orhc.2012.07.001
- [21] Láinez, J.M., Schaefer, E. and Reklaitis, G.V., Challenges and opportunities in enterprise-wide optimization in the pharmaceutical industry, *Comput. Chem. Eng.*, 47, pp. 19-28, 2012. DOI: 10.1016/j.compchemeng.2012.07.002
- [22] Narayana, S.A., Kumar-Pati, R. and Vrat, P., Managerial research on the pharmaceutical supply chain – A critical review and some insights for future directions, *J. Purch. Supply Manag.*, 20(1), pp. 18-40, 2014. DOI: 10.1016/j.pursup.2013.09.001
- [23] Susarla, N. and Karimi, I.A., Integrated supply chain planning for multinational pharmaceutical enterprises, *Comput. Chem. Eng.*, 42, pp. 168-177, 2012. DOI: 10.1016/j.compchemeng.2012.03.002
- [24] D.D. Dobrzykowski, Saboori, V., Hong, P.C. and Kim, S.-C., A structured analysis of operations and supply chain management research in healthcare (1982–2011), *Int. J. Prod. Econ.*, 147(2/Part B), pp. 514-530, 2014. DOI: 10.1016/j.ijpe.2013.04.055
- [25] Sousa, R.T., Shah, N. and Papageorgiou, L.G., Global supply chain network optimisation for pharmaceuticals, pp. 1189-1194, 2005.
- [26] Masoumi, A.H., Yu, M. and Nagurney, A., A supply chain generalized network oligopoly model for pharmaceuticals under brand differentiation and perishability, *Transp. Res. Part E Logist. Transp. Rev.*, 48(4), pp. 762-780, 2012. DOI: 10.1016/j.tre.2012.01.001
- [27] Castrellón-Torres, J.P., Torres-Acosta, J.H. and Adarme-Jaimes, W., Model for the logistics distribution of medicines in the Colombian public health program, *DYNA*, 81(187), pp. 257-266, 2014.
- [28] Zamora-Aguas, J.P., *Diseño metodológico para la gestión del riesgo en el proceso de aprovisionamiento de la cadena de suministro, caso de estudio IPS de oncología de Bogotá, D.C.*, Tesis. Universidad Nacional de Colombia, 2013.
- [29] Liu, Q. and Chen, J., A coordination mechanism of supply chain under asymmetric information based on supply hub, in *2010 International Conference on E-Product E-Service and E-Entertainment*, 2010, pp. 1-4. DOI: 10.1109/ICEEE.2010.5661096

D.A. Balcázar-Camacho, completed his BSc. in Industrial Engineering from Universidad Francisco José de Caldas, Bogotá, Colombia, in 2009, and his MSc. in Industrial Engineering with emphasis in Logistics and Operations Research at the Universidad Francisco José de Caldas, Bogotá, Colombia. He has experience as a business consultant, a distribution information systems analyzer in massive consumption companies, and as a researcher in the area of logistics and supply chain management by preparing, reviewing and designing research projects and scientific publications. He is a PhD Engineering, Manufacturing and Organizations student at Universidad Nacional de Colombia, Bogotá, Colombia, and his research interests include the development of methodologies and models for operations management, logistics and supply chains as well as knowledge management and organizational learning.

ORCID: 0000-0002-2300-5846

C.A. Lopez-Bello completed his degree in Industrial Engineering at Universidad Francisco José de Caldas in 1986; a specialization in Production Engineering at Universidad Francisco José de Caldas; and a MSc. in Industrial Engineering at Universidad de los Andes. He is an experienced professor and researcher in related production, logistics and operations research areas at Universidad Distrital and Universidad de la Sabana. He currently works as a professor and researcher as part of the Industrial Engineering program. His research interests are related to mathematical programming and operations research applied to decision-making.
ORCID: 0000-0001-7123-0469

W. Adarme-Jaimes, completed his BSc. in Industrial Engineering in 1993 at Universidad Industrial de Santander, Colombia; a Sp. in Production Engineering and Continuous Improvement in 1997 at Universidad Pedagógica y Tecnológica de Colombia, Colombia; a MSc. in Logistics Engineering at Universidad del Valle, Colombia, and a PhD in Engineering, Manufacturing and Organizations at Universidad Nacional de Colombia, Bogotá, Colombia. He has experience in the metallurgic and textile sector, and currently works as an associate professor at Universidad Nacional de Colombia, Bogotá, Colombia. He manages of a research group called Society Economy and Productivity SEPRO - Logistics Line, where he conducts research and extension projects in regional, national and international contexts, all related to logistics and supply chain management. His areas of interest include the coordination of actors and public policies in logistics.

ORCID: 0000-0001-7401-223X



UNIVERSIDAD NACIONAL DE COLOMBIA

SEDE MEDELLÍN
FACULTAD DE MINAS

Área Curricular de Ingeniería Administrativa e
Ingeniería Industrial

Oferta de Posgrados

Especialización en Gestión Empresarial
Especialización en Ingeniería Financiera
Maestría en Ingeniería Administrativa
Maestría en Ingeniería Industrial
Doctorado en Ingeniería - Industria y Organizaciones

Mayor información:

E-mail: acia_med@unal.edu.co
Teléfono: (57-4) 425 52 02

Variables relevantes para la medición de la calidad percibida del servicio bancario

Carmen Regina Berdugo-Correa, Rodrigo Alberto Barbosa-Correa & Lina Margarita Prada-Angarita

Departamento de Ingeniería Industrial, Universidad del Norte, Barranquilla, Colombia. cberdugo@uninorte.edu.co, rbarbosa@uninorte.edu.co, pradal@uninorte.edu.co

Received: January 27th, de 2016. Received in revised form: March 8th, 2016. Accepted: April 25th, 2016

Resumen

El estudio de la calidad percibida de los servicios ha sido ampliamente estudiado pero poco aplicado en el sector bancario por lo cual despierta gran interés por parte de las instituciones prestadoras del servicio e investigadores. Dada la complejidad y gran cantidad de variables involucradas, su estimación se convierte en un reto, debido a su característica multidimensional. El presente artículo tiene como objetivo determinar las variables que definen la calidad percibida de los servicios bancarios, mediante el análisis de los modelos propuestos en el área y bajo una metodología de clasificación que contempla veinticuatro artículos científicos. Para la clasificación se tuvo en cuenta: modelo, tipo de variable, año de la propuesta y contexto de aplicación. Se hace el análisis desde la perspectiva del cliente del servicio bancario, resultando la siguiente propuesta de variables de calidad: elementos tangibles, garantía, capacidad de respuesta, seguridad, empatía, tecnología y conectividad, cobertura, y accesibilidad.

Palabras clave: Calidad de servicio; SERVQUAL; Servicios Bancarios.

Relevant variables in the measuring of the perceived quality in bank service

Abstract

The study of the perceived quality of services has been widely studied but little applied in the banking sector which attracts great interest from the institutions providing service and researchers. Given the complexity and large number of variables involved, the estimate becomes a challenge because of its multidimensional feature. This article aims to determine the variables that define the perceived quality of banking services, by analyzing the models proposed in the area and under a classification methodology that includes twenty scientific articles. For classification took into account: model variable, year of the proposal and application context. The analysis is done from the client's perspective of banking services, resulting in the following proposal for quality variables: tangibles, assurance, responsiveness, assurance, empathy, technology and connectivity, coverage and accessibility

Keywords: Quality service; SERVQUAL; Bank Services

1. Introducción

La importancia de la calidad ha sido demostrada en el campo de las operaciones [1-3], pero poco estudiada y aplicada en el sector de servicios, ya que en este último existe una diferencia entre el cumplimiento de las expectativas del cliente y la satisfacción de sus necesidades. Para las organizaciones es un obstáculo medir la calidad del servicio ofrecido a sus clientes

porque a estos se les dificulta emitir un juicio de algo intangible. De hecho, la heterogeneidad de los servicios lleva a que cada persona perciba su calidad de manera diferente porque no todos la valoran igual. La calidad se convierte en un reto para este tipo de organizaciones ya que su sostenibilidad se basa en la recompra y recomendación de los mismos, entre otros.

Si bien es categórico que toda organización tiene como meta ganar dinero, la velocidad de cambio del mundo actual ha

How to cite: Berdugo-Correa, C.R., Barbosa-Correa, R.A. and Prada-Angarita, L.M., Variables relevantes para la medición de la calidad percibida del servicio bancario. DYNA 83(197), pp. 213-222, 2016.

obligado a las mismas a crear ventaja competitiva mediante el aumento de la satisfacción del cliente. Esto requiere que se concentren no sólo en el proceso, sino en la manera en que son entregados. Desde el punto de vista del consumidor, la calidad se puede definir como el grado de aceptación de un producto o servicio, así a mayor aceptación, mayor calidad [4]. La calidad del servicio se entiende como una actitud o juicio global del cliente acerca de la superioridad del mismo y difiere de la satisfacción del cliente, aunque la naturaleza exacta de esta diferencia genera discusión [5].

Otros autores definen la calidad del servicio como la diferencia entre lo percibido y lo esperado, es decir, una comparación entre las expectativas y el desempeño percibido [6]. Otros argumentan que es una comparación del desempeño con los estándares ideales [7], mientras que hay autores que proponen que es la percepción del desempeño en sí misma la que define la calidad [8].

Prakash y Mohanty [9] aseguran que la calidad de los servicios genera buenos dividendos cuando se hace bien, por lo que invertir dinero y tiempo en su mejoramiento, aumenta las ventas, retiene al clientes y logra la exclusividad de uso por parte de los usuarios. La calidad de servicio también se define como la percepción que tiene el cliente sobre el servicio recibido [10]; dicha percepción es producto de la discrepancia entre las expectativas que tenía antes de recibir el servicio y el desempeño actual. No obstante, los clientes no perciben la calidad de servicio como un concepto unidimensional, si no que hacen una diferenciación de la información en varias dimensiones [10].

Adicionalmente, ofrecer servicios de calidad se ha convertido en una necesidad dada la gran variedad de alternativas disponibles que tienen los clientes. Los competidores son cada vez más veloces en atender las nuevas demandas por lo que la calidad debe convertirse en una ventaja competitiva de toda organización. Por todo esto, la calidad del servicio debe estar inmersa en las estrategias de las organizaciones; un servicio de calidad conlleva a fortalecer relaciones con los clientes, aumenta su satisfacción y lealtad, por ende las ventas y la participación en el mercado, generando mayor crecimiento y rentabilidad.

El principal propósito de este estudio es determinar las variables claves para la medición de la calidad del servicio bancario, utilizando una metodología para la identificación, clasificación y análisis de las variables hasta ahora estudiadas.

Se tomó como referente la metodología propuesta por Galván [11], que facilitó la evaluación de los variables que impactan la percepción de la calidad de los servicios, los modelos, las diferentes metodologías utilizadas, los alcances y limitaciones encontradas y los sectores de aplicación para comprender el panorama mundial y en Colombia, específicamente de la región Caribe, acerca de la conceptualización, medición y análisis de la calidad de los servicios bancarios.

2. Metodología

Las principales corrientes sobre la conceptualización de la calidad se basan en la diferencia entre lo que los clientes esperan y lo que perciben; otras argumentan que este parámetro no es el más adecuado y proponen otros. La principal raíz de la problemática consiste en determinar las variables que miden la calidad del servicio con el fin de tener un conocimiento real sobre lo que reciben los clientes y lo que esperan recibir. Se encuentran en la literatura varios

modelos conceptuales y prácticos propuestos que se pueden clasificar en dos grandes corrientes: la escuela nórdica o norte europea y la escuela norteamericana [12].

La escuela nórdica está encabezada por los aportes realizados por Grönroos [13,14] y Lehtinen y Lehtinen [15]. Sus aportes se fundamentan en el concepto de la tridimensionalidad de la calidad de servicio. La medida de dicha calidad la obtienen a partir de la opinión del cliente, quien determinará si el servicio recibido ha sido superior o inferior a lo esperado. Esta escuela se ha focalizado principalmente en el concepto de calidad de servicio sin entrar a buscar evidencias empíricas que lo soporten. Esta ha sido la principal razón por la que no ha sido ampliamente desarrollada.

Por otra parte, la escuela norteamericana de calidad de servicio, se encuentra encabezada por los aportes realizados por Parasuraman, Zeithaml y Berry [16,6] quienes propusieron el modelo SERVQUAL. Este define la calidad de servicio como la diferencia entre las expectativas previas al consumo del servicio y la percepción del servicio prestado. Esta escuela es la precursora de muchos modelos aplicados a diversos sectores.

En la determinación de las variables claves para la medición de la calidad del servicio bancario, se realizaron una serie de pasos que incluyeron consecutivamente, el análisis de los modelos de la medición de la calidad del servicio en general encontrados en la literatura, y posteriormente, las propuestas dirigidas específicamente al sector bancario, luego se listaron dichas variables, teniendo en cuenta su definición y frecuencia de aparición para consecutivamente desarrollar un análisis de Pareto que permitió seleccionar y proponer los atributos finales de la calidad del servicio bancario con su respectiva definición. A continuación se describen estas etapas.

2.1. Modelos conceptuales de calidad del servicio

2.1.1. Modelo de Grönroos

El modelo técnico y funcional propuesto por Grönroos [13,14] define la calidad percibida del servicio como un proceso evaluativo en el cual el cliente compara sus expectativas del servicio con lo que el percibe que ha recibido en el servicio, proponiendo tres dimensiones básicas que influyen en la calidad del servicio: calidad técnica (corresponde al soporte físico, los medios materiales, la organización interna, en conclusión, lo “qué” el consumidor recibe), calidad funcional (la forma en que el consumidor es tratado en el desarrollo del proceso de prestación del servicio, en conclusión, el “cómo” se recibe) e imagen corporativa (la forma en que los clientes perciben la organización).

2.1.2. Modelo de Bitner

El modelo propuesto por Bitner [17] describe la calidad de servicio percibida como una consecuencia de la experiencia satisfacción/insatisfacción. En esta teoría se presenta un paradigma por la disconformidad que existe entre el resultado del servicio percibido en una transacción, con base en determinadas características, y las expectativas de servicio, las cuales se forma el cliente de acuerdo con la

actitud inicial que tenga antes de recibir el servicio. Esta disconformidad afecta la experiencia de satisfacción/insatisfacción del cliente y, posteriormente, la calidad de servicio percibida, derivándose las conductas post-compra.

2.1.3. Modelo de Nguyen

El modelo de calidad de los servicios propuesto por Nguyen [18] se estructura a partir de las relaciones entre los componentes: imagen de la empresa, resultado del personal en contacto, organización interna de la servucción, soporte físico de la servucción y la satisfacción del cliente, todos estos componentes explican la evaluación de la calidad de un servicio hecha por el cliente.

2.1.4. Modelo de Bolton y Drew

El modelo de disconformidad de Bolton e Drew [19] sostiene que la calidad de servicio es una actitud frente a la empresa y todos sus servicios, mientras que la satisfacción se expresa con respecto al servicio ofrecido localmente y es de la calidad de servicio percibida, de la que se derivan las intenciones comportamentales. Los autores asumen que las expectativas del cliente se forman por la comunicación boca-oído, las necesidades personales del cliente y sus experiencias pasadas, mientras que las percepciones se evalúan con base en atributos y dimensiones del servicio y que dichos atributos se pueden clasificar en organizacionales y de ingeniería. De la disconformidad entre las expectativas de los clientes y las percepciones en función de dichos atributos y dimensiones, el cliente quedará satisfecho o insatisfecho y, resultado de varias satisfacciones, el cliente percibirá calidad en el servicio.

2.1.5. Modelo de Chakrapani

El modelo tridimensional para la medición de la calidad propuesto por Chakrapani [20] hace diferenciación sobre la calidad del servicio medida a partir de lo que denominan servicio personal y funcional, siendo el servicio personal aquel que hace referencia a la relación con el cliente y el funcional a aquel relacionado con características que para ser modificadas o mejoradas no necesitan a los clientes, similares a lo que podrían ser características funcionales de un producto.

Siguiendo esta línea, los autores realizan el siguiente dimensionamiento: (a) Primera: servicio o producto que se ofrece, teniendo en cuenta si es aquel que el cliente quiere, si satisface sus necesidades; (b) Segunda: Dependencia, se refiere a la confianza que el cliente tiene en que el servicio o producto se ofrecerá; y (c) Tercera: exceder expectativas, se refiere a si el servicio ofrecido excede las expectativas del cliente y lo hace ver importante.

Este modelo es básicamente conceptual; propone analizar el mercado como un conjunto de líneas, cuadrados y esferas que corresponden a las dimensiones anteriormente mencionadas. Si la organización prestadora del servicio solo se comporta como una línea(es decir únicamente es fuerte en la primera dimensión) mientras su entorno se comporta como cubos y esferas, no se alcanzará el éxito en la prestación del servicio.

2.1.6. Modelo de Rust y Oliver

El modelo de los tres componentes propuesto por Rust y Oliver [21] establece tres elementos para la calidad del servicio que son: (a) Servicio y sus características, dadas en el diseño del servicio antes de ser entregado; (b) Entrega del servicio y (c) Entorno o ambiente del servicio interno en donde se incluye la cultura organizacional y el externo que contiene el ambiente físico que se utiliza para la prestación del servicio. Su planteamiento inicial fue para productos físicos pero fue redefinido para los servicios. Según los autores, el servicio y sus características se refieren al diseño antes de ser entregado al cliente.

2.1.7. Modelo de Parasuraman, Zeithaml y Berry

Por otro lado, el modelo SERVQUAL: modelo del gap (SERVICE QUALITY) propuesto por Parasuraman, Zeithaml e Berry [16] se centra en el punto de vista del cliente ya que define la calidad del servicio como la diferencia entre las expectativas y la percepción del desempeño que tiene el cliente con respecto al servicio, de esta forma mide la calidad del servicio a través de la sustracción entre la puntuación que los clientes le dan al desempeño y a las expectativas. El modelo propuesto se basa en el análisis de las brechas o "Gaps". Estos "gaps" se describen de la siguiente manera:

- Gap 1: discrepancia entre las expectativas de los clientes sobre un servicio concreto y las percepciones o creencias que se forman los directivos sobre lo que espera el cliente de ese servicio.
- Gap 2: diferencia entre las percepciones de los directivos y las especificaciones o normas de calidad.
- Gap 3: discrepancia entre las especificaciones o normas de calidad del servicio y la prestación del mismo.
- Gap 4: diferencia entre la prestación del servicio y la comunicación hacia los clientes acerca de la entrega del servicio.
- Gap 5: diferencia entre las expectativas del cliente y la percepción del servicio. Este "Gap" depende de la dimensión y el sentido de las otras cuatro brechas asociadas con la entrega del servicio por parte de la empresa. ($GAP5 = GAP1 + GAP2 + GAP3 + GAP4$)

En [6] los autores establecen las dimensiones de la calidad del servicio como: confiabilidad, responsabilidad, tangibilidad, seguridad y empatía. A partir del modelo SERVQUAL surgieron una buena parte de trabajos realizados por otros investigadores.

2.1.8. Modelo de Cronin Jr e Taylor.

Por otro lado, el modelo SERVPERF propuesto por Cronin Jr y Taylor [8] es basado en el desempeño en contraposición del modelo SERVQUAL. SERVPERF se basa únicamente en el desempeño o la percepción del desempeño. Determina la calidad del servicio como antecedente de la satisfacción del cliente.

El modelo SERVPERF consiste en una escala de 22 ítems, que produce como resultado una puntuación de la calidad general del servicio. Se compone de los mismos ítems que el modelo SERVQUAL pero se diferencia en que solo hace énfasis en la parte de las percepciones, lo que quiere

decir que en la práctica únicamente se realiza una parte de las dos que componen el cuestionario de SERVQUAL.

2.1.9. Modelo de Teas

El modelo de desempeño evaluado de Teas [7] desarrolló y contrastó empíricamente las escalas desempeño evaluado y calidad normalizada como instrumentos alternativos para medir la calidad de servicio percibida. El desempeño evaluado se produce por la necesidad de conceptualizar las expectativas como puntos ideales en los modelos actitudinales; y la calidad normalizada integra el concepto de punto ideal clásico, con el de expectativa revisada. Este modelo plantea que el incremento de la diferencia entre las percepciones y las expectativas puede no reflejar necesariamente un incremento continuo en los niveles de calidad percibida, como implica SERVQUAL.

2.1.10. Modelo de Philip e Hazlett

El modelo PCP de Philip e Hazlett [22] (PCP-Pivotal-Core-Peripheral) está estructurado en tres niveles jerárquicos con el que se propone un marco de trabajo más general que el ofrecido por SERVQUAL. La estructura se basa en las tres clases principales de atributos: fundamental o pivote (salidas), nuclear y periférica, que conjuntamente representan insumos y procesos.

Los atributos fundamentales son considerados como los determinantes sobre la razón por la que el cliente decidió acercarse a una organización en particular y están definidos como “producto terminado” o “salida”, es decir, lo que el consumidor espera obtener y recibir. Los atributos nucleares o básicos pueden ser descritos como el conjunto de las personas, los procesos y la estructura organizacional de servicios a través del cual los clientes tienen que interactuar y/o negociar para que puedan lograr recibir el atributo fundamental. Los atributos periféricos se definen como aquellos que complementan la entrega del servicio y permiten que la experiencia del cliente sea satisfactoria. Cuando el cliente evalúa la calidad del servicio recibido, se basa principalmente en los atributos fundamentales.

2.1.11. Modelo de Koelemeijer, Roest y Verhallen

El modelo integrador propuesto por Koelemeijer, Roest y Verhallen [23] aporta un esquema con una estructura integradora, en la que el resultado y las expectativas se encuentran en tres planos: (A) Macro, indica expectativas y percepciones de resultado respecto a clase-productos; (B) Meso, indica expectativas y percepciones de resultado con respecto al proveedor de servicio; y (C) Micro, indica expectativas y resultado percibidos con respecto a una transacción individual. La comparación entre expectativas y resultado percibido en calidad y precio, resulta en valor percibido para cada nivel.

2.1.12. Modelo de Brady y Cronin Jr

El modelo jerárquico multidimensional propuesto por Brady y Cronin Jr [24] plantean que la medición de la calidad se realiza a partir de dimensiones que a su vez son

conformadas por subdimensiones, proponiendo un análisis multinivel y estructural. El modelo no tiene una estructura estandarizada, es básicamente la integración de instrumentos y modelos anteriormente propuestos por [14, 6, 21, 25] para conformar una estructura global de medición de la calidad del servicio. A través de los diferentes modelos propuestos, es evidente que la calidad percibida es una variable multidimensional por lo que se manifiesta a través de una serie de constructos con un alto grado de correlación.

2.2. Modelos para la medición de la calidad del servicio bancario

Las dimensiones o factores no son universales como algunos autores como Parasuraman, Zeithaml e Berry [6] han sostenido, sino que son específicos del tipo de servicio [26]. En el estudio presentado por Sangeetha y Mahalingam [27] se identifican varios modelos aplicados al sector bancario, sus fortalezas y limitaciones. Se muestra a continuación una exploración de los modelos, identificando las variables consideradas y las principales observaciones de los mismos.

2.2.1. Modelo de Mersha y Adlakha

El modelo de atributos de calidad del servicio desde la perspectiva de los clientes propuesto por Mersha y Adlakha de [28] tiene como objetivo de estudio, identificar y clasificar los atributos de calidad desde la perspectiva de los consumidores (los autores modificaron y extendieron el instrumento SERVQUAL incluyendo preguntas adicionales para evaluar el núcleo del servicio, la personalización del servicio y el conocimiento e información). En este caso, los atributos de la “buena” y “mala” calidad de los servicios se identifican y clasifican por separado. Cinco servicios fueron seleccionados para el estudio: médicos, banca minorista, mantenimiento de automóviles, escuelas / universidades y restaurantes de comida rápida.

Según el estudio, los seis atributos de buena calidad en el servicio son: 1 el conocimiento del servicio; 2 exhaustividad / precisión; 3 la consistencia / fiabilidad; 4 costo razonable; 5 la disposición para corregir errores; y 6 servicio oportuno / rápido. También se plantearon los seis atributos de mala calidad del servicio más relevantes son: 1 falta de conocimiento sobre el servicio; 2 indiferencia del empleado o actitud “no me importa; 3 renuencia a corregir errores; 4 inconsistencia del servicio; 5 servicio descuidado o poco riguroso; y 6 altos costos.

Para el servicio de la banca minorista, las características más importantes de buena calidad encontradas fueron: 1 voluntad para corregir errores; 2 exhaustividad / precisión; 3 la consistencia / fiabilidad; y 4 habilidad de conocimiento. Se encontró también que las características más importantes de la mala calidad del servicio son: 1 renuencia a corregir errores; 2 indiferencia del empleado; 3 falta de conocimiento sobre el servicio; y 4 servicio descuidado o poco riguroso.

2.2.2. Modelo de Ennew, Reed y Binks

En el análisis de la Importancia-Desempeño y la calidad del servicio de Ennew, Reed y Binks [29] los autores examinan los problemas relacionados con la medición de la

calidad de los servicios y presentan un conjunto de índices para proporcionar medidas de expectativas, percepciones y la satisfacción general.

Las 11 dimensiones de la calidad de los servicios utilizados en este análisis fueron: 1. conocimiento del negocio; 2. conocimiento de la industria; 3. conocimiento del mercado; 4. otorga consejos útiles; 5. amplia gama de servicios; 6. tasas de interés competitivas; 7. cargos/costos competitivos; 8. rapidez de la decisión; 9. diseño de las finanzas; 10. Negociar con una persona; y 11. fácil acceso a oficial de préstamo.

2.2.3. Modelo de Avkiran

La escala de la calidad del servicio al cliente propuesta por Avkiran [30] tuvo como propósito desarrollar un instrumento multidimensional que pudiera ser aplicado para la medición de la calidad del servicio al cliente según su percepción de la red de sucursales bancarias. Los autores tomaron como punto de partida el trabajo desarrollado por Parasuraman, Zeithaml y Berry [16], El trabajo resultó en un instrumento de 17 puntos en 4 dimensiones para medir la calidad del servicio al cliente en la red de sucursales bancarias, con un enfoque en la banca minorista. Las cuatro dimensiones que surgieron son: conducta del personal, credibilidad, comunicación y acceso a los servicios de cajero.

2.2.4. Modelo de Blanchard y Galloway

El modelo de la calidad del servicio de Blanchard y Galloway [31] determina la percepción de los clientes respecto a los requerimientos de calidad de servicio en la banca minorista usando SERVQUAL [6]. Debido a que dicho modelo no fue exitoso para proporcionar alguna idea acerca de cómo mejorar el servicio, los autores intentaron desarrollar un modelo alternativo de mayor utilidad. El modelo propuesto se basa en tres dimensiones: el proceso / resultado (diseño de servicios y los resultados), subjetivo / objetivo (medida del grado en que la calidad del servicio puede ser objetivamente valorada) y “soft” / “hard” (“Hard” representa aspectos físicos del servicio, mientras que “Soft” representa la interacción interpersonal).

2.2.5. Modelo de Johnston

El modelo de los factores de calidad del servicio basado en la satisfacción propuesto por Johnston [32] provino cuando el autor encontró una paradoja inquietante en el sector bancario del Reino Unido dada la cantidad de reportes de insatisfacción de los clientes, a pesar de los grandes esfuerzos por tratar de mejorar el servicio. El estudio clasifica los factores de calidad en términos de su importancia relativa y su efecto sobre la satisfacción y la insatisfacción. Los 18 factores determinantes de la calidad de los servicios propuestos son: 1. Accesibilidad; 2. Estética; 3. Atención; 4. Disponibilidad; 5. Cuidado; 6. Limpieza/Pulcritud; 7. Comfort; 8. Compromiso; 9. Comunicación; 10. Competencia; 11. Cortesía; 12. Flexibilidad; 13. Amabilidad; 14. Funcionalidad; 15. Integridad; 16. Fiabilidad; 17. Capacidad de respuesta; y 18. Seguridad.

2.2.6. Modelo de Joseph, McClure y Joseph

El modelo desarrollado por estos autores está basado en la tecnología utilizada para la entrega de la calidad del servicio en donde, el uso de esta tecnología en la prestación de servicios bancarios, es cada vez más frecuente ya que permite reducir los costos y eliminar incertidumbres. Este estudio se centra en el papel que desempeña la tecnología en la banca australiana y su impacto en la prestación de la calidad del servicio percibida.

Joseph, McClure y Joseph [33] utilizaron la cuadrícula de Hemmasi, Strong y Taylor [34] e identificaron un modelo de seis factores consistentes en: 1. Conveniencia / precisión; 2. Realimentación feedback. / Gestión de reclamaciones; 3. Eficiencia; 4. Gestión de colas; 5. Accesibilidad; y 6. Personalización.

2.2.7. Modelo de Bahia y Nantel

En el modelo para la calidad del servicio bancario de Bahia y Nantel [35] los autores sostienen que los instrumentos disponibles incluyen una escala diseñada para bancos específicos, o instrumentos no diseñados específicamente para el servicio bancario sino para una amplia gama de servicios. Dado lo anterior, los autores proponen una escala de 31 puntos para medir la calidad percibida de los servicios bancarios que abarcan seis dimensiones: 1. Efectividad y seguridad; 2. Accesibilidad; 3. Precio; 4. Tangibles; 5. Portafolio de servicios; y 6. Fiabilidad.

2.2.8. Modelo de Sureshchandar, Rajendran y Kamalanabhan

Estos autores [36] revisaron exhaustivamente el modelo de [4] SERVQUAL y sus 22 puntos lo que implicó analizarlos desde la interacción/intervención de los humanos en la entrega del servicio, además de las facetas tangibles de servicio. El instrumento parece haber pasado por alto algunos otros factores importantes, por ejemplo: Servicio principal (producto), sistematización / estandarización de la prestación de servicios (el elemento no humano) y la responsabilidad social de la organización. En un esfuerzo por conceptualizar la calidad del servicio, estos autores identificaron cinco factores de calidad de servicio que son: 1. Servicio principal producto.; 2. Elemento Humano en la prestación del servicio; 3. Sistematización de la prestación del servicio; 4. Tangibles; y 5. Responsabilidad social de la organización.

2.2.9. Modelo de Aldlaigan y Buttle

El Modelo SYSTRA propuesto por Aldlaigan y Buttle [37] es un instrumento para medir la calidad percibida basado en el modelo propuesto por [13] realizado especialmente para los servicios bancarios, concluyendo en una escala de 21 ítems basados en la calidad a nivel organizacional y transaccional. Desarrolló las siguientes 4 dimensiones: 1. Calidad del sistema de servicio, 2. Calidad de comportamiento del servicio, 3. Calidad de servicio de las máquinas y 4. Exactitud transaccional del servicio.

2.2.10. Modelo de Al-Hawari, Hartley y Ward

Para Al-Hawari, Hartley y Ward [38] la calidad del servicio automatizado se plantea como el factor que determina el éxito o el fracaso del comercio electrónico. Los autores afirman que los modelos disponibles en la actualidad están limitados en su enfoque, porque abarcan sólo un canal electrónico – Internet – haciendo caso omiso de los demás canales de servicio automatizadas. En relación con el sector bancario, las investigaciones han identificado que los clientes tienden a utilizar una combinación de canales. Por lo tanto, los autores se esfuerzan por desarrollar un modelo integral teniendo en cuenta las características únicas de cada canal de distribución y otras dimensiones que tienen una influencia potencial sobre la calidad. Los autores proponen cinco factores: 1. Servicio de cajeros automáticos; 2. Servicio de banco por Internet; 3. Servicio de banco por teléfono; 4. Servicio básico; y 5. Percepción del cliente sobre el precio.

2.2.11. Modelo de Karatepe, Yavas y Babakus

La escala de calidad de servicio para la banca propuesta por Karatepe, Yavas y Babakus [39], los autores desarrollaron un instrumento de 20 puntos para medir las percepciones de los clientes sobre la calidad del servicio bancario en el norte de Chipre. Los resultados mostraron que la calidad del servicio podría ser conceptualizada y medida a través de cuatro dimensiones: 1. Ambiente del servicio; 2. Calidad de la interacción; 3. Empatía y 4. Confiabilidad.

2.2.12. Modelo de Ehigie

EL modelo de las expectativas y servicio percibido por parte de los clientes de Ehigie [40] en la que los autores desarrollaron una escala mediante la técnica cualitativa para medir las variables que intervienen en las expectativas del cliente, la calidad de servicio percibida y la satisfacción que impactan la lealtad de los clientes hacia los bancos en Nigeria, se encontraron las siguientes variables: 1. trabajadores del banco con el dominio de las habilidades requeridas; 2. trabajadores del banco con el conocimiento y experiencia requerida; 3. continuidad del servicio al cliente en los próximos años; 4. entender las necesidades de los clientes; 5. oferta de un servicio rápido y eficiente; 6. proporcionar seguridad física a los clientes; 7. confidencialidad de las transacciones; 8. actitud positiva del personal de servicios al cliente; 9. fiabilidad de banco; 10. buena reputación del banco; 11. amabilidad del personal; 12. mantener a los clientes informados; 13. escuchar a los clientes; 14. la inclusión de servicio bancario los sábados; 15. horarios bancarios extendidos; y 16. la cobertura del seguro para los clientes.

3. Resultados

3.1. Comparativo de atributos de la calidad

En resumen, se encuentran una variedad de dimensiones en común a través de los diferentes modelos y aplicaciones desarrollados para medir la calidad del servicio. En la Tabla 1 se muestra un resumen de las variables encontradas en la

literatura. Cabe resaltar que aunque los títulos o nombres de las variables no coincidían exactamente entre los modelos, el concepto desarrollado por cada autor sí, por tanto fueron agrupadas las siguientes categorías de variables: Tangible, Garantía, Capacidad de Respuesta, Seguridad, Empatía, Tecnología y Conectividad, Cobertura, Accesibilidad, Precio, Responsabilidad Social e Imagen Corporativa.

En la Fig. 1 se resumen el porcentaje de atributos considerados a través de los diferentes modelos analizados, teniendo en cuenta que fueran mencionados explícitamente o en su concepto en la literatura.

Tabla 1.
Atributos considerados por los modelos

Referencia	Variable equivalente (correspondencia conceptual)									
	Tangible	Garantía	Capacidad de Respuesta	Seguridad	Empatía	Tecnología y Conectividad	Cobertura	Accesibilidad	Precio	Responsabilidad Social Imagen Corporativa
[10,11]	x	x	x		x	x		x		x
[4]	x	x	x	x	x					
[14]	x				x		x	x	x	
[15]	x	x	x		x					x
[16]	x	x	x	x	x	x				
[17]	x	x	x	x	x					
[6]	x	x	x	x	x					
[5]	x	x	x	x	x					
[18]	x	x	x	x	x		x	x		
[20]	x	x	x	x	x		x	x	x	
[19]	x	x	x	x	x			x		
[21]	x	x	x	x	x			x	x	x
[25]		x	x	x	x				x	
[26]		x	x	x	x		x		x	
[27]		x	x	x	x					
[28]	x	x	x	x	x					
[29]	x	x	x	x	x					
[30]		x	x	x	x			x		
[32]	x	x		x	x		x	x	x	
[33]	x	x	x	x	x	x				x
[34]	x	x	x	x	x			x		
[35]						x		x	x	
[36]	x	x	x	x	x					
[37]	x	x	x	x	x					x

Fuente: Los Autores

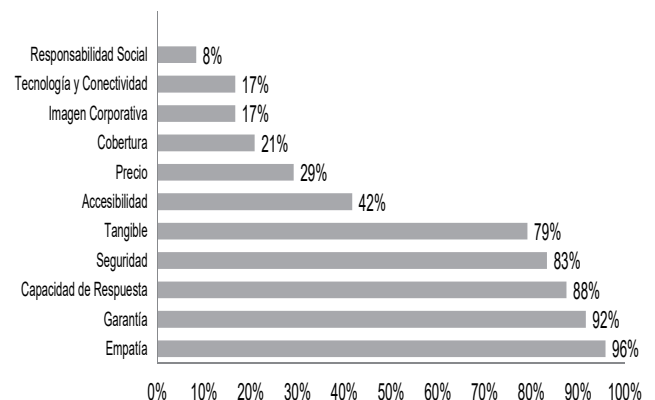


Figura 1: Porcentaje de atributos considerados a través de los modelos relevantes sobre la calidad del servicio

Fuente: Los Autores

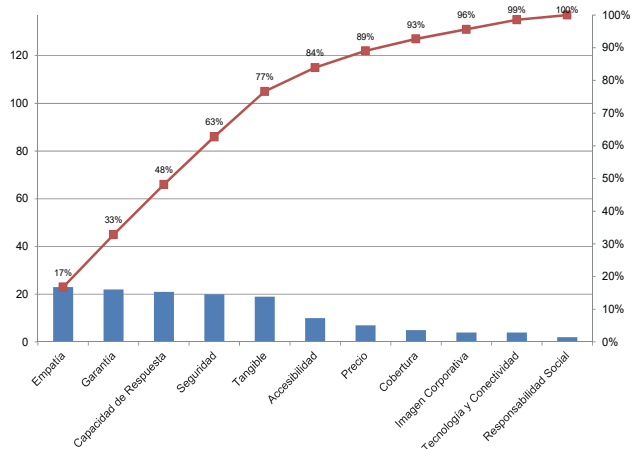


Figura 2: Diagrama de Pareto: variables medición de la calidad del servicio.
Fuente: Los Autores

3.2. Variables relevantes en la medición de la calidad del servicio bancario

Resumida la frecuencia de aparición de las variables relevantes para la medición de la calidad del servicio bancario, se propone un análisis de Pareto para establecer las variables más importantes. El principio de Pareto se basa en la proporción 80-20 que indica que el grupo minoritario, formado por un 20% de población, ostenta el 80% de algo y el grupo mayoritario, formado por un 80% de población, el 20% de ese mismo algo. Como lo muestra la Figura 2, se encuentran en el grupo minoritario, las variables Garantía, Empatía, Seguridad, Capacidad de Respuesta, Tangible y Accesibilidad. En el grupo mayoritario las variables Precio, Cobertura, Imagen Corporativa, Tecnología y Conectividad y Responsabilidad Social.

Aunque las variables Cobertura y Tecnología y Conectividad quedan en segundo plano de acuerdo con el análisis de Pareto, se considera que para el sector bancario no pueden ser descartadas y se proponen como variable relevante para medir la calidad del servicio bancario.

Se plantean las siguientes variables para medir la calidad del servicio bancario: 1. Tangible; 2. Garantía; 3. Capacidad De Respuesta; 4. Seguridad; 5. Empatía; 6. Tecnología y Conectividad; 7. Cobertura; y 8. Accesibilidad. La conceptualización de cada una de ellas, se describe a continuación.

3.2.1. Variable elementos tangibles

Debido a la naturaleza intangible de los servicios bancarios, el cliente asocia los elementos relacionados con el ambiente físico cuando evalúa la calidad de servicio [41]. En el modelo [32] proponen el factor estética y lo definen como el grado en que los servicios son agradables para el cliente, incluyendo tanto la apariencia como el ambiente del entorno de servicio, el aspecto y la presentación de las instalaciones, bienes y personal. Por su parte, en el modelo [36], el factor tangible del servicio contempla los equipos, la señalización, la apariencia de los empleados y el ambiente físico hecho por el hombre que rodea el servicio. En el modelo [39] se plantea la dimensión de ambiente de servicios que se refiere a la

apariciencia y el aspecto tanto interior como exterior de las instalaciones del banco. La variable tangible está propuesta en el modelo SERVQUAL de [42] definida como la apariencia de las instalaciones físicas, equipos, personal y elementos de comunicación. Se adapta esta definición y se ajusta al contexto bancario:

La variable tangible se define como los elementos visibles que tiene el banco relacionados con las condiciones físicas, apariencia y cantidad de los recursos de: instalaciones, equipos, personal y dispositivos de comunicación y operación necesarios para proporcionar un buen servicio.

3.2.2. Variable garantía

La dimensión técnica del modelo propuesto en [13,14] incluye los factores de habilidades técnicas del personal, conocimiento del empleado, soluciones técnicas y sistemas computarizados. En [28] para el servicio de la banca minorista plantean como características importante de buena calidad la exhaustividad / precisión y la consistencia / fiabilidad; y en [29] la dimensión de negociar con una persona está relacionada con la garantía. Por su parte en [30] se plantea la dimensión Conducta del personal definida como la capacidad de respuesta, la conducta civilizada y la presentación del personal que proyecte una imagen profesional a los clientes.

En [32] proponen los factores de compromiso, integridad y fiabilidad relacionados con la garantía. En [40] se plantea la continuidad del servicio al cliente en los próximos años relacionados con la garantía.

En el modelo de [37] se plantea la exactitud transaccional del servicio que mide la exactitud percibida por los clientes del servicio y sus empleados. Por su parte el modelo de [39] plantea la confiabilidad que se refiere a la fiabilidad del servicio y la exactitud de los registros e información. La variable garantía está propuesta en el modelo SERVQUAL de [34] definida como la capacidad de realizar el servicio prometido de forma fiable y precisa. Se adapta esta definición y se ajusta al contexto bancario:

La variable garantía se define como la capacidad que tiene el banco para cumplir los compromisos realizados con el cliente, incluyendo tiempo, conocimiento, confianza y precisión necesaria para proporcionar un buen servicio.

3.2.3. Variable capacidad de respuesta

En [28] uno de los atributos de la buena calidad en el servicio es la rapidez y oportunidad en la entrega del servicio y en [96] las dimensiones de otorgar consejos útiles y rapidez de la decisión están relacionadas con la capacidad de respuesta. Por su parte en [30] se plantea la dimensión credibilidad, en [32] proponen los factores de atención, disponibilidad, cuidado, amabilidad y capacidad de respuesta. En el estudio de [40] se contemplan las variables oferta de un servicio rápido y eficiente, actitud positiva del personal de servicios al cliente y amabilidad del personal.

La variable capacidad de respuesta está propuesta en el modelo SERVQUAL de [42] definida como la voluntad de ayudar a los clientes y ofrecer un servicio rápido. Se adapta esta definición y se ajusta al contexto bancario:

La variable capacidad de respuesta se define como la capacidad que tiene el banco para dar solución rápida y

eficiente ante los requerimientos del cliente, y que son necesarios para que el banco proporcione un buen servicio.

3.2.4. Variable seguridad

En [29] las dimensiones de conocimiento del negocio, conocimiento de la industria y conocimiento del mercado están relacionadas con la seguridad.

Por su parte en [30] se plantea la dimensión comunicación entendida como la asesoría financiera y notificación oportuna a los clientes. En [32] proponen los factores de comunicación, competencia, cortesía y seguridad. En el estudio [33] se identificaron los factores eficiencia y gestión de colas, así como también en [35] se contemplan las variables efectividad y seguridad.

En el estudio de [40] se contemplan las variables de trabajadores del banco con el dominio de las habilidades requeridas, trabajadores del banco con el conocimiento y experiencia requerida, proporcionar seguridad física a los clientes, confidencialidad de las transacciones, fiabilidad de banco y mantener a los clientes informados.

La variable seguridad está propuesta en el modelo SERVQUAL de [42] definida como el conocimiento y la cortesía de los empleados y su capacidad de transmitir confianza y seguridad. Se adapta esta definición y se ajusta al contexto bancario:

La variable seguridad se define como la percepción que tiene el cliente acerca de la privacidad de su información y transacciones en la prestación del servicio, incluyendo el conocimiento, cortesía y confianza que transmite el trabajador hacia el cliente.

3.2.5. Variable empatía

En [28] uno de los atributos de la buena calidad en el servicio es la disposición para corregir errores. Por su parte en [30] se plantea la dimensión acceso a los servicios de cajero refiriéndose al adecuado número de empleados que atienden a clientes en todo horario y durante las horas pico. En el modelo de [31] proponen la dimensión “Soft” que representa la interacción interpersonal.

En [32] proponen los factores de acceso, confort y flexibilidad. En la propuesta de [36] se plantea el factor del elemento Humano en la prestación del servicio y en el modelo de [37] se plantea la dimensión calidad de comportamiento del servicio referente a la prestación del servicio por parte de los empleados, que incluye empleados educados, con cortesía, amigables y actitudes positivas.

Por su parte en el modelo de [39] se plantea la empatía como la atención individualizada a los clientes y la disposición del personal para ayudar a los clientes y resolver sus problemas oportunamente. En el estudio de [40] se contemplan las variables entender las necesidades de los clientes y escucharlos. La variable empatía está propuesta en el modelo SERVQUAL de [42] definida como el cuidado, la atención individualizada que la compañía ofrece al cliente. Se adapta esta definición y se ajusta al contexto bancario:

La variable empatía se define como la capacidad que tiene el banco de personalizar los servicios al ajustarse a las necesidades de los clientes.

3.2.6. Variable tecnología y conectividad

En el estudio [36] los autores proponen el factor sistematización de la prestación del servicio como elemento importante de la calidad del servicio bancario. Por su parte, [38] contemplan los factores servicio de cajeros automáticos, servicio de banco por Internet y servicio de banco por teléfono. Adicionalmente, en [13] la dimensión técnica incluye la disposición de soluciones técnicas, sistemas computarizados y calidad de las máquinas como elementos de la calidad del servicio. Teniendo en cuenta lo propuesto por otros autores, se adapta esta definición y se ajusta al contexto bancario:

La variable tecnología y conectividad se define como la disponibilidad de medios tecnológicos para facilitar la conectividad del cliente con el banco en cuanto a atención del mismo y realización de transacciones.

3.2.7. Variable cobertura

En el modelo propuesto por [35] se contempla la dimensión portafolio de servicios. Los autores sostienen que este es un elemento adicional que el modelo SERVQUAL, ampliamente utilizado, no contempla. En [43] uno de los factores propuestos es el ofrecimiento de diferentes tipos de préstamos ajustados a las necesidades del cliente. Adicionalmente, [41] contempla la oferta de una amplia gama de productos y servicios bancarios acorde con las últimas innovaciones. Por su parte, en [29] una de las dimensiones propuestas fue la oferta de una amplia gama de servicios. Teniendo en cuenta lo propuesto por otros autores, se adapta esta definición y se ajusta al contexto bancario:

La variable cobertura se define como la percepción del cliente sobre la amplitud de los servicios estableciendo si es lo suficientemente diversificado según sus necesidades.

3.2.8. Variable accesibilidad

En el estudio [33] se identificó el factor accesibilidad, así como también en [35]. En la propuesta de [37] y en [38] se contempla la ubicación de las sucursales bancarias y de los cajeros automáticos en lugares seguros y convenientes como elemento relevante de la calidad. Teniendo en cuenta lo propuesto por otros autores, se adapta esta definición y se ajusta al contexto bancario:

La variable accesibilidad se define como la percepción del cliente sobre el grado de facilidad para acceder a los servicios proporcionados por el banco.

4. Conclusiones

En el presente estudio se desarrolló una metodología de identificación, análisis y clasificación de las variables utilizadas para la medición de la calidad del servicio bancario en los últimos años. Se obtiene un listado de variables con las definiciones propuestas por otros autores y se plantea la propia adaptándola al contexto bancario. Este análisis se convierte en un referente valioso para investigaciones y aplicaciones asociadas al estudio de la calidad del servicio bancario, su medición e impacto en las organizaciones.

Se determinaron las variables relevantes en la medición

de la calidad del servicio bancario, destacándose las cinco dimensiones propuestas en el modelo SERVQUAL (metodología que contempla los elementos: tangible, garantía, capacidad de respuesta, seguridad y empatía). Adicionalmente se proponen las variables de tecnología y conectividad, cobertura y accesibilidad.

Bibliografía

- [1] Niebles, R.P.P.-B., Trespacios, Ó.H.O., Hernández, S.P. C. y Solano, E.G., Metodología para la implementación del diseño económico y/o estadístico de cartas de control x-barra con parámetros variables (VP). DYNA, 81(184), pp. 150-157, 2014.
- [2] Pulido-Rojano, A.D.J., Methodological design for the prevention of risk in production processes. DYNA, 82(193), pp. 16-22, 2015.
- [3] Rios-Griego, J.H., Design of a single sampling attributes in search of a social optimum. DYNA, 78(169), pp. 53-61, 2011.
- [4] Deming, W.E., Out of the Crisis. MIT press, 2000. ISBN 0262541157.
- [5] Robinson, S., Measuring service quality: current thinking and future requirements. Marketing Intelligence & Planning, 17(1), pp. 21-32, 1999. DOI: 10.1108/02634509910253777
- [6] Parasuraman, A., Berry, L.L. and Zeithaml, V.A., Servqual: A multiple-item scale for measuring consumer perceptions of service quality. Journal of Retailing, 64(1), pp. 12-37, 1988.
- [7] Teas, R.K., Expectations, performance evaluation, and consumers' perceptions of quality. The Journal of Marketing, pp. 18-34, 1993.
- [8] Cronin Jr, J.J. and Taylor, S.A., Measuring service quality: A reexamination and extension. The Journal of Marketing, pp. 55-68, 1992. ISSN 0022-2429.
- [9] Prakash, A. and Mohanty, R.P., Understanding service quality. Production Planning & Control, n. ahead-of-print, pp. 1-16, 2012. DOI: 10.1080/09537287.2011.643929
- [10] Zeithaml, V.A. and Bitner, M.J., Service marketing: Integrating customer focus across the firm. Boston: McGraw Hill, 2000. ISBN 0077131711.
- [11] Galvan, J.L., Writing literature reviews: A guide for students of the social and behavioral sciences, Pyrczak, 2006. ISBN 1884585663
- [12] Brogowicz, A.A., Delene, L.M. and Lyth, D.M., A synthesised service quality model with managerial implications. International Journal of Service Industry Management, 1(1), pp. 27-45, 1990. DOI: 10.1108/09564239010001640
- [13] Grönroos, C., A service quality model and its marketing implications. European Journal of Marketing, 18(4), pp. 36-44, 1984. DOI: 10.1108/EUM00000000004784
- [14] Grönroos, C., Service quality: The six criteria of good perceived service quality. Review of Business, 9(3), 1988.
- [15] Lehtinen, U. and Lehtinen, J.R., Two approaches to service quality dimensions. Service Industries Journal, 11(3), pp. 287-303, 1991. DOI: 10.1080/02642069100000047
- [16] Parasuraman, A., Zeithaml, V.A. and Berry, L.L., A conceptual model of service quality and its implications for future research. The Journal of Marketing, pp. 41-50, 1985. DOI: 10.2307/1251430
- [17] Bitner, M. J. Evaluating service encounters: the effects of physical surroundings and employee responses. The Journal of Marketing, p. 69-82, 1990. ISSN 0022-2429. DOI: 10.2307/1251871
- [18] Nguyen, N., Un modèle explicatif de l'évaluation de la qualité d'un service: une étude empirique. Recherche et Applications en Marketing, 6(2), pp. 83-98, 1991. ISSN 0767-370.
- [19] Bolton, R.N. and Drew, J.H., A multistage model of customers' assessments of service quality and value. Journal of Consumer Research, pp. 375-384, 1991. ISSN 0093-5301.
- [20] Chakrapani, C., Service quality measurement and the myth of service, [Online]. 1991, 5 P. Available at: www.chuckchakrapani.com/articles/PDF/91040542Chakrapani.pdf
- [21] Rust, R.T. and Oliver, R.L., Service quality: New directions in theory and practice. Sage Publications, 1994. ISBN 0803949197.
- [22] Philip, G. and Hazlett, S.-A., The measurement of service quality: A new PCP attributes model. International Journal of Quality & Reliability Management, 14(3), p. 260-286, 1997. DOI: 10.1108/02656719710165482
- [23] Koelemeijer, K., Roest, H. and Verhallen, T.M., An integrative framework of perceived service quality and its relations to satisfaction/dissatisfaction, attitude and repurchase intention: A multilevel approach. Marketing for the new Europe: Dealing with complexity. The 22nd Annual Conference of the European Marketing Agency, Barcelona, Spain, May 25-28, 1993, pp. 683-699.
- [24] Brady, M.K. and Cronin Jr, J.J., Some new thoughts on conceptualizing perceived service quality: a hierarchical approach. The Journal of Marketing, pp. 34-49, 2001. DOI: 10.1509/jmkg.65.3.34.18334
- [25] Dabholkar, P.A., Thorpe, D.I. and Rentz, J.O., A measure of service quality for retail stores: scale development and validation. Journal of the Academy of Marketing Science, 24(1), pp. 3-16, 1996. ISSN 0092-0703.
- [26] Buttle, F., SERVQUAL: Review, critique, research agenda. European Journal of marketing, 30(1), pp. 8-32, 1996. DOI: 10.1108/03090569610105762
- [27] Sangeetha, J. and Mahalingam, S., Service quality models in banking: A review. International Journal of Islamic and Middle Eastern Finance and Management, 4(1), pp. 83-103, 2011. DOI: 10.1108/17538391111122221
- [28] Mersha, T. and Adlakha, V., Attributes of service quality: The consumers' perspective. International Journal of Service Industry Management, 3(3), pp. 34-45, 1992. DOI: 10.1108/09564239210015157
- [29] Ennew, C.T., Reed, G.V. and Binks, M.R., Importance-performance analysis and the measurement of service quality. European Journal of Marketing, 27(2), pp. 59-70, 1993. DOI: 10.1108/03090569310026402
- [30] Avkiran, N.K., Developing an instrument to measure customer service quality in branch banking. International Journal of Bank Marketing, 12(6), pp. 10-18, 1994. DOI: 10.1108/02652329410063223
- [31] Blanchard, R. and Galloway, R., Quality in retail banking. International Journal of Service Industry Management, 5(4), pp. 5-23, 1994. DOI: 10.1108/09564239410068670
- [32] Johnston, R., Identifying the critical determinants of service quality in retail banking: Importance and effect. International Journal of Bank Marketing, 15(4), pp. 111-116, 1997. DOI: 10.1108/02652329710189366
- [33] Joseph, M., McClure, C. and Joseph, B., Service quality in the banking sector: the impact of technology on service delivery. International Journal of Bank Marketing, 17(4), pp. 182-193, 1999. DOI: 10.1108/02652329910278879
- [34] Hemmasi, M., Strong, K.C. and Taylor, S.A., Measuring service quality for strategic planning and analysis in service firms. Journal of Applied Business Research, 10(4), pp. 24-33, 1994.
- [35] Bahia, K. and Nantel, J., A reliable and valid measurement scale for the perceived service quality of banks. International Journal of Bank Marketing, 18(2), p. 84-91, 2000. DOI: 10.1108/02652320010322994
- [36] Sureshchandar, G., Rajendran, C. and Kamalanabhan, T., Customer perceptions of service quality: a critique. Total Quality Management, 12(1), pp. 111-124, 2001. DOI: 10.1080/09544120020010138
- [37] Aldlaigan, A.H. and Buttle, F.A., SYSTRA-SQ: A new measure of bank service quality. International Journal of Service Industry Management, 13(4), pp. 362-381, 2002. DOI: 10.1108/09564230210445041
- [38] Al-Hawari, M., Hartley, N. and Ward, T., Measuring banks' automated service quality: A confirmatory factor analysis approach. Marketing Bulletin, 16(1), pp. 1-19, 2005.
- [39] Karatepe, O.M., Yavas, U. and Babakus, E., Measuring service quality of banks: Scale development and validation. Journal of Retailing and Consumer Services, 12(5), pp. 373-383, 2005. DOI: 10.1016/j.jretconser.2005.01.001
- [40] Ehigie, B.O., Correlates of customer loyalty to their bank: a case study in Nigeria. International Journal of Bank Marketing, 24(7), pp. 494-508, 2006. DOI: 10.1108/02652320610712102
- [41] Lewis, B.R., Service quality: An international comparison of bank customers' expectations and perceptions. Journal of Marketing Management, 7(1), pp. 47-62, 1991. DOI: 10.1080/0267257X.1991.9964139

- [42] Parasuraman, A., Berry, L.L. and Zeithaml, V.A., Understanding customer expectations of service. *Sloan Management Review*, 32(3), pp. 42, 1991.
- [43] Paswan, A.K. et al., Search quality in the financial services industry: A contingency perspective. *Journal of Services Marketing*, 18(5), pp. 324-338, 2004. DOI: 10.1108/08876040410548267

C.R. Berdugo-Correa, es Ing., MSc y Dra en Ingeniería Industrial de la Universidad del Norte, Barranquilla, Colombia, Sp. en Ingeniería de Procesos Industriales. Ha desarrollado estudios de consultoría e investigación en la optimización de procesos que impacte en la productividad y competitividad de las organizaciones. Docente de las áreas de métodos cuantitativos, gestión operacional y de organizaciones. Participó en varios proyectos de acreditación. Profesora de la Universidad del Norte. Trabajó en la industria en áreas de mejoramiento de procesos, costos, eficiencia y manejo de almacenes.
ORCID:0000-0002-4803-2816

R. Barbosa-Correa, es Ing. Químico, especializado en Gerencia de la Calidad y en Normalización de procesos. Universidad de los Andes MSc. en Informática y Análisis de Datos y Dr. en Ingeniería. ENSP- Francia, es profesor titular, profesor investigador del Departamento de Ingeniería Industrial de la Universidad del Norte, Barranquilla, Colombia, miembro del grupo de investigación de Productividad y Competitividad de dicha institución.
ORCID: 0000-0001-7726-1012

L.M. Prada-Angarita, es Ing. y MSc. en Ingeniería Industrial, Universidad del Norte, Barranquilla, Colombia. Experta internacional en teoría de restricciones del Goldratt Schools. IDEAL Scholar 2014. Experiencia de 7 años en procesos académicos y administrativos y 4 años como profesora universitaria. Participó en varios proyectos de acreditación nacional e internacional bajo el estándar de ABET. Experiencia en calidad percibida de los servicios, diseño del proceso de desarrollo de productos basado en ingeniería concurrente y de estrategias de cooperación en logística de la cadena de suministro del Caribe Colombiano. Se desempeña como docente de la Universidad del Norte.
ORCID: 0000-0003-1670-5171



UNIVERSIDAD NACIONAL DE COLOMBIA

SEDE MEDELLÍN
FACULTAD DE MINAS

Área Curricular de Ingeniería Administrativa e
Ingeniería Industrial

Oferta de Posgrados

Especialización en Gestión Empresarial
Especialización en Ingeniería Financiera
Maestría en Ingeniería Administrativa
Maestría en Ingeniería Industrial
Doctorado en Ingeniería - Industria y Organizaciones

Mayor información:

E-mail: acia_med@unal.edu.co
Teléfono: (57-4) 425 52 02

Coordinated recharge of electric vehicles in real time

Cristian C. Mendoza ^a, Adriana M. Quintero ^b, Francisco Santamaria ^c & Alexander Alarcon ^d

^a Universidad Distrital Francisco José de Caldas, Bogotá, Colombia, ccmendozag@correo.udistrital.edu.co

^b Universidad Distrital Francisco José de Caldas, Bogotá, Colombia, amquinterom@correo.udistrital.edu.co

^c Universidad Distrital Francisco José de Caldas, Bogotá, Colombia, fsantamariap@udistrital.edu.co

^d Universidad Distrital Francisco José de Caldas, Bogotá, Colombia, jaalarconv@udistrital.edu.co

Received: January 24th, de 2016. Received in revised form: February 23th, 2016. Accepted: March 4th, 2016

Abstract

The electric vehicle is a transportation technology that requires a management system for the electric networks of the future in order to avoid overloads and adverse effects on voltage level and power quality. This article presents a management algorithm to adjust the operation and recharge conditions of electric vehicles in real time, optimizing the use of network capacity and improving the voltage profile through the coordination of recharges depending on loss sensitivity; an approach to identify nodes where power losses can decrease or increase under a specific load condition. The results obtained by applying the strategy are presented through a random consumption model of vehicles based on existing mobility patterns using a distribution circuit of 628 nodes. This allows the identification of electric vehicle penetration after which it becomes necessary to implement the proposed management scheme.

Keywords: Electric vehicle, management algorithm, recharge coordination, selection of maximum sensitivities, sensitivity ranges, voltage level, power losses, maximum capacity.

Coordinación de la recarga de vehículos eléctricos en tiempo real

Resumen

El vehículo eléctrico es una tecnología de transporte que requiere un sistema de gestión para las redes eléctricas del futuro para evitar sobrecargas y efectos negativos en el nivel de voltaje y la calidad de potencia. Por lo tanto, este artículo presenta un algoritmo de gestión para ajustar las condiciones de operación y recarga de los vehículos eléctricos en tiempo real, optimizando el uso de la capacidad de la red y mejorando el perfil de voltaje a través de la coordinación de las recargas dependiendo de la sensibilidad de pérdidas, un enfoque que identifica los nodos donde las pérdidas de potencia pueden disminuir o aumentar bajo condiciones de carga específica. Los resultados obtenidos aplicando la estrategia, son presentados a través de un modelo de consumo aleatorio de los vehículos basado en los patrones de movilidad existentes, usando un circuito de distribución de 628 nodos. Esto permite la identificación del nivel de penetración después del cual se hace necesario implementar el esquema de gestión propuesto.

Palabras clave: Vehículo Eléctrico; Algoritmo de gestión; recarga coordinada; selección de sensibilidades máximas; rango de sensibilidades; nivel de voltaje; pérdidas de potencia; capacidad máxima.

1. Introduction

The current electric sector combines the concepts of energy efficiency, demand response, and load management. All of them aim to change energy consumption patterns through changes in facilities, infrastructure, devices, customs or prices [1]. For this reason, countries with projects incorporating new generation resources [2,3] or implementing technologies such as electric vehicles, adopt

strategies to manage demand, since this represents economic benefits for the end user and the energy supplier [4,5], significantly reducing energy demand at peak hours, and thus preserving the stability of the electric system and service continuity [6].

Some research studies have shown that to perform the simultaneous recharge of electric vehicles (EV) without any control, depending on EV penetration, can cause power losses, voltage drops, harmonics and transformer overloads,

How to cite: Mendoza, C.C., Quintero, A.M., Santamaria, F., and Alarcon, A., Coordinated recharge of electric vehicles in real time. DYNA 83(197), pp. 223-232, 2016

among others, which increment the system operation and maintenance costs [7-10]. Therefore, in order to achieve the massive introduction of electric vehicles without threatening the system, it is important to consider a strategy that will coordinate the recharges to mitigate or prevent the described impacts.

According to the above, this paper proposes a management algorithm for recharging electric vehicles in real time, based on the operation conditions of the distribution system, the priority of connection and the selection of nodes contributing to the increment of power losses. This will enable optimization in the use of the network capacity, a factor that can have an impact on network reliability, security, and performance. In Section 2, a background of the impact of electric vehicles on distribution networks is presented. A methodology for managing the electric vehicle recharge in real time is described in Section 3. In Section 4 a management algorithm for the simultaneous recharge of electric vehicles is presented, and Section 5 outlines the implementation of the proposed management algorithm in a case study. Conclusions are drawn in section 6.

2. Background of the impact of electric vehicles on distribution networks

The massive introduction of electric vehicles in a distribution network forces those responsible for its operation and maintenance to verify the impact these can cause. However, these loads can have many effects on the network and they depend largely on factors such as EV penetration, the implemented recharge strategy, the battery size, the number of users, the load profiles, the distribution transformer capacity, the size of the primary feeder, among others [11-13].

An effect of the simultaneous connection of nonlinear loads such as battery chargers of electric vehicles is the injection of harmonics to the network [14-17]. In particular, distribution transformers are designed to withstand linear loads with a specific frequency; therefore, current harmonics will contribute to increased losses of up to 40% and a rise in equipment temperature [7,18]. Therefore, the life cycle of the device decreases exponentially and the investment costs could increase by up to 15% [19]. Recent research states that power losses are directly proportional to EV penetration [8,20,21].

Another point of interest is the voltage profile, where a decisive factor is load location, given that although the number of vehicles increases, if the distance between the load and the substation feeding the circuit is short, the voltage drops will be up to 10% lower in comparison to the corresponding ones for distant loads. However, since distribution systems typically have a voltage imbalance caused by the load variation in each of the phases, the phase to which a certain number of vehicles is connected will experience a large voltage drop while others will experience a rise in voltage [20,22-25].

In addition, some studies show that in a scenario where the penetration rate is between 20% and 30%, the maximum transformer capacity could be exceeded by up to 36% [22,26-29], without considering the constantly changing technology for batteries and chargers. In these works, the authors

calculate the energy required by each electric vehicle connected to the distribution system using a probabilistic model, which allows the impact analysis of massive EV recharge. In the case study presented in [10], it was found that the voltage profile could drop by up to 8% below the established limit, the power losses may increase by 3-5 times compared to a scenario without electric vehicles, and the maximum capacity of the transformer feeding the circuit would be exceeded by about 13.16%.

Several management strategies have been evaluated in this research, such as the use of objective functions for minimizing power losses through centralized communication systems [30] or demand response strategies, where the user schedules the recharge according to his preferences. This, taking into account that management is based on a stochastic method considering the driving patterns and the time of arrival to the house [31].

[32-33] consider the recharge of a fleet of electric vehicles based on the amount of required energy, the selling price, the contracted power and the time they must be ready to operate. In [34], the authors seek to recharge the vehicles maximizing the satisfaction of owners and minimizing the network operation costs, without violating the restrictions of that network. Meanwhile, in [35], the management system considers generation from renewable sources in order to maximize the net income of charging stations, by scheduling the energy required for the recharge on the previous day and controlling this process in real time. Therefore, the proposed management schemes usually seek to optimize the use of the existing electric networks, fulfill the load requirements of users and maintain the provided quality service, avoiding an alteration in voltage levels, load curve of the system, and power losses [36-38].

3. Methodology for managing the electric vehicle recharge in real time

This research proposes a methodology to coordinate the recharge of electric vehicles based on the minimization of power losses and considering voltage restrictions and distribution system capacity, thereby mitigating the impact they could generate on network reliability and security.

The developed methodology has been divided into four phases, which are described in this section and are summarized in Fig. 1. These phases include the construction of the consumption model for electric vehicles, the impact assessment, the proposed management scheme and the results analysis.

3.1. Step 1: Characterization of electric vehicle consumption

This phase seeks statistical correlation between the variables that determine the electric vehicle consumption (number of connections, battery capacity and traveled distances) and mobility behavior in the city, in a scenario of introducing electric vehicles proportionally to the number of users of the distribution system. In this way, it is expected that loading conditions and power demand required by the network vary randomly according to a probability distribution of the energy required by the battery.

3.2. Step 2: Identification of the massive recharge impact

From the model of electric vehicle consumption in the previous phase, the technical information of the distribution system, and the residential demand of the circuit, it is possible to calculate the load flow for each hour and to determine the power in all branches, the voltage levels and the losses in the distribution system. Once the operation limits are defined, it is possible to assess the impact on the recharge condition of electric vehicles "without control", for which it is possible to develop a computer application where several case studies can be programmed and the statistical analysis of these variables, can be performed [10].

3.3. Step 3: Characterization of electric vehicle consumption

With the identification of the network problems associated with massive electric vehicle recharge, the third phase of the methodology seeks to maintain the operation conditions of the distribution circuit based on the minimization of technical losses in the system. The aim is to increase the network efficiency while reducing operation and maintenance costs. Under this approach, the concept of demand response is involved as a control strategy when

facing unpredictable loading conditions generated by the consumption model developed in the first phase.

The main tool of the proposed management scheme is the loss sensitivity approach [39,40], through which it is possible to identify the nodes with sensitivity indices representing the largest contribution to system losses according to the power consumed by the electric vehicle connected there. Later, these nodes are selected to shift the battery recharge to another time, when the loading conditions of the system are not exceeded or during off-peak demand on the load curve, so that both the voltage profile and the capacity of the transformer feeding the distribution circuit do not exceed the permitted limits.

For this purpose, it is convenient again to develop and implement the proposed management algorithm within the software tool, in order to observe the behavior of the analyzed variables in three scenarios: the first, without electric vehicles; the second, with electric vehicles without management; and the third, performing the management by implementing the proposed scheme. It is convenient to use the proposed application, since the process must be repeated many times, depending on the number of nodes in the system and the number of vehicles connected to this.

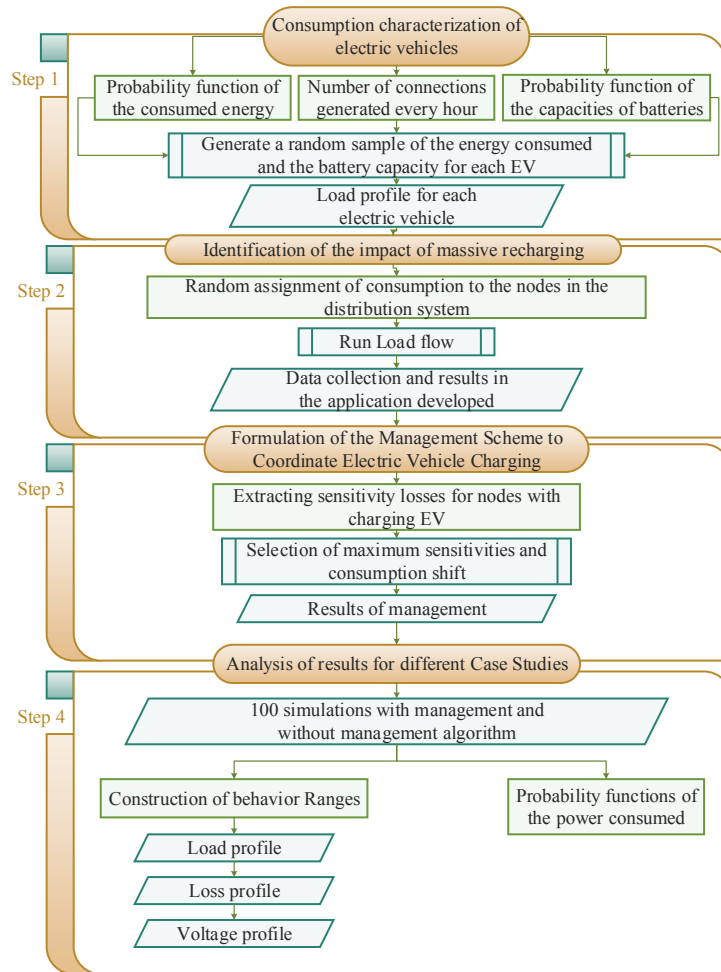


Figure 1. Conceptual map of the methodology used to manage the recharge of electric vehicles.
Source: The Authors

3.4. Step 4: Results analysis for different case studies

After implementing the management scheme proposed in the previous phase, several case studies are constructed by varying the percentage of EV penetration in the distribution system. N simulations are performed using the Monte Carlo method [41]. Using these simulations, voltage profiles, load curves and power loss curves are evaluated to find the range of values defining the behavior of the system under study. This allows us to determine the level of penetration from which the power demand exceeds the capacity of the circuit, and the variations in the voltage profile are below the accepted limit.

Given the objective of this article, the management algorithm proposed to coordinate the electric vehicle recharge in real time is explained in Section 4.

4. Management algorithm for the simultaneous recharge of electric vehicles

Taking into account that the load conditions can change from hour to hour during a day, the proposed management algorithm coordinates recharges through the selection of maximum sensitivity of the active power to find nodes where the load shedding leads to decreased losses, according to the restrictions of the objective function as presented below.

4.1. Objective function

In order to mitigate the impact of recharging electric vehicles, an objective function is defined that fulfills the operation conditions of the electric grid in a system of equations for minimizing power losses according to the voltage restrictions and maximum capacity of the distribution circuit feeder, as shown in eq. (1)-(3).

$$\min P_{losses,h} = \sum_{k=1}^{n-1} P_{losses(k,k+1),h} \quad (1)$$

Subject to:

$$\sum_{k=1}^n S_{load,h} < D_{max} \quad (2)$$

$$V_{k,min} p.u. \leq |V_k| \leq V_{k,max} p.u. \quad (3)$$

where:

$P_{losses,h}$: total active power losses in the system at a specific time of day.

$P_{losses(k,k+1),h}$: active power losses in a branch comprised between k and $k + 1$, at a specific time.

$S_{load,h}$: apparent power of the load in each node for a specific time.

D_{max} : peak demand level, defined by the capacity assigned to the circuit from the substation transformer.

$|V_k|$: voltage magnitude in node k in p.u.

n : total number of system nodes.

This objective function only has a solution in the operation points fulfilling the two constraints (voltage and peak demand); i.e., in periods when one of the conditions is not fulfilled the recharge of electric vehicles must be managed.

4.2. Proposed management algorithm

One of the concepts used by the proposed management algorithm is selective load shedding, since under a condition of consumed power in the system, the charge of a set of electric vehicles is postponed, in order to fulfill the constraints of the objective function described above.

On the other hand, the management scheme must have a criterion for deciding which system loads must be postponed, so that the problem lies in finding a method to select the loads with the greatest contribution to power losses according to system constraints. To do this, we used the loss sensitivity approach and the maximum sensitivities selection technique.

4.2.1. Loss sensitivity approach

The loss sensitivity approach allows us to quantify through indices the contribution each system node provides to the increment of losses, according to the present load condition [42].

The sensitivity calculation is derived from the Jacobian matrix of the Newton-Raphson method for solving power flows by expressing the voltage magnitude and angle (state variables) in terms of system incremental losses [43].

Finally, the loss sensitivity; i.e., the variation of the power losses in a branch of the system with respect to the variation in the power injected into a node is calculated with eq. (4), where $[J^T]^{-1}$, represents the transpose and inverse matrix of the Jacobian obtained from the load flow in its latest iteration [43].

4.2.2. Maximum sensitivities selection

From the calculation of loss sensitivity, it is possible to find different indices for categorizing the system nodes. In some research works, the selection of the most representative sensitivities corresponds to finding places in the system where power losses can decrease or increase, according to the injection of active or reactive power in the nodes [40,42].

Considering that residential distribution systems typically have radial topologies and the increment of losses is proportional to the distance between their branches, variations in active power are positive. Since these loading conditions vary by the hour, in this study, a tool is managed, through the selection of maximum sensitivities of active power, to find nodes where load shedding leads to loss reduction, according to the constraints of the objective function. The proposed management algorithm uses the selection of maximum sensitivities to postpone the recharge of a specific set of electric vehicles (EV). The process is summarized in the steps shown in Fig. 2, where:

$h = h_{start}: h_{end}$: Time interval of the load curve analysis.

Y_{bus} : Admittance matrix of distribution system under study.

- The sensitivity indices decrease in nodes where electric

vehicle recharge shifts to the next hour. Under this premise, the selection criteria can set sensitivity ranges, which decrease as the electric vehicles recharge is postponed, implying us to: *Select the nodes with the highest sensitivity indices within an established range and shift the electric vehicle recharge to the next hour.*

- When shifting the power that a set of electric vehicles would consume during recharge, the loading conditions in the distribution system will change, which means: *For a specific loading condition in the system, specific sensitivity indices are established.*
- Since the calculation of sensitivities comes from the Jacobian matrix, when the loading conditions change, it is necessary to run the power flow again, then: *The load flow is run each time load shedding occurs in the system.*
- Finally, this process is repeated until the constraints of the objective function are fulfilled, so that the sensitivity ranges established at each iteration of the process decrease as the selective load shedding is carried out. Similarly, it is important to note that these steps must be repeated for each hour of the load curve when any voltage or circuit capacity restriction is not fulfilled, meaning the management algorithm must act whenever the loading conditions change until achieving the coordinated recharge.

$$\begin{bmatrix} \frac{\partial P_{losses}}{\partial P_n} \\ \frac{\partial P_{losses}}{\partial Q_n} \end{bmatrix} = [J^T]^{-1} \begin{bmatrix} \frac{\partial P_{losses}}{\partial \theta} \\ \frac{\partial P_{losses}}{\partial V} \end{bmatrix} \quad (4)$$

5. Implementation of the proposed management algorithm

To evaluate the effect of recharging electric vehicles and the implementation of the proposed management algorithm, a MATLAB application, including a graphical interface was developed. Using this software application, several case studies are simulated.

The system selected to implement the proposed management algorithm corresponds to a typical residential distribution circuit of the city of Bogota with 628 nodes, of which 500 are residential users. It also has 17 distribution transformers and it is characterized by having medium and low voltage underground cables [44].

The limits used as reference to demonstrate the impact of the connection of electric vehicles and to evaluate the performance of the management algorithm are defined below:

- In accordance with the standard NTC 1340 [45] and the standard ANSI C84.1 [46], in normal operating conditions, the voltage for levels 1, 2 and 3 must be between +5% and -10% of the nominal voltage.
- The substation has three 30 MVA - 115/11.4 kV transformers. Each transformer supplies power to eight circuits with a maximum capacity of 3.8 MVA.

In order to predict the behavior of the voltage level and the capacity of the electric network, different simulations are performed to assess the effects on these variables depending on EV penetration. Based on these results, it is possible to

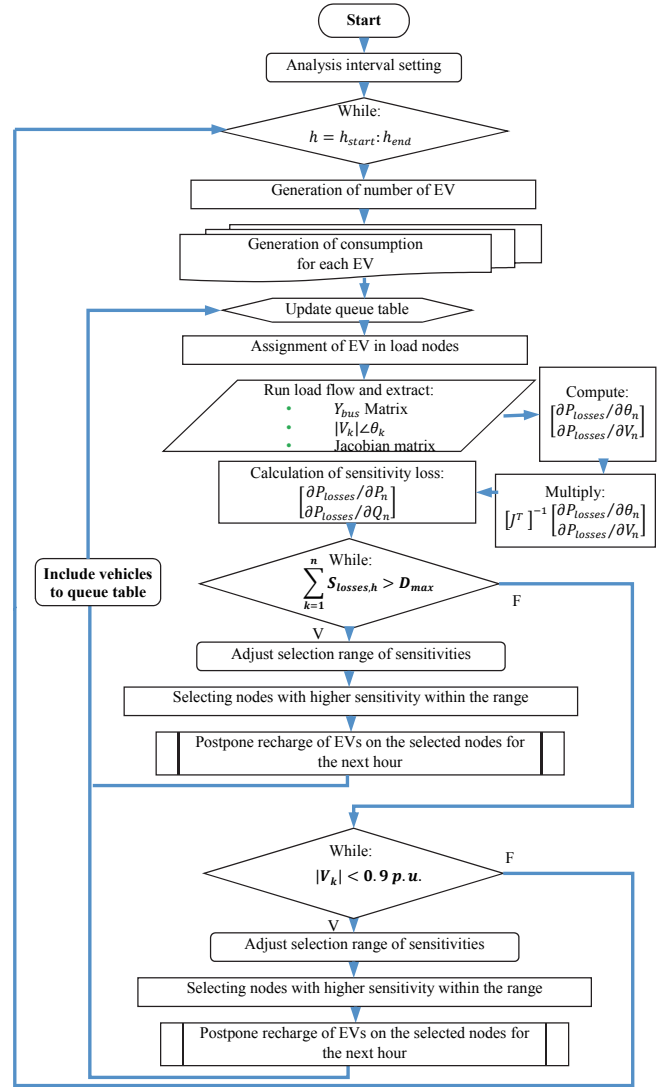


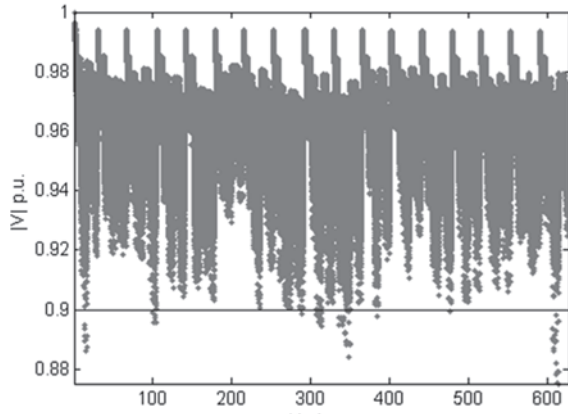
Figure 2. Flowchart of the methodology used by the management algorithm. Source The Authors

predict for a scenario with 38% EV penetration that the voltage levels will be below the established limit, even if the maximum capacity of the circuit is below 3.8 MVA. However, both voltage and capacity are out of limits with 72% EV penetration.

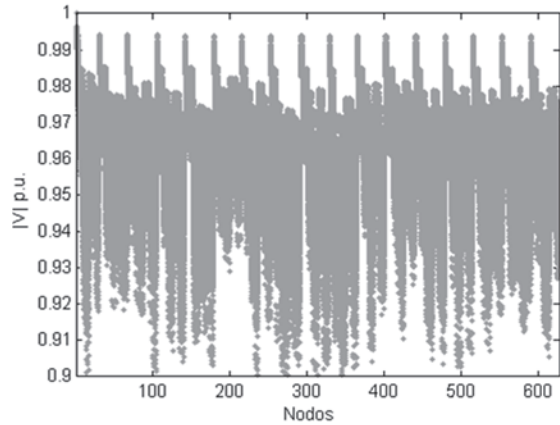
Therefore, in the following sections, the parameters of voltage, maximum capacity, and losses under three EV penetration scenarios will be assessed: 38%, 72% and 100%, also analyzing their behavior when implementing the management algorithm proposed in Section 3.

5.1. Case Study 1: 38% EV penetration

For this case study, 100 simulations corresponding to a day and a 38% EV penetration were performed. Fig. 3 shows the behavior of the voltage level at each node in the distribution system, whereby in (a) we can observe that the lower limit is exceeded reaching values of 0.875 p.u.; however, when applying the proposed management algorithm (b) shows that the voltage drops caused by the recharge are avoided.

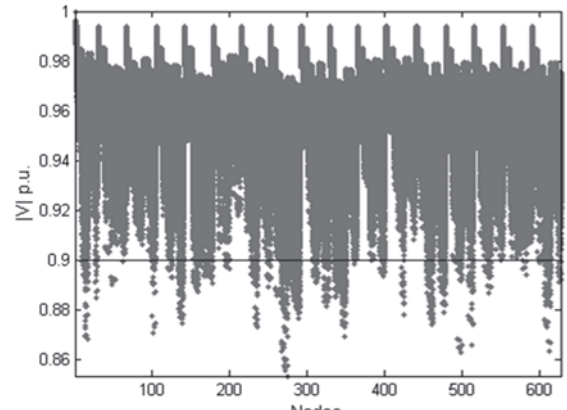


(a)

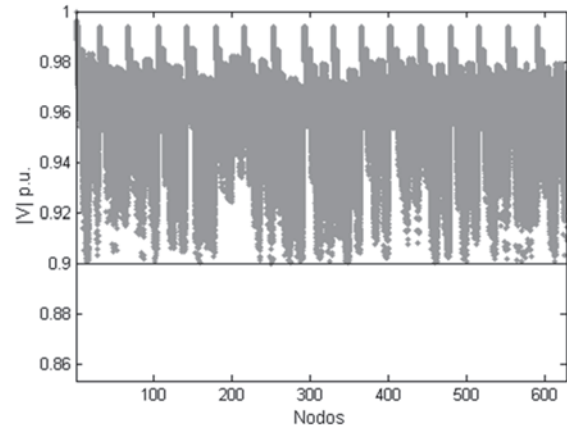


(b)

Figure 3. Behavior of the voltage level with 38% EV penetration level: (a) without any management and (b) with the proposed management algorithm. Source: the Authors



(a)



(b)

Figure 4. Behavior of the voltage level with 72% EV penetration level: (a) without any management and (b) with the proposed management algorithm. Source: The Authors.

When analyzing the maximum capacity assigned to the circuit, it is clear that this variable is not affected with the number of vehicles without any management, and when the management is implemented, a maximum value of 3.3 MVA is reached. As the capacity of the transformer supplying power to the circuit was not exceeded, the power losses are not modified even after implementing the management.

5.2. Case Study 2: 72% of electric vehicles penetration

As in the previous case, 100 simulations corresponding to a day were carried out. The demand was managed with a penetration level of 72%. The obtained data presented in Fig. 4 indicate the voltage level reaches values close to 0.855 p.u. without any control on vehicle recharge (a). However, the management scheme ensures that the minimum limit of 0.9 p.u. is fulfilled regardless of the number of users connected to the system.

On the other hand, Fig. 5 shows that a 72% of electric vehicles' penetration into the electric grid makes the maximum capacity reach values close to 4.2 MVA, representing an overload of up to 400 kVA for the transformer supplying power to the load (a). However, the management scheme avoids this situation, which could endanger system stability and security, postponing the needed recharges taking into account the priority of connection and keeping below the capacity value assigned to the circuit (b).

Evaluating the behavior of power losses, when executing the management algorithm a loss reduction of 25.72% is achieved, while maintaining the voltage level and system capacity, which confirms that the algorithm works.

5.3. Case Study 3: 100% of electric vehicles penetration

Assuming the most critical case, i.e., 100% of electric vehicle penetration, this scenario is simulated 100 times, allowing the voltage level to reach values of close to 0.82 p.u. when the recharges are not managed, meaning they are 8% below the limit established in the regulations. Otherwise, with the operation of the management algorithm the recharge is carried out without exceeding 0.9 p.u.

By comparing the behavior of the system load curves, Fig. 6 shows that 100% EV penetration represents a load close to 4.7 MVA (a), i.e., 500 kVA more than with 72% penetration and 1400 kVA more than in case study 1. By applying the methodology proposed in the paper to manage the recharges, load curves flatten when the maximum power demand occurs ensuring the permitted limit and proper system operation (b).

If technical losses are evaluated against case study 2, Fig. 7 shows an increase of 50 kW when there is no control of the charging time, reaching a value of close to 220 kW (a). In

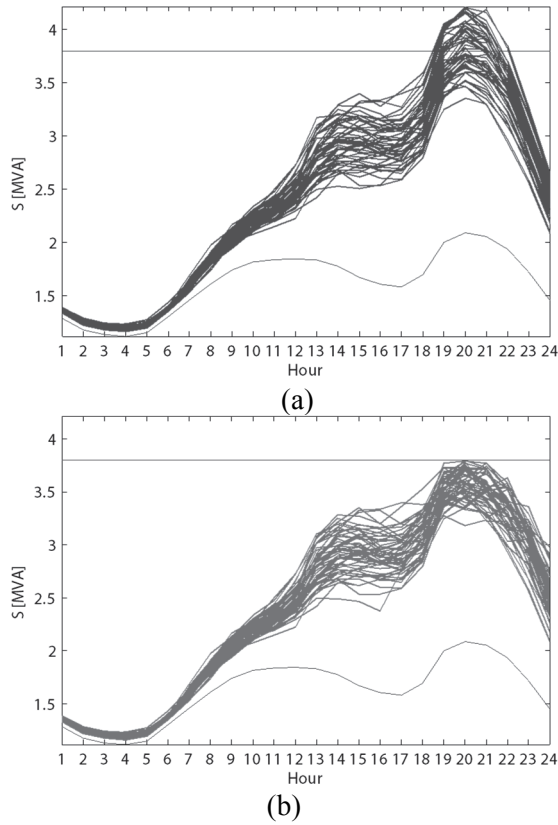


Figure 5. Behavior of load curves with 72% EV penetration level: (a) without any management and (b) with the proposed management algorithm. Source: The Authors.

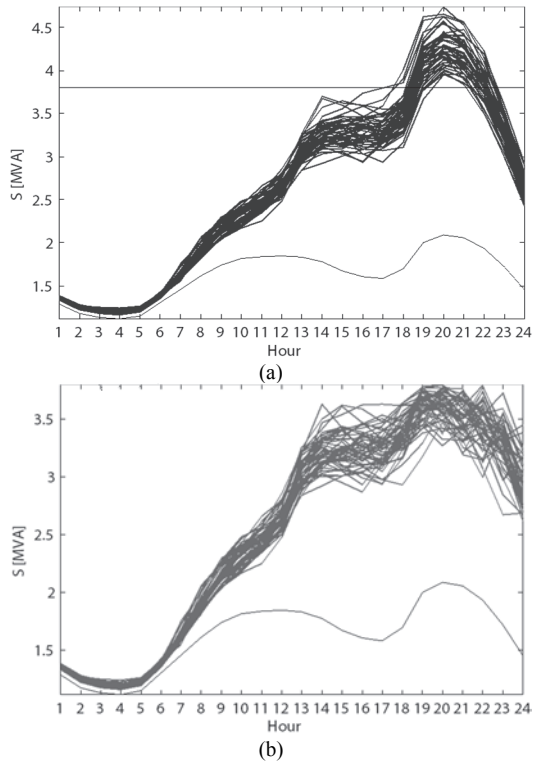


Figure 6. Behavior of load curves in the presence of 100% EV penetration level: (a) without any management and (b) with the proposed management algorithm. Source: The Authors.

addition, when compared to the scenario without electric vehicles connected to the network, it is possible to state that with 100% of electric vehicle penetration, the increase in losses would be equivalent to 3 or 5 times. After managing the battery recharge of electric vehicles connected to the system, a reduction of 40.37% is achieved in the power losses (b), which would mean lower operation costs for the network and the optimization of the objective function under which the algorithm was proposed.

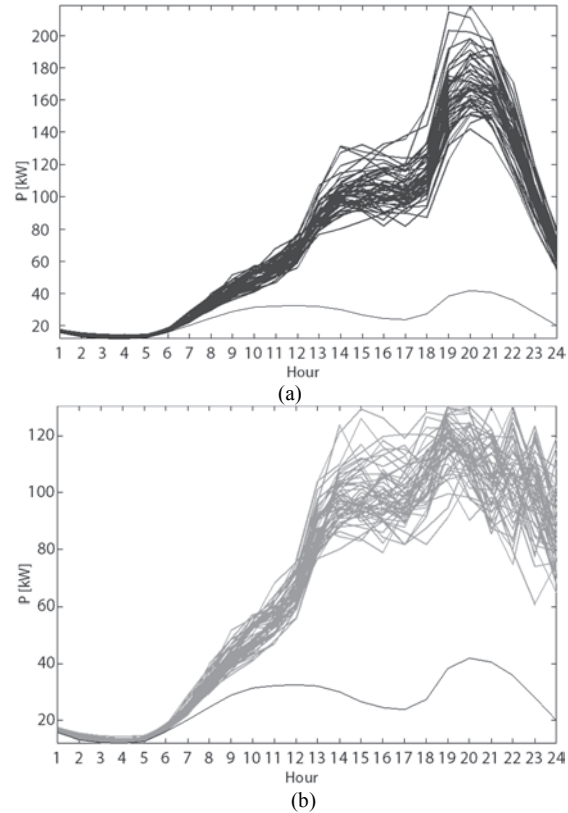


Figure 7. Behavior of the power loss curves with 100% EV penetration level: (a) without any management and (b) with the proposed management algorithm. Source: The Authors.

In order to analyze the effect of the proposed management algorithm on the variables under study, Table 1 summarizes the results obtained in the case studies presented above.

Table 1.
Results from the Case Studies

EV [%]	Minimum Voltage Level [p.u.]		Maximum Capacity [MVA]		Power Losses [kW]		Reduction of power losses [%]
	Without management	With management	Without management	With management	Without management	With management	
38	0.87	0.9	3.3	3.3	10.2	10.2	0
72	0.85	0.9	4.2	3.8	175	130	25.72
100	0.82	0.9	4.7	3.8	220	131.2	40.37

Source: The Authors

Table 2.

Comparison table of computing costs and management algorithm features

Algorithm features	Description
Average time management	16.93 s for 628 nodes including time to generate consumption*
Penetration level	Programmable for any level of penetration with the consumer model.
Distribution system topology	Radial topology for real distribution circuit of 628 nodes with different line lengths [9].
Characteristic of battery charge	Implementation of a variable load curve over time, which depends on the rate of battery recharge, its capacity and energy required to perform a displacement.
Behavior of the sensitivity losses	In contrast to [35], it is proposed to choose the nodes with the highest level of sensitivity within a set range to consider the load variability of line-lengths in the distribution system.

*Intel Core i5 3.4 GHz 16 GB RAM MATLAB R2013b

Source: The Authors

Finally, in Table 2 a comparison of computational costs and features of the proposed management scheme is presented

6. Conclusions

From the obtained results, it is possible to state that when applying the proposed management algorithm, both voltage and capacity were within the allowed limits due to the shift in the recharge of vehicles connected to the nodes with greater loss sensitivity indices. Likewise, the reduction on power losses in the branches are close to 26% in a scenario with 72% EV penetration, but these losses decrease approximately twice with 100% EV penetration, which would represent a reduction in investment, operation, and maintenance costs of networks and equipment for energy distribution companies.

Although the negative impacts on the network can occur at any time and under certain penetration levels, the methodology proposed in this study can be applied to manage any scenario.

The loss sensitivity approach used in the proposed management algorithm allows the coordination of the recharges in a random consumption scenario through minimizing losses, while voltage levels in all nodes significantly improve and peak demand control is carried out, increasing the efficiency and avoiding the need to expand the distribution system infrastructure.

The voltage level is critical in the branches which: a) are at a greater distance from the distribution network transformer, b) are the load nodes farthest from the connected distribution transformer, c) are branches with longer sections.

The impact of massive charging of electric vehicle on power losses is a function of the number of EV connections, the capacity and charging characteristic of the batteries, the parameters of conductors, the sections length and the system topology, as well as the behavior of residential demand.

References

- [1] Godlman, C., Reid, M., Levy R. and Silverstein, A., Coordination of energy efficiency and demand response: A resource of the national action plan for energy efficiency, USA, 2010.
- [2] Ministerio de Industria, Turismo y Comercio, Estrategia integral para el impulso del vehículo eléctrico en España, España, 2010.
- [3] Carvajal, S., Arango, A., Arango, S. and Younes, C., A study of incentives to increase the use of DG in Colombia based on a System Dynamics modeling, *Ingeniería e Investigación*, 31(2), pp. 58-65, 2011.
- [4] Huang, S. and Infield, D., Demand side management for domestic plug-in electric vehicles in power distribution system operation, The 21st International Conference on Electricity Distribution (CIRED), Frankfurt, Paper 0701, 2011.
- [5] Finn, P., Fitzpatrick, C. and Connolly, D., Demand side management of electric car charging: Benefits for consumer and grid, *Energy*, 42(1), pp. 358-363, 2012. DOI: 10.1016/j.energy.2012.03.042
- [6] Vandael, S., Claessens, B., Hommelberg, M., Holvoet, T. and Deconinck, G., A scalable three-step approach for demand side management of plug-in hybrid vehicles, *IEEE Transactions on Smart Grid*, 4(2), pp. 720-728, 2013. DOI: 10.1109/TSG.2012.2213847
- [7] Pieltain, L., Gómez, T., Cossent, R., Domingo, C. and Frías, P., Assessment of the impact of plug-in electric vehicles on distribution networks, *IEEE Transactions on Power Systems*, 26(1), pp. 2066-213, 2011. DOI: 10.1109/TPWRS.2010.2049133
- [8] Kazerooni, M., New load demand for electric vehicles and its harmonic impacts on power system distribution transformers, Msc. Thesis, Department of Electrical Engineering, University of Windsor, Ontario, Canada, 2012.
- [9] Grahn, P., Electric vehicle charging impact on load profile, BSc. Thesis, Department of Electrical Engineering, Royal Institute of Technology, Stockholm, Sweden, 2013.
- [10] Mendoza, C., Quintero, A., Santamaria, F. and Alarcon, J., Impact analysis of plug-in electric vehicles on an electric distribution network, *Transmission & Distribution Conference and Exposition - Latin America*, Medellín, 2014. DOI: 10.1109/TDC-LA.2014.6955258
- [11] Gerkensmeyer, C., Kintner-Meyer, M. and DeSteele, J., Technical challenges of plug-in hybrid electric vehicles and impacts to the US power system: Distribution system analysis, United States Department of Energy, USA, 2010.
- [12] De Nigris, M., Gianinoni, S., Grillo, S., Massucco, S. and Silvestro, F., Impact evaluation of plug-in electric vehicles (PEV) on Electric Distribution networks, 14th International Conference on Harmonics and Quality of Power (ICHQP), Bergamo, Italy, 2010. DOI: 10.1109/ICHQP.2010.5625433
- [13] Yao, W., Guo, Q., Hongbin, S. and Zhengshuo, L., An investigation into the impacts of the crucial factors on EVs charging load, *de IEEE PES ISGT Asia*, Asia, 2012. DOI: 10.1109/ISGT-Asia.2012.6303111
- [14] Gómez, J.C. and Morcos, M.M., Impact of EV battery chargers on the power quality of distributed systems, *IEEE Transactions Power Delivery*, 18(3), pp. 975-981, 2003. DOI: 10.1109/TPWRD.2003.813873
- [15] Basu, M., Gaughan, K. and Coyle, E., Harmonic distortion caused by EV battery chargers in the distribution systems network and its remedy, 39th Universities Power Engineering Conference (UPEC), 2, pp. 869-873, Bristol UK, 2004.
- [16] Jiang, C., Torquato, R., Salles, D. and Xu, W., Method to assess the power quality impact of plug-in electric vehicles, 2014 IEEE 16th International Conference on Harmonics and Quality of Power (ICHQP), pp. 170-180, Bucharest, Romania, 2014. DOI: 10.1109/ICHQP.2014.6842835
- [17] Niitsoo, J., Taklaja, P., Palu, I. and Kiitam, I., Modelling EVs in residential distribution grid with other nonlinear loads, 2015 IEEE 15th International Conference on Environment and Electrical Engineering (EEEIC), pp. 1543-1548, Rome, 2015. DOI: 10.1109/EEEIC.2015.7165401
- [18] Rosso, Á., Evaluación del impacto de los vehículos eléctricos en las redes de distribución, BSc. Tesis, Departamento de Ingeniería Industrial, Universidad Pontificia Comillas, Madrid, España, 2010.
- [19] Grahn, P., Rosenlind, J., Hilber, P., Alvehag, K. and Soder, L., A method for evaluating the impact of electric vehicle charging on

- transformer hotspot temperature, 2nd IEEE PES International Conference and Exhibition on Innovative Smart Grid Technologies (ISGT Europe), Manchester, United Kingdom, 2011. DOI: 10.1109/ISGTEurope.2011.6162755
- [20] Sharma, I., Canizares, C. and Bhattacharya, K., Smart charging of PEVs penetrating into residential distribution systems, IEEE Transactions on Smart Grid, 5(3), pp. 1196-1209, 2014. DOI: 10.1109/TSG.2014.2303173
- [21] Leou, R.C., Su, C.-L. and Lu, C.-N., Stochastic analyses of electric vehicle charging impacts on distribution network, IEEE Transactions on Power Systems, 29(3), pp. 1055-1063, 2014. DOI: 10.1109/TPWRS.2013.2291556
- [22] Tie, C.H., Gan, C.K. and Ibrahim, K., The impact of electric vehicle charging on a residential low voltage distribution network in Malaysia, 2014 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia), pp. 272-277, Kuala Lumpur, 2014. DOI: 10.1109/ISGT-Asia.2014.6873802
- [23] Neagoe-Stefana, A., Eremia, M., Toma, L. and Neagoe, A., Impact of charging electric vehicles in residential network on the voltage profile using Matlab, 2015 9th International Symposium on Advanced Topics in Electrical Engineering (ATEE), pp. 787-791, Bucharest, Romania, 2015. DOI: 10.1109/ATEE.2015.7133909
- [24] Geth, F., Leemput, N., Van Roy, J. and Buscher, J., Voltage droop charging of electric vehicles in a residential distribution feeder, 3rd IEEE PES International Conference and Exhibition on Innovative Smart Grid Technologies (ISGT Europe), Berlin, Germany, 2012. DOI: 10.1109/ISGTEurope.2012.6465692
- [25] Rodríguez-García, L.F., Pérez-Londoño, S.M. and Mora-Flórez, J.J., Load area aggregation considering integration of electric vehicles to the system, Revista Ingeniería e Investigación, 35(Supl.1), pp. 42-49, 2015. DOI: 10.15446/ing.investig.v35n1Sup.53652
- [26] International Energy Agency, 2014 Annual Report, IEA, Paris, 2014.
- [27] Salah, F., Ilg, J.P., Flath, C.M., Bassec, H. and van Dinther, C., Impact of electric vehicles on high-voltage grids: A Swiss case study, Switzerland, 2013. DOI: 10.2139/ssrn.2209841
- [28] Stoeckl, G., Witzmann, R. and Eckstein, J., Analyzing the capacity of low voltage grids for electric vehicles, IEEE Electrical Power and Energy Conference (EPEC), pp. 415-420, Winnipeg, MB, Canada, 2011. DOI: 10.1109/EPEC.2011.6070236
- [29] Shafiee, S., Fotuhi-Firuzabad, M. and Rastegar, M., Investigating the impacts of plug-in hybrid electric vehicles on power distribution systems, IEEE Transactions on Smart Grid, 4(3), pp. 1351-1360, 2013. DOI: 10.1109/TSG.2013.2251483
- [30] Huachun, H., Haiping, X., Zengquan, Y. and Yingjie, Z., Interactive charging strategy of electric vehicles connected in smart grids, IEEE 7th International Power Electronics and Motion Control Conference (IPEMC), 3, pp. 2099-2103, Harbin, China, 2012. DOI: 10.1109/IPEMC.2012.6259168
- [31] Shao, S., Pipattanasomporn, M. and Rahman, S., Grid integration of electric vehicles and demand response with customer choice, IEEE Transactions on Smart Grid, 3(1), pp. 543-550, 2012. DOI: 10.1109/TSG.2011.2164949
- [32] Wu, D., Aliprantis, D. and Ying, L., Load scheduling and dispatch for aggregators of plug-in electric vehicles, IEEE Transactions on Smart Grid, 3(1), pp. 368-376, 2012. DOI: 10.1109/TSG.2011.2163174
- [33] Xi, X. and Sioshansi, R., Using price-based signals to control plug-in electric vehicle fleet charging, IEEE Transactions on Smart Grid, 5(3), pp. 1451-1464, 2014. DOI: 10.1109/TSG.2014.2301931
- [34] Shaaban, M., Ismail, M., El-Saadany, E. and Zhuang, W., Real-Time PEV Charging/Discharging coordination in smart distribution systems, IEEE Transactions on Smart Grid, 5(4), pp. 179-1807, 2014. DOI: 10.1109/TSG.2014.2311457
- [35] Guo, Y., Xu, S. and Su, W., Smart management system of plug-in electric vehicle charging service, 2015 IEEE Transportation Electrification Conference and Expo (ITEC), Detroit, USA, 2015. DOI: 10.1109/ITEC.2015.7165815
- [36] Deilami, S., Masoum, A., Moses, P. and Masoum, M., Real-Time coordination of Plug-In electric vehicle charging in smart grids to minimize power losses and improve voltage profile, IEEE on Smart Grids, 2(3), pp. 456-467, 2011. DOI: 10.1109/TSG.2011.2159816
- [37] Peng, H., Jinkuan, W. and Yinghua, H., Dynamic-priority-based real-time charging management for plug-in electric vehicles in smart grid, 2012 31st Chinese Control Conference (CCC), pp. 2432-2436, China, 2012.
- [38] Singh, M., Thirugnanam, K., Kumar, P. and Kar, I., Real-Time coordination of electric vehicles to support the grid at the distribution substation level, IEEE Systems Journal, 9(3), pp. 1000-1010, 2015. DOI: 10.1109/JSYST.2013.2280821
- [39] Yao, K.C. and Yao, W.T., New network sensitivity-based approach for real-time complex power flow calculation, IET Generation, Transmission & Distribution, 6(2), pp. 109-120, 2011. DOI: 10.1049/iet-gtd.2011.0429
- [40] Bhattacharyya, B. and Goswami, S.K., Sensitivity based evolutionary algorithms for reactive power dispatch, Third International Conference on Power Systems, Kharagpur, 2009. DOI: 10.1109/ICPWS.2009.5442664
- [41] Bingliang, Z., Yutian, S., Li, B. and Li, L.J., A modeling method for the power demand of electric vehicles based on Monte Carlo simulation, Power and Energy Engineering Conference (APPEEC), Shanghai, 2012. DOI: 10.1109/APPEEC.2012.6307215
- [42] Khatod, D.K., Pant, V. and Sharma, J., A novel approach for sensitivity calculations in the radial distribution system, IEEE transactions on Power Delivery, 21(4), pp. 2048-2057, 2006. DOI: 10.1109/TPWRD.2006.874651
- [43] Peng, Y., Shun, H., Zou, J. and Wang, D., Improved method based on second order loss sensitivity for voltage collapse, International Conference on Electrical Engineering (ICEE), 2005.
- [44] CODENSA, Generalidades de la Norma LA, Bogotá, [En línea] 2009. Disponible en: <http://likinormas.micodensa.com/>
- [45] ICONTEC, NTC 1340: Tensiones y frecuencia nominales en sistemas de energía eléctrica en redes de servicio público, Segunda Ed., Instituto Colombiano de Normas Técnicas y Certificación (ICONTEC), Bogotá D.C., 2004.
- [46] American National Standards Institute, Inc., ANSI C84.1 - 2011: For Electric Power Systems and Equipment - Voltage Ratings (60 Hz), National Electrical Manufacturers Association, United States of America, 2011.

C.C. Mendoza-Grosso, completed his BSc. Eng in Electrical Engineering at Universidad Distrital Francisco José de Caldas, Bogotá, Colombia, in 2013. Since 2012, he has been member of the Electromagnetic Compatibility and Interference Research Group at the Universidad Distrital Francisco José de Caldas (GCEM). He works in the electric sector as an electrical engineer in the high voltage area.
ORCID: 0000-0002-8017-5374

A.M. Quintero-Montenegro, completed her BSc. Eng Electrical Engineer at Universidad Distrital Francisco José de Caldas, Bogotá, Colombia, in 2013 and her degree in Telecommunications Technology at SENA – Servicio Nacional de Aprendizaje, Bogotá, in 2005. Since 2012, she has been a member of the Electromagnetic Compatibility and Interference Research Group at the Universidad Distrital Francisco José de Caldas (GCEM). She has experience in the area of telecommunications in oil companies, and control and protections in high voltage Substations. She works in the electric sector as an electrical engineer in the extra high voltage and high voltage substations design.
ORCID: 0000-0002-9794-4870

F. Santamaria, completed his BSc, MSc and PhD degrees in Electrical Engineering at Universidad Nacional de Colombia in 2002, 2007 and 2013, respectively. From 2004 to 2010 he was a full time research engineer with the Electromagnetic Compatibility Research Group at the Universidad Nacional de Colombia (EMC-UNC) where he carried out research on the lightning phenomena, electromagnetic compatibility, gas discharges and pulsed power systems. In 2010, he joined the staff of the Department of Electrical Engineering at Universidad Distrital Francisco Jose de Caldas, where he currently works as an associate professor and member of the Electromagnetic Compatibility and Interference Research Group (GCEM). His research interests include experimental studies in gas discharges, lightning protection, electromagnetic interference, electric vehicles and distributed generation.
ORCID: 0000-0002-0391-4508

J.A. Alarcón-Villamil, completed his BSc. and MSc. degrees in Electrical Engineering at the Universidad Nacional de Colombia, Bogotá, in 2002 and 2013, respectively. From 2004 to 2008, he was a research engineer with the Electromagnetic Compatibility Research Group, at Universidad Nacional de Colombia (EMC-UNC), where he investigated the corona effect, electromagnetic compatibility, dielectric materials, and surge protection devices. In 2010, he joined the Department of Electrical Engineering, Universidad Distrital Francisco José de Caldas, where he is currently an Assistant Professor and Member of the Electromagnetic Compatibility and Interference Research Group (GCEM). His research interests include high-voltage engineering, electric vehicles and smart grids, renewable energies, and electromagnetic compatibility.
ORCID: 0000-0002-8718-2542



UNIVERSIDAD NACIONAL DE COLOMBIA

SEDE MEDELLÍN
FACULTAD DE MINAS

Área Curricular de Ingeniería
Eléctrica e Ingeniería de Control

Oferta de Posgrados

Maestría en Ingeniería - Ingeniería Eléctrica

Mayor información:

E-mail: ingelcontro_med@unal.edu.co
Teléfono: (57-4) 425 52 64

Caracterización físicoquímica de cuatro especies de agaves con potencialidad en la obtención de pulpa de celulosa para elaboración de papel

Edith Jiménez-Muñoz ^a, Francisco Prieto-García ^{a*}, Judith Prieto-Méndez ^b, Otilio A. Acevedo-Sandoval ^b
& Rodrigo Rodríguez-Laguna ^b

^aInstituto de Ciencias Básicas e Ingeniería. Universidad Autónoma del Estado de Hidalgo, Pachuca, México. *Autor de correspondencia.
prietog@uaeh.edu.mx, bioedith13@gmail.com

^bInstituto de Ciencias Agropecuarias. Universidad Autónoma del Estado de Hidalgo, Pachuca, México. Jud_292003@yahoo.com.mx,
acevedo@uaeh.edu.mx, rodris71@yahoo.com

Received: July 31th, 2015. Received in revised form: April 22th, 2016 Accepted: April 27th, 2016

Resumen

El objetivo general de este trabajo fue encontrar un nuevo valor agregado a las hojas de las agaváceas, desecho de la agroindustria. Se estudiaron cuatro especies de agaves de tres estados de la república mexicana: *Agave angustifolia* (Oacaxa), *Agave tequilana* (Jalisco), *Agave salmiana* y *Agave lechuguilla* (Hidalgo). Luego de una clasificación taxonómica, fenológica, ecológica y etnobotánica (basados en morfología comparada), se sometieron a diversos análisis físicos (humedad, molienda, tamizaje, granulometría y características foliares, MEB) y químicos (cenizas grasas y aceites, proteínas, análisis TGA y DTGA, así como FTIR). Las especies de *A. angustifolia* y *A. tequilana* son las que presentan un óptimo tamaño de fibra para la elaboración de papel. Corresponde también a los mayores porcentajes de celulosa (44.3% y 36.8%) por DTGA. Del análisis proximal, se evidencia que las dos especies definidas como mejores para la obtención de pulpa de celulosa, presentan pequeñas diferencias significativas en contenidos de grasas, proteínas, material solubilizable en agua caliente (AT) como en solución moderadamente alcalina y/o alcohólica.

Palabras clave: Agave ssp., análisis proximal, DTGA, FTIR, celulosa.

Physicochemical characterization of four species of agaves with potential in obtaining pulp for paper making

Abstract

The overall objective of this work was to find a new value to the leaves of agaváceas, waste of agribusiness. Four species of agaves were studied in three states of the Mexican Republic: *Agave angustifolia* (Oacaxa), *Agave tequilana* (Jalisco), *Agave salmiana* and *Agave lechuguilla* (Hidalgo). After a classification taxonomy, phenology, ecological and ethnobotanical (morphology), various physical analysis (moisture, grinding, screening, grading and leaf characteristics, SEM) and chemical (fat and oils, ash, proteins, TGA and DTGA and FTIR were subjected). The species of *A. angustifolia* and *A. tequilana* are those with an optimum size of fiber for papermaking. It also corresponds to the highest percentages of cellulose (44.3% and 36.8%) by DTGA. The proximate analysis, it appears that the two species defined as obtaining better for pulp, significant differences have small content of fat, protein, water solubilizable hot material (AT) and moderately alkaline and/or alcoholic solution.

Keywords: Agave ssp., proximate analysis, DTGA, FTIR, cellulose.

How to cite: Jiménez-Muñoz, E., Prieto-García, F., Prieto-Méndez, J., Acevedo-Sandoval, O.A. y Rodríguez-Laguna, R., Caracterización físicoquímica de cuatro especies de agaves con potencialidad en la obtención de pulpa de celulosa para elaboración de papel. DYNA 83(197), pp. 233-243, 2016

1. Introducción

La celulosa, principal componente de las paredes celulares de las plantas, es un polisacárido formado por moléculas de cadenas lineales de $\beta(1,4)$ -D-glucopiranosas, que tienden a formar microfibrillas con diámetros de 1 a 10 nm y longitudes de 25 nm. Es el polímero más abundante de la naturaleza y utilizado como materia prima en industrias de papel y textil [1].

Es conocida la producción y obtención de celulosa, a partir de fuentes maderables, y existe una tendencia actual a la búsqueda de nuevas fuentes de obtención de celulosa. El agave es una especie no maderable que puede ser utilizada para extracción de celulosa. Existen diversas investigaciones sobre elaboración de fibras a base de residuos de agaves [2] en las cuales obtuvieron pulpa de bagazo de *A. tequilana* mediante técnicas mecánicas, químicas y biológicas. Recientemente se reporta [3] la elaboración de papel a partir de fibras crudas de seis especies de *Agave*; tres de hoja angosta (*A. lechuguilla*, *A. angustifolia* y *A. tequilana*) y tres de hoja ancha (*A. americana*, *A. salmiana* y *A. mapisaga*), utilizando diferentes métodos para la extracción de fibras. El autor señala que *A. salmiana* y *A. mapisaga* fueron las especies que presentaron mejores características. El objetivo general de este estudio, fue encontrar un nuevo valor agregado a las hojas residuales de especies de agaves de la agroindustria tequilera, mezcalera y/o pulquera, y su potencialidad para obtener pulpa de celulosa como opción para la elaboración de papel de fibras cortas.

2. Metodología

2.1. Zona de estudio y toma de muestras

Las especies muestreadas aparecen en la Tabla 1. *A. tequilana* Weber, Var. Azul, sólo es cultivable en los estados mexicanos de Jalisco, Guanajuato, Michoacán, Nayarit y Tamaulipas, a pesar de que se adapta a un amplio rango altitudinal [4]; *A. tequilana* se desarrolla óptimamente de 700 a 1900 msnm [5].

2.2. Realización de ensayos

Las muestras fueron sometidas a determinación de humedad, posteriormente se molieron y tamizaron.

Se evaluaron sus dimensiones y características foliares, se determinó distribución y tamaños de partículas de las fibras obtenidas, mediante un analizador por difracción de rayos Laser en equipo LS13-320 de la firma Beckman Coulter ®. La muestra fue suspendida en agua destilada y, en agitación continua, se midieron los diferentes tamaños y distribución. Las muestras secas fueron analizadas por Microscopía Electrónica de Barrido (MEB) para precisar morfologías de las fibras. Se realizaron análisis termogravimétrico (TGA) y

derivativo (DTGA) y de espectroscopía infrarroja (FTIR). El análisis morfológico se realizó en un MEB de la firma JEOL, modelo JSM-820®. Los análisis térmicos se llevaron a cabo en analizador METTLER-TOLEDO®, modelo TGA/SDTGA-851, crisoles de aluminio y un barrido de 25°C a 600°C a velocidad de 10°C/min. FTIR se realizó acorde a [6] en muestras secas; analizadas en pastillas con KBr en equipo Perkin Elmer®, entre 400-4000 cm^{-1} y 4 cm^{-1} de resolución.

La concentración de azúcares totales se determinó por el método Fenol-Sulfúrico [5]. Los extractos fueron obtenidos a partir de 2 g de fracción de hojas de agaves, secadas y trituradas en un vaso de precipitados (maceradas), se les añadió agua destilada relación 1:6 (o en su caso, solución de NaOH 1 % y/o solución de etanol-agua 1:1 con calentamiento y agitación a 70°C por 30 minutos. Al término de 30 minutos, se centrifugó a 5000 rpm y el líquido claro se reservó como extracto de fracciones solubles. Se determinaron grasas y aceites, proteínas y lignina [5, 6].

3. Resultados y discusión

3.1. Caracterización física

Se observa en la Tabla 2 los resultados de caracterización de muestras de hojas de *A. tequilana* Weber, Var. Azul. Por sus dimensiones (98,58 cm de largo, 8,33 cm de ancho y 2,08 cm de espesor) se asemejan a lo reportado por algunos autores [7]; se puede analizar que en sus espesores resultaron ligeramente más gruesas y por tanto con mayores volúmenes y densidades volumétricas, lo cual da una idea de que las mismas se encuentran algo pasadas en madurez para el jimado (proceso de corte de las hojas para quedar sólo la piña central), también puede ser debido a que la recolección de las piñas se realizó en época de lluvia.

De igual forma la altura de las plantas se considera similar a lo reportado [7] aunque ligeramente menores en diámetro de la roseta. Los contenidos de humedad en las hojas resultaron un poco más elevadas (83,43%) con relación a lo que reportan [8], debido a haberse recolectado en época lluviosa. Se puede considerar un poco intenso el color verde de las hojas, al ser comparados en una escala de 1-6 con un valor promedio de 4,31, y según establecen [9], este valor debe estar entre 3-4 como indicativo de una suficiente madurez de la planta, que ha de corresponder a una edad entre 8-10 años (edad óptima para la jima (proceso de corte de las hojas para quedar sólo la piña central) y producción de tequila). Si los colores son menores de 3 la escala, presupone no madurez suficiente. En cuanto a la altura de la espina apical, se corresponde con lo reportado por otros autores [9,10] pero un poco más pequeña que lo que indica [8].

Tabla 1.
Localización geográfica de los puntos de toma de muestras de agavaceas

Variedad	Localidad; Estado	Coordenadas	Altitud (msnm)
<i>Agave angustifolia</i> Haw.	Zimatlán, Oaxaca	16°45'05.4" N y 96°49'47.8" O	1654
<i>Agave lechuguilla</i> Torr.	San Agustín Tlaxiaca, Hidalgo	20°05'26.5" N y 98°52'19.5" O	2853
<i>Agave salmiana</i> B. Otto ex Salm-Dick	Epazoyucan, Hidalgo	20°00'02.5" N y 98°36'28.9" O	2220
<i>Agave tequilana</i> Weber, Var. Azul	Tequila, Jalisco	21°12'30" N y 103°36'00" O	850

Elaboración propia

Tabla 2.

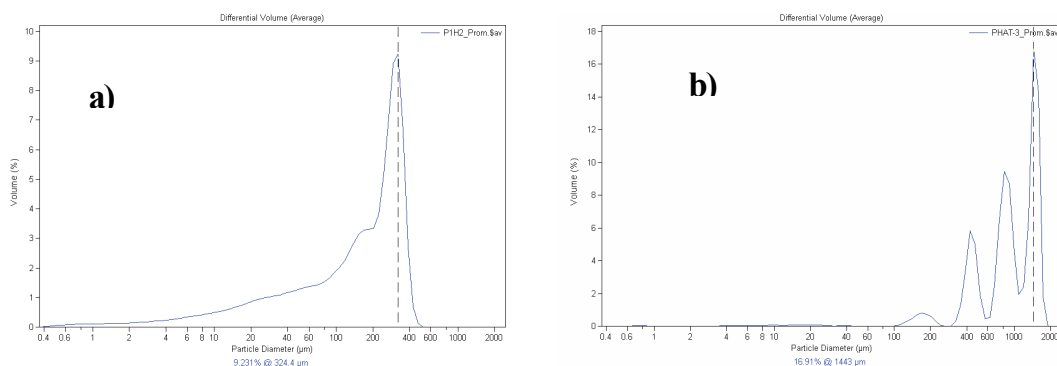
Caracterización física de *Agave tequilana* Weber, Var. Azul, tomada en el municipio de Tequila, Jalisco.

Hoja	Hoja (n=10 hojas)					Roseta (n=6 plantas)			
	largo (cm)	ancho (cm)	espesor (cm)	masa (g)	Volumen (cm ³)	densidad (g/cm ³)	Área foliar (cm ²)	Altura Planta (mm)	Diam. Planta (mm)
media	98,58	8,33	2,08	641,64	1741,92	0,38	827,59	1417,43	1469,48
desv.									
Est.	9,011	0,833	0,207	149,094	439,885	0,065	152,652	40,044	57,931
%CV	9,14	10,00	9,95	23,24	25,25	17,20	18,45	2,83	3,94
t cal	2,055	2,187	7,341	1,403	3,896	5,120	0,240	1,377	18,042
Dif. Sig	NO	NO	SI	NO	SI	SI	NO	NO	SI
Ref [11]	100,00	7,50	1,60	575,47	1200,00	0,48	816,00	1400,00	1800,00

Hoja	Humedad (%)	<0,5 mm (%)	0,5-1,0 mm (%)	>1,0mm (%)	Color (1-6)	Nº Espinas	Alt. E. apical (cm)
media	83,43	13,62	84,83	1,56	4,3	151,05	1,51
desv. Est.	3,405	5,315	5,914	0,870	0,458	21,650	0,070
%CV	4,08	39,03	6,97	55,86	10,66	14,33	4,64
t cal	3,182	-	0,091	-	2,070	0,153	22,136
Dif. Sig	SI	-	NO	-	NO	NO	SI
Ref [11]	80,00	NR	> 85%	NR	4,00	150,00	2,00

Nota t tabulada: 2,26; cuando t cal < t tabulada no hay diferencias significativas, NR: No Reportado.

Fuente: Elaboración propia

Figura 1. Distribución y tamaños de partículas secas de la a) fracción minoritaria (<0,5 mm) y b) fracción 0,5-2,0 mm, de hojas de *Agave tequilana* Weber, Var. Azul en el municipio de Tequila, estado de Jalisco, México.

Fuente: Elaboración propia.

Se analizó la composición por fracciones de tamaños de partículas; se obtuvo 98,5% < de 1 mm siendo la fracción mayoritaria entre 0,5-1,0 mm (84,3%) lo cual se corresponde

con lo reportado por [12]. La fracción < 0,5 mm varía entre 8-10% y se obtuvo un valor promedio ligeramente superior (13,6%). En la Fig. 1a se observa el tamaño y distribución de

Tabla 3.

Caracterización física de *Agave salmiana* B. Otto ex Salm-Dick, tomada en el municipio de Epazoyucan, Hidalgo.

Hoja	Hoja (n=10 hojas)					Roseta (n=6 plantas)			
	largo (cm)	ancho (cm)	espesor (cm)	masa (g)	Volumen (cm ³)	densidad (g/cm ³)	Área foliar (cm ²)	Altura Planta (mm)	Diam. Planta (mm)
media	159,00	21,23	6,54	9179,36	22821,48	0,43	3397,78	2236,17	2488,83
desv.				2281,69					
Est.	15,593	1,555	1,248	3	7474,732	0,125	567,370	161,310	336,553
%CV	9,81	7,33	19,09	24,86	32,75	28,71	16,70	7,21	13,52
t cal	1,825	6,659	5,165	1,912	2,632	0,903	1,545	4,630	2,714
Dif. Sig	NO	SI	SI	NO	SI	NO	NO	SI	SI
Ref [10]	150,00	24,50	4,50	7800,00	16600,00	0,47	3675	2000	2200

Hoja	Humedad (%)	<0,5 mm (%)	0,5-1,0 mm (%)	>1,0mm (%)	Color (1-6)	Nº Espinas	Alt. E. apical (cm)
media	92,31	25,16	70,18	4,66	3,75	111,40	5,85
desv. Est.	1,886	3,119	4,723	2,520	0,403	8,888	0,709
%CV	2,04	12,40	6,73	54,07	10,75	7,98	12,12
t cal	12,255	-	9,923	-	1,961	4,056	0,669
Dif. Sig	SI	-	SI	-	NO	SI	NO
Ref [10]	85	NR	> 85%	NR	4	100	6

Nota t tabulada: 2,26; cuando t cal < t tabulada no hay diferencias significativas, NR: No Reportado.

Fuente: Elaboración propia

partículas de la fracción más fina (<0,5 mm). Como se puede apreciar el tamaño prevaleciente fue de 324,4 μm y muestra una distribución casi homogénea. En la Fig. 1b se observa el tamaño y distribución de partículas de la fracción 0,5-2,0 mm. El tamaño prevaleciente fue de 1443 μm .

La Tabla 3 refleja los resultados de caracterización física de la variedad *A. Salmiana* Otto ex Salm-Dyck. No existen diferencias significativas en cuanto al largo de sus hojas y corresponde con lo reportado algunos autores [5,10-12] que señalan valores de 150 cm. Sin embargo el valor promedio en el ancho de las hojas (21,23 cm) resultó ligeramente menor que lo reportado algunos de los autores de 24cm [10, 11]. Una variación importante espesor de las hojas que en promedio fue de 6,5 cm, muy superior a lo reportado (4.5 cm) [12]. La masa promedio se encontró que también fue superior (9179 g) a lo esperado, atribuido a la mayor cantidad de agua retenida, debido a la época de lluvias y correspondiendo con los mayores anchos y espesores de las hojas que resultaron 1,37 veces más voluminosas que lo que se reporta en las referencias consultadas (1600-1700 cm^3).

Se observó mayor esbeltez de la planta y mayor diámetro de la roseta en surco, aproximadamente unos 200 cm más altas y más anchas las rosetas.

La humedad promedio (92,3%) en la variedad *A. Salmiana* presentó un 7% por encima de los valores que reportan [10], debido probablemente a la temporada de lluvias.

Los tamaños y distribución de partículas de las fracciones minoritarias (<0,5 mm) y mayoritarias (0,5-1,0 mm), se muestran en las Fig. 2a y 2b respectivamente; para la fracción mayoritaria fue de un 70.8%. Se observa en la Fig. 2b distribución del 25,2% de tamaño de partículas <0,5 mm que se espera igualmente no afecte demasiado en el pulpeo). En la Fig. 2b se aprecia una distribución casi homogénea con tamaño de partícula predominante entre 0.5 mm y 1.9 mm.

En cuanto al color de las hojas, se consideraron con edad de madurez adecuada (valores entre 3-4, según escala de color de [10]) con valor promedio de 3.70. Presentaron un número mayor de espinas laterales (111 como promedio) que lo reportado (≈ 100) por [12], sin embargo la altura promedio de la espina apical (5,85 cm) si se corresponde con el valor que estos mismos autores refieren ($\approx 6,0$ cm).

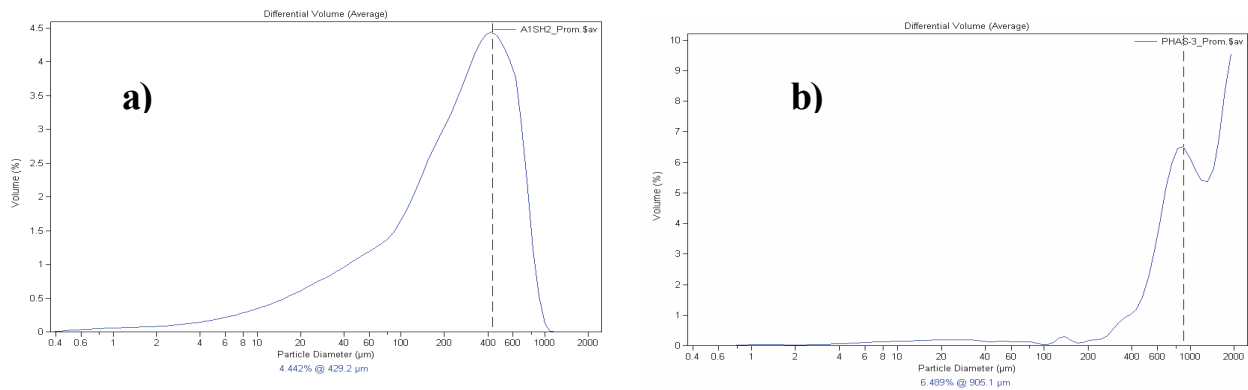


Figura 2. Distribución y tamaños de partículas secas de la a) fracción minoritaria (<0,5 mm) y b) fracción 0,5-2,0 mm, de hojas de *Agave salmiana* B. Otto ex Salm-Dick, tomada en el municipio de Epazoyucan, Hidalgo, México.

Fuente: Elaboración propia

Tabla 4.

Caracterización física de *Agave angustifolia* Haw, tomada en el municipio de Zimatlán, Oaxaca.

Hoja	Hoja (n=10 hojas)					Roseta (n=6 plantas)			
	largo (cm)	ancho (cm)	espesor (cm)	masa (g)	Volumen (cm^3)	densidad (g/cm^3)	Área foliar (cm^2)	Altura Planta (mm)	Diam. Planta (cm)
media	102,00	6,16	2,12	427,67	1349,22	0,313	636,45	1484,17	2225,35
desv.									
Est.	9,754	0,650	0,087	116,497	282,172	0,030	126,335	130,539	193,818
%CV	9,56	10,54	4,09	27,24	20,91	9,46	19,85	8,80	8,71
t cal	7,456	1,650	4,260	2,235	1,112	10,127	2,101	2,806	3,677
Dif. Sig	SI	NO	SI	NO	NO	SI	NO	SI	SI
Ref [13]	125,00	6,50	2,00	510,00	1250,00	0,408	552,50	1600,00	2000

Hoja	Humedad (%)	<0,5 mm (%)	0,5-1,0 mm (%)	>1,0mm (%)	Color (1-6)	Nº Espinas	Alt. E. apical (cm)
media	72,87	19,89	65,17	5,69	4,0	79,75	2,42
desv. Est.	6,593	5,825	6,366	2,618	0,150	3,894	0,1166
%CV	9,05	29,29	9,77	46,05	3,80	4,88	4,82
t cal	1,375	-	4,883	-	1,054	0,203	2,169
Dif. Sig	NO	NR	SI	NR	NO	NO	NO
Ref [13]	70	-	75	-	4	80	2,5

Nota t tabulada: 2,26; cuando t cal < t tabulada no hay diferencias significativas, NR: No Reportado.

Fuente: Elaboración propia

En la Tabla 4 se muestran los resultados de la caracterización física de la variedad *A. angustifolia* Haw. Se aprecia que existen diferencias significativas en cuanto largo y espesor de las hojas con relación a lo reportado por otros autores [4,13]; en este trabajo se obtuvo hojas de 102 cm de largo, más cortas que lo reportado (125 cm), sin embargo ligeramente mayor espesor (2,12 cm) con relación a lo que se reporta (2,0 cm) por los anteriores autores. Por otro lado la densidad de estas hojas fue menor (0,313 g/cm³). En relación a las demás mediciones sólo se aprecian menores valores en cuanto a la altura y diámetro de la roseta de las plantas.

La humedad se reporta ligeramente por encima de lo señalado [13], probablemente debido a la época de abundantes lluvias. Los tamaños y distribución de partículas de la fracción mayoritaria entre 0,5-1,0 mm fue de 65,17%, para la menor fracción (<0,5 mm) fue de aproximadamente de un 20%, la distribución y tamaños de estas últimas se observan en la Fig. 3a. Se aprecia una distribución que resulta casi homogénea pero con un tamaño de partículas predominante ligeramente mayor de 517,2 µm, mientras que para el tamaño y distribución de partículas de la fracción 0,5 mm y hasta 2,0 mm se muestra en la Fig. 3b. Como se puede observar el tamaño prevaleciente fue de 993,6 µm.

La Tabla 5 muestra resultados de la caracterización física

de la variedad *A. lechuguilla* Torr. Se aprecia que existen varias diferencias significativas, salvo para el ancho y la masa de las hojas, en todos los restantes parámetros hay diferencias. Con relación a lo reportado [11], en este trabajo se obtienen hojas más cortas, menor espesor, menor volumen, mayor densidad y mayor área foliar. En relación a las mediciones de la altura y el diámetro de la roseta de las plantas no existen diferencias significativas. Los contenidos de humedad de igual manera resultaron más altos de lo esperado.

Se aprecia en la Fig. 4a una distribución casi homogénea pero con un tamaño de partículas predominante ligeramente mayor de 567,7 µm, similar a la variedad *A. Angustifolia*. En la Fig. 4b se observa el tamaño y distribución de partículas de la fracción 0,5-2,0 mm con tamaño prevaleciente de 1909 µm.

En cuanto al color de las hojas, según [9] con valor promedio de 3,50, se consideran en edad madura y excelente para ser utilizadas. Estas hojas presentaron el menor número de espinas laterales (27 como promedio) que todas las demás variedades, esto debido a su menor tamaño relativo, la altura promedio de la espina apical (1,87 cm) si se corresponde con el reportado por [9] (≈ 2,0 cm).

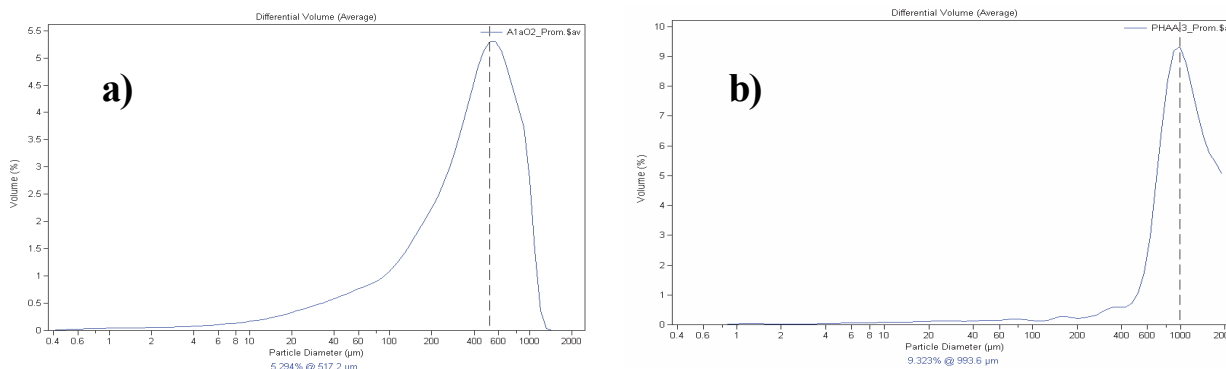


Figura 3. Distribución y tamaños de partículas secas de la a) fracción minoritaria (<0,5 mm) y b) fracción 0,5-2,0 mm de hojas de *Agave angustifolia* Haw., tomada en el municipio de Zimatlán, Oaxaca, México.

Fuente: Elaboración propia

Tabla 5.

Caracterización física de *Agave lechuguilla* Torr., tomada en el municipio de San Agustín Tlaxiaca, Hidalgo.

Hoja	Hoja (n=10 hojas)					Roseta (n=6 plantas)			
	largo (cm)	ancho (cm)	espesor (cm)	masa (g)	Volumen (cm ³)	densidad (g/cm ³)	Área foliar (cm ²)	Altura Planta (mm)	Diam. Planta (mm)
media	39,83	4,00	0,66	91,04	104,83	0,87	159,40	58,67	91,33
desv.									
Est.	2,177	0,229	0,080	12,529	16,619	0,050	12,352	2,156	3,194
%CV	5,46	5,71	12,23	13,76	15,85	5,67	7,75	3,68	3,50
t cal	7,506	0,061	5,698	0,263	7,454	15,856	15,739	1,955	1,320
Dif. Sig	SI	NO	SI	NO	SI	SI	SI	NO	NO
Ref [13]	45,00	4,00	0,80	90,00	144	0,63	97,92	60,00	90,00

Hoja	Humedad (%)	<0,5 mm (%)	0,5-1,0 mm (%)	>1,0mm (%)	Color (1-6)	Nº Espinas	Alt. E. apical (cm)
media	76,08	41,80	46,92	11,29	3,5	27,33	1,87
desv. Est.	1,230	6,527	8,266	7,299	0,0	2,540	0,227
%CV	1,62	15,62	17,62	64,65	0,0	9,29	12,16
t cal	15,638	-	10,744	-	0,0	-	1,858
Dif. Sig	SI	-	SI	-	NO	-	NO
Ref [13]	70,00	NR	75,00	NR	3,5	NR	2

Nota t tabulada: 2,26; cuando t cal < t tabulada no hay diferencias significativas, NR: No Reportado.

Fuente: Elaboración propia

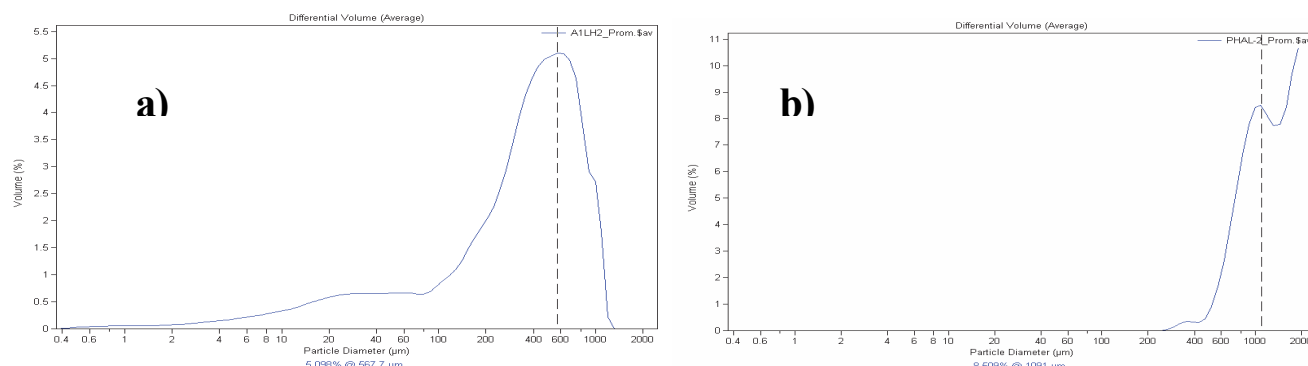


Figura 4. Distribución y tamaños de partículas secas de la a) fracción minoritaria (<0,5 mm) y b) fracción mayoritaria (0,5-2,0 mm) de hojas de Agave lechuguilla Torr., tomada en el Municipio de San Agustín Tlaxiaca, Hidalgo, México.

Fuente: Elaboración propia.

3.2. Microscopía electrónica de barrido (MEB)

En la Fig. 5 a-d se observa la morfología de hojas previa a la trituration, formadas por fibrillas de celulosa aglomeradas mediante hemicelulosa y lignina que actúan como cementantes, que deben ser eliminados durante el proceso de obtención de fibras o pulpa de celulosa. En la Fig. 6 a-d se aprecian las microfotografías de las fibras de hojas, posterior a la trituration en la fracción 0,5-2,0 mm. Se muestran conformaciones similares de estas fibras para las cuatro variedades. Los tamaños se corroboran con los evaluados en la etapa anterior.

3.3. Análisis térmicos

Por los análisis térmicos (TGA y DTGA) de las especies de agaves, se pudo apreciar cinco momentos de cambios o transformaciones (Fig. 7 y Fig. 8). La biomasa lignocelulósica, es una mezcla compleja de polímeros de carbohidratos conocidos como celulosa, hemicelulosa, lignina y pequeñas cantidades de otras sustancias como extractables y cenizas, contenidos en la pared celular de las plantas. Según [14], la descomposición térmica de la madera, por análisis termogravimétrico, revela dos regímenes de descomposición, y que el pico en la zona de descomposición a la menor temperatura está asociado a la presencia de humedad, seguido por la pirolisis de la hemicelulosa, y el pico a alta temperatura a la descomposición de la celulosa.

Raveendran (1996), desarrolló estudios de descomposición de residuos agrícolas, distinguiendo cinco intervalos de temperatura en los cuales plantea que tiene lugar la descomposición de los diferentes componentes de la biomasa [15,16].

En todas las figuras se observa un primer cambio ($1 \Delta p$) que ocurre entre $47 \pm 5^\circ \text{C}$ y $120 \pm 3^\circ \text{C}$ para las cuatro especies. Este cambio se asocia a la pérdida por evaporación de humedad absorbida por las fibras, debido a su carácter hidrofílico [17]. Estas pérdidas de humedad, representaron entre 2,59-3,54% en masa, con pequeñas diferencias significativas. Estos resultan dos veces más bajos que los que reporta [18] en fibras de henequén (6%) y menor que 4,2% a lo reportado por [19] para fibras de Leguminosa unguiculata.

La humedad absorbida está correlacionada con la cantidad de lignina; atribuible a la presencia de los grupos carboxílicos

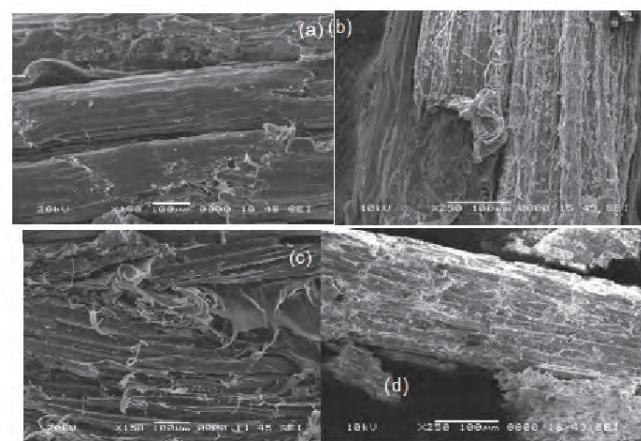


Figura 5. Microfotografías que muestran las morfologías de las fibras de hojas de agaves secadas, previo a la molienda. a) *A. angustifolia*, b) *A. salmiana*, c) *A. tequilana* y d) *A. lechuguilla*

Fuente: Autores

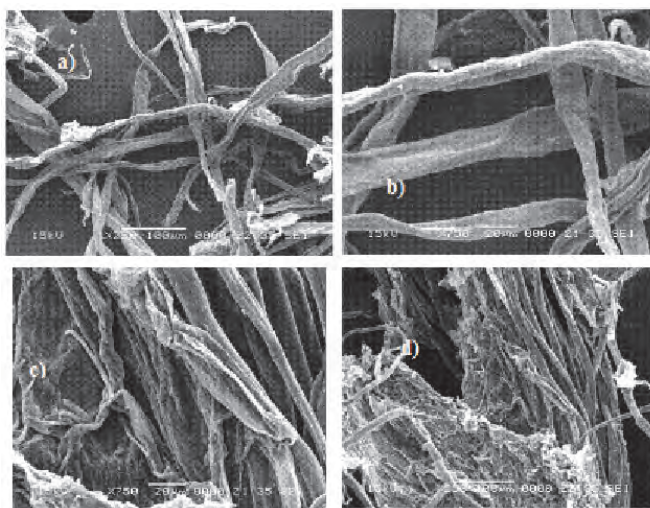


Figura 6. Microfotografías que muestran las morfologías de las fibras obtenidas por proceso de molienda y tamizaje para las cuatro variedades de hojas de agaves en estudio a) *A. angustifolia*, b) *A. salmiana*, c) *A. tequilana* y d) *A. lechuguilla*

Fuente: Autores

y carbonilos que se introducen en la lignina mediante el sistema enzimático de la lacasa, que inhibe los grupos hidroxilos e incrementan el carácter hidrofílico [20]. De ahí que en el proceso de pulpeo alcalino se mineralice la lignina haciéndola solubilizable.

El segundo cambio (2 Δ p), menos pronunciado que el anterior, se observó en el intervalo de temperatura de 140,1 \pm 5,7 °C a 177,1 \pm 2,1 °C con una pérdida de masa del 2,57-3,47%, que se atribuyó a la degradación de la hemicelulosa, la cual siendo semicristalina, presenta menos estabilidad térmica que la celulosa [16,21,22].

Mientras que la celulosa es más cristalina, fuerte y resistente a la hidrólisis, la hemicelulosa tiene una estructura semicristalina al azar con poca fuerza, fácilmente hidrolizada por ácido diluido o base, así como por enzimas hemicelulasas innumerables.

Los xilanos son cuantitativamente dominantes en la fracción hemicelulósica. El más abundante es el arabinoxilano caracterizado por la unión mediante un enlace O-glucosídico β -(1 \rightarrow 3) de moléculas de α -L-arabinofuranosas a la cadena de xilano principal [23]. Por su estructura química, al poseer unidades de pentosas, la distribución dentro de la estructura de la planta es en forma de microtúbulos entrelazados de modo superficial de una cadena a otra, sólo por enlaces por puente de hidrógeno. De lo anterior se puede deducir, que ante un proceso de despolimerización, como lo constituye el aumento de temperatura, se necesitará menor energía térmica para fragmentarlo [22].

La celulosa forma cadenas largas entrelazadas unas con otras por enlaces de hidrógeno, de un modo más compacto que en la hemicelulosa. Esto se debe a los grupos -CH₂OH alternados por arriba y por debajo del plano de la molécula, que permiten una organización espacial en forma de bloques a través de puentes de hidrógeno, provocando un proceso de despolimerización de las capas superiores hacia el interior del paquete, lo que explica la mayor estabilidad térmica con respecto a la hemicelulosa [22,23].

El tercer y principal cambio observado (3 Δ p) es el que representa la degradación de la celulosa presente en fibras vegetales de agaves, con una pérdida de masa entre 1.08-1.50 mg, representando entre 34.23-44.25%. Esta pérdida ocurre en un rango de temperatura de 250 \pm 8 °C y 360 \pm 3 °C. Este intervalo es prácticamente igual al que reportan [19] para la degradación de la celulosa presente en el bagazo de caña o para la cascarilla de café (entre 200-400 °C) con un porcentaje en masa similar de 44% [19]. Para residuos de hojas de tabaco se observó por los mismos autores dos zonas de temperaturas, una entre 227-320 °C y otra entre 320-460 °C. De la Fig. 8 se observa que para la celulosa pura el rango de temperatura fue de 302-368 °C lo cual corrobora lo anterior.

El cuarto cambio (4 Δ p) se observó entre temperaturas de 428 \pm 7 °C y 503 \pm 3 °C y se atribuyó a la degradación de la lignina. Se apreció una pérdida de masa entre 2,11-2,86%. Estos resultados se corresponden con los que reporta [11], quienes encontraron 2.61% en *A. salmiana*. Según Otero, citado por [23], los contenidos de lignina en agaves cultivados en Cuba (*A. fourcroydes*, Lem), pueden oscilar entre 3,1-9,9%; en este caso los valores resultaron más bajos. Por su parte [24] ha reportado en cascarilla y paja del cultivo de cebada contenidos de lignina mucho más altos (10,6-19,7%).

El quinto cambio observado (5 Δ p) entre 524 \pm 20 °C y 596 \pm 4 °C se atribuye a las “cenizas” derivadas de la degradación de polímeros más complejos y sales inorgánicas, con una pérdida de masa entre 7,14 % y 10,10 %. A temperaturas superiores a 600 °C es posible que se logre el resto de las transformaciones (degradación y/o fusión de sales).

En la bibliografía se reportan contenidos de hemicelulosa para plantas no maderables entre 15-32%, de lignina entre 7-24% y de celulosa entre 22-85% [24], mucho más elevados que los que se han encontrado. Para especies maderables los contenidos reportados de lignina resultan más elevados (17-48%) [25].

3.4. Solubilidades y análisis químicos

Las hojas de los agaves constituyen residuos agrícolas del cultivo a pesar de su alto contenido de azúcares totales (AT) y a los grandes volúmenes que anualmente se generan [7,26]. Las hojas representan del 45 al 50% del peso total de la planta. Es en las hojas secadas se encuentran residuos que pueden ser solubles en agua fría y/o caliente u otros medios; atribuidos a la presencia de azúcares totales (AT), parte de la lignina, entre otros. Anualmente la industria tequilera demanda aproximadamente un millón de toneladas de piñas de *A. tequilana* Weber Var. azul, actividad que genera una cantidad similar en peso de hojas y que no son utilizadas en la actualidad [7,27].

Por su estructura química, los carbohidratos se dividen en dos grupos principales, azúcares y no azúcares. Los azúcares más simples (monosacáridos), en general, son solubles en agua, escasamente en etanol e insolubles en éter y poseen propiedades reductoras. Las fructanas son polímeros de fructosa que presentan un solo residuo de glucosa en el extremo de la molécula; constituyen unos de los principales carbohidratos de las plantas [4,28], también se les conoce como polifruetosilsacarosos debido a que son polímeros sintetizados a partir de la molécula de sacarosa. Se caracterizan por pertenecer a la fracción solubilizable en agua caliente [29].

En la Tabla 6 se muestran los resultados de porcentajes de cenizas, grasas y proteínas, así como material solubilizable en agua caliente, expresados como %AT, material solubilizables en NaOH y CH₃CH₂-OH. Como se aprecia, las tres especies de hoja angosta mostraron similares contenidos de cenizas a diferencia de la especie *A. salmiana* de hoja ancha. Asociado a la posible mayor capacidad de acumular sales minerales por sus mayores dimensiones. Este valor encontrado en *A. salmiana* es 1,6 veces más elevado que el reportado por [4], lo que se atribuye a la toma de muestras en épocas de lluvia con mayor probabilidad de incrementar las sales disponibles de los suelos.

Los contenidos de grasas y aceites resultan ser valores relativamente bajos, siendo el menor de ellos el correspondiente a la especie *A. lechuguilla* (0.08 %). El valor más elevado correspondió a la variedad *A. salmiana* (1.85 %) y resulta similar al que reportan [6] en 1.82%. Peña y colaboradores [30], indican que la presencia de grasas en los agaves (*A. tequilana*, *A. salmiana* y *A. angustifolia*) se atribuye a la presencia de terpenos y ácidos grasos y concuerdan los resultados con los encontrados cuando señala [31] que la especie *A. salmiana* es más rica en ácidos grasos.

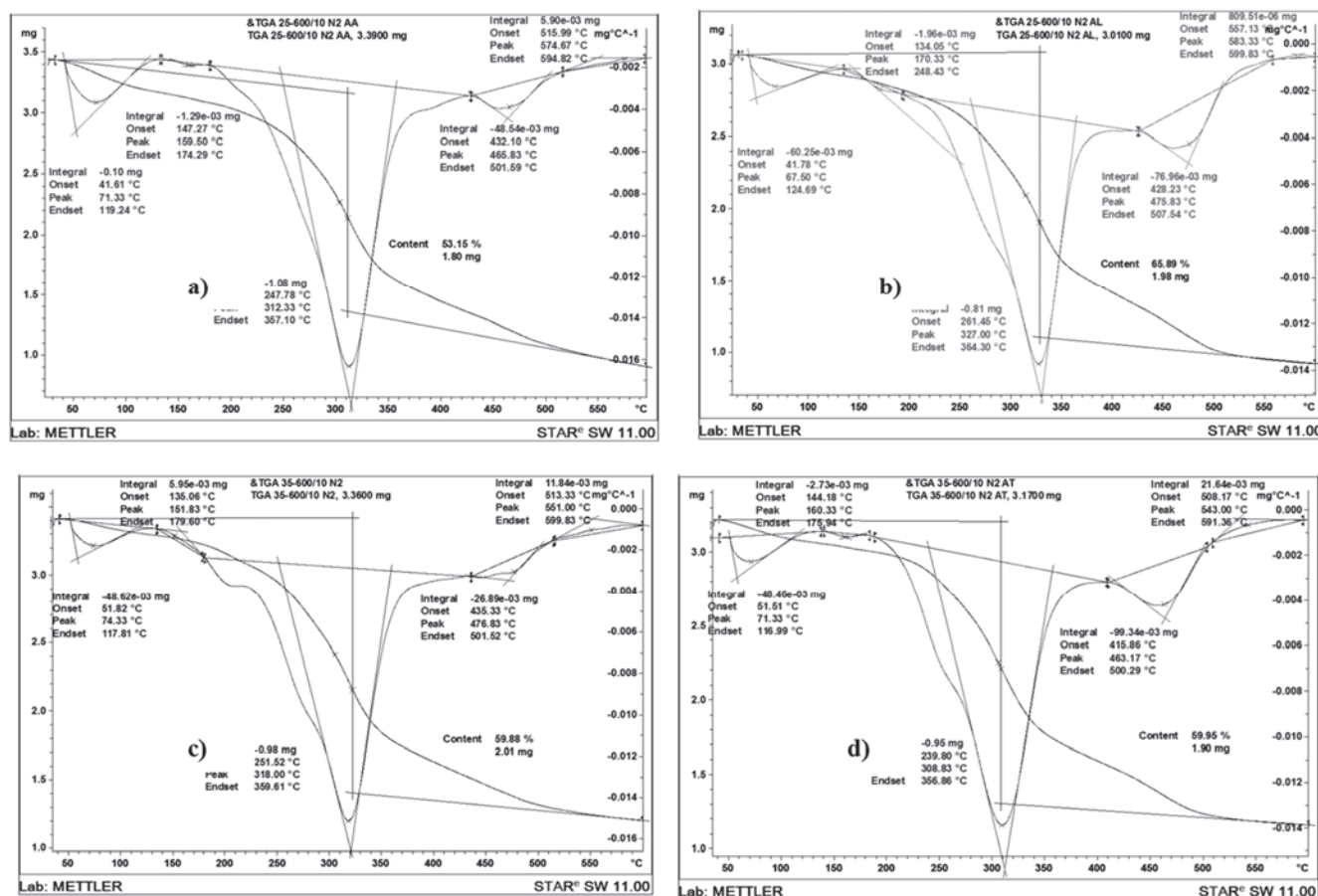


Figura 7. Análisis térmicos de las fibras de a) *A. angustifolia*, b) *A. lechuguilla*, c) *A. salmiana* y d) *A. tequilana*. Fuente: Elaboración propia

En cuanto a los contenidos en proteínas resultan ser menores de un 5% (entre 2-3%) lo cual no hace de interés a estas especies es este sentido. De nueva cuenta la variedad más pobre en proteínas fue *A. lechuguilla* (0,37 %). El valor encontrado de 3,32 % en *A. salmiana* es similar al que reporta [4] de 3,24 %.

Para los contenidos de solubilizables en agua caliente (AT), se realizó la curva de calibrado a partir de estándar de manosa ($y = 0,004x - 0,0091$, y con un regresión de 0,9964).

Los resultados variaron entre 17-21% de AT, siendo las especies más rica la *A. tequilana* y la *A. salmiana*. Algunos autores señalan que en general los AT están constituidos por inulinas y fructanos no identificados, como principales carbohidratos solubles de reserva además de sacarosa, fructosa y glucosa, los cuales pueden participar en los procesos de protección de la planta, almacenamiento de fuentes de energía, estabilizadores de presiones intracelulares y como compuestos de difusión para un almacenamiento más eficiente de los productos originados por el proceso fotosintético, ya que estos compuestos tienden a encontrarse en mayor proporción en el floema [32]. Sin embargo por los bajos contenidos en materia solubilizables en extracto etanólico, se asume que son relativamente bajos los contenidos de inulina. Esto se corresponde con lo reportado para la especie *A. atrovirens* Kart [33], donde los autores

también indican que los resultados son dependientes de la edad del cultivo. Los contenidos solubles en solución 1% de NaOH y etanol-agua 1:1, se asocian a una fracción solubilizable de una parte de los contenidos de lignina presentes y que son integrantes de la fibra cruda total. Una parte de la lignina total puede ser parcialmente disuelta por álcalis y/o solventes orgánicos. Luego de una hidrólisis ácida, la lignina puede ser removida por un proceso de

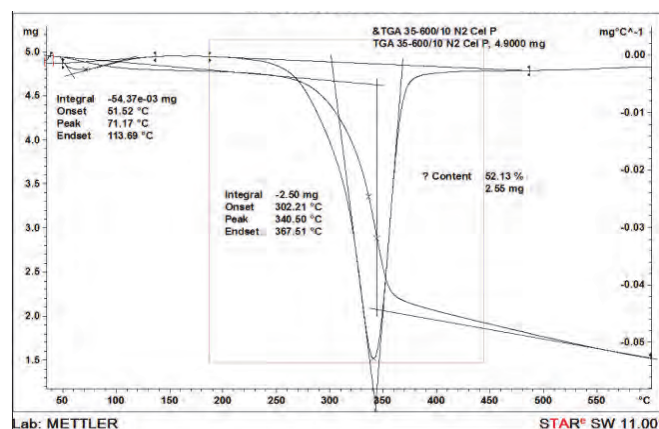


Figura 8. Análisis térmicos de Celulosa pura. Fuente: Elaboración propia

Tabla 6.
Resultados de análisis proximales de las hojas de agaves en estudio

Especie: <i>Agave angustifolia</i> Haw (Mezcal) (Oaxaca)						
Hoja	Cenizas (%)	Grasas (%)	Proteína (%)	Sol. H ₂ O (% AT)	Sol. NaOH 1% (%)	Sol. CH ₃ -CH ₂ -OH:agua 1:1 (%)
media	10,09 ^a	0,697 ^a	2,289 ^a	18,741 ^a	2,055 ^b	0,344 ^b
desv. Est.	0,044	0,085	0,200	0,255	0,095	0,025
%CV	0,439	12,20	8,74	1,36	4,62	7,27
Especie: <i>Agave tequilana</i> Weber, Var. Azul (Tequila) (Jalisco)						
Hoja	Cenizas (%)	Grasas (%)	Proteína (%)	Sol. H ₂ O (% AT)	Sol. NaOH 1% (%)	Sol. CH ₃ -CH ₂ -OH:agua 1:1 (%)
media	10,08 ^a	0,97 ^b	2,64 ^b	20,630 ^b	3,021 ^c	0,322 ^b
desv. Est.	0,210	0,103	0,205	0,254	0,104	0,019
%CV	2,087	10,57	7,77	1,23	3,443	5,90
Especie: <i>Agave salmiana</i> B. Otto ex Salm-Dick (Pulque) (Hidalgo)						
Hoja	Cenizas (%)	Grasas (%)	Proteína (%)	Sol. H ₂ O (% AT)	Sol. NaOH 1% (%)	Sol. CH ₃ -CH ₂ -OH:agua 1:1 (%)
media	12,93 ^b	1,85 ^c	3,32 ^c	20,333 ^b	1,875 ^a	0,284 ^a
desv. Est.	0,321	0,115	0,186	0,232	0,112	0,029
%CV	2,49	6,22	5,60	1,14	5,97	10,21
Especie: <i>Agave lechuguilla</i> Torr. (fibras) (Hidalgo)						
Hoja	Cenizas (%)	Grasas (%)	Proteína (%)	Sol. H ₂ O (% AT)	Sol. NaOH 1% (%)	Sol. CH ₃ -CH ₂ -OH:agua 1:1 (%)
media	10,08 ^a	0,08 ^d	0,369 ^d	16,889 ^c	2,176 ^a	0,472 ^c
desv. Est.	0,181	0,013	0,067	0,328	0,176	0,069
%CV	1,79	16,25	18,16	1,94	8,09	14,62

Nota: Medias con letras diferentes en columna indican diferencias significativas ($p < 0,05$) según Tuckey.

Fuente: Elaboración propia

deslignificación alcalina con NaOH; sin previa hidrólisis ácida, solo es removida solo una pequeña fracción de la lignina total. Este tipo de tratamiento simple con álcalis es reportado en estudios para mejorar la digestibilidad de tejidos vegetales (por ejemplo con $\text{Ca}(\text{OH})_2$) donde es eliminada parcialmente la lignina [34].

3.5. Análisis por FTIR

En la Fig. 9 se muestran los espectros FTIR de las cuatro especies en estudio y de la celulosa pura. Se observa un gran número de bandas, donde se superponen absorciones características de la celulosa, hemicelulosas y la lignina, con preponderancia de las bandas de la celulosa, que es el componente mayoritario de las fibras de los agaves.

Los espectros FTIR de polisacáridos son generalmente difusos, con fuerte solapamiento de las bandas, de modo que las diferencias espectrales entre ellos no se encuentran bien definidas. Esto se atribuye a la complejidad de los diversos tipos de interacciones intermoleculares e intramoleculares [34]. Se observan de manera importante bandas entre 3500-3390 cm^{-1} asociadas a la tensión del enlace O-H con formación de puentes de hidrógeno y de doblaje en las señales entre 1000 y 1200 cm^{-1} y que están presentes en la lignina, la hemicelulosa y la celulosa; estas últimas se traslapan con señales de vibraciones de valencia de enlaces glicosídicos C-O-C en la zona de 1050 a 1200 cm^{-1} .

Las bandas de vibración entre 2915-2920 cm^{-1} (simétrica) y 2848-2850 cm^{-1} (asimétrica), así como las bandas en 1420-1425 cm^{-1} y 1315-1320 cm^{-1} correspondientes a la tensión del enlace H-C-H del grupo $-\text{CH}_2$ o $-\text{CH}_3$ se corresponden con lo reportado [38] para la celulosa; las bandas en 1735 cm^{-1} (hombro) y 1636 cm^{-1} , que se atribuye a la tensión del enlace C-O en C=O de éster lactónico y C=C vinílico o aromático.

Se aprecia también absorciones características de anillos aromáticos de la lignina en la zona de 1400 a 1600 cm^{-1} , mismas que se ven disminuidas significativamente en la muestra de celulosa pura. Es interesante distinguir la señal de absorción entre 875-895 cm^{-1} como absorción más típica de la configuración estereoquímica β de las D-glucopiranosas, reportada para la celulosa como una banda débil [34]. La naturaleza de esta absorción del espectro IR de la celulosa y otros polisacáridos tiene un carácter complejo según señala Zbankovb en 1966, citado por [34]. En la celulosa natural esta banda es débil, pero su intensidad varía con el tratamiento a que se someta la misma; lo cual es un aspecto de seguimiento en el proceso de pulpeo, donde esta intensidad debe ir en aumento [34]. En efecto para la celulosa pura se puede apreciar dicha banda débil pero de carácter más intenso en relación a las especies en estudio.

4. Conclusiones

Los promedios de todas las especies en cuanto al tamaño de las partículas en las fracciones minoritarias ($<0,5$ mm), aquellas que pudieran resultar no beneficiosas al momento del pulpeo, se concluye que con tamaños entre 0,34-0,45 mm corresponden los porcentajes más elevados (25,2% y 41,8%) para las especies *A. lechuguilla* y *A. salmiana*; ambas especies cultivadas en el estado de Hidalgo. Se corresponde con las especies de hoja más larga, más ancha y de mayor espesor (*A. salmiana*) y la más corta, más angosta y de menor espesor (*A. lechuguilla*). Corresponde igualmente a las dos especies con los porcentajes más bajos encontrados en celulosa (35,9% y 34,2%, respectivamente), según DTGA. Las especies *A. angustifolia* y *A. tequilana* fueron las de menores porcentaje en esta fracción $<0,5$ mm, y mayor porcentaje en partículas 0,5-2,0 mm lo cual representa una

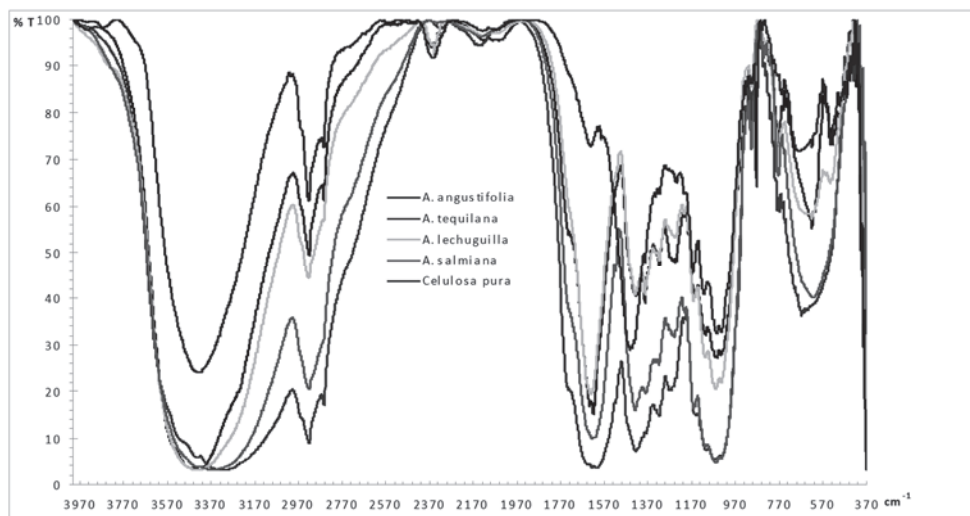


Figura 9. Espectros FTIR en pastillas de KBr de las hojas de agave secadas y trituradas.
Fuente: Elaboración propia

ventaja sobre las otras especies, concluyendo que éstas son las que presentan un óptimo tamaño de fibra para la elaboración de pulpa. Corresponde también a los mayores porcentajes de celulosa (44,3% y 36,8%) por DTGA.

Del análisis proximal, se evidencia que las dos especies definidas por análisis físicos como mejores para la obtención de pulpa de celulosa, presentan pequeñas diferencias significativas en contenidos de grasas, proteínas, material soluble en agua y en solución moderadamente alcalina y/o alcohólica. En las especies *A. salmiana* y *A. lechuguilla* las diferencias significativas resultan más marcadas.

Los espectros FTIR se presentan con fuerte solapamiento de las bandas, de modo que las diferencias espectrales entre diferentes especies no se encuentran bien definidas. Sin embargo se logran identificar señales propias que están presentes en la lignina, la hemicelulosa y la celulosa. Es importante distinguir la señal de absorción en 895 cm⁻¹ como la absorción más típica de la configuración estereoquímica β de las D-glucopiranosas, reportada para la celulosa.

Referencias

- [1] Franco-Guarnizo, A., Martínez-Yépes, P.N. y Sánchez-Hoover, A., Pretratamientos de la celulosa y biomasa para la sacarificación. *Revista Scientia Et Technica*, [En línea]. 15(42), pp. 284-289, 2009. Disponible en: <http://www.redalyc.org/pdf/849/84916714053.pdf>.
- [2] Parra, L.A., del Villar, P. y Prieto, A., Extracción de fibras de agave para elaborar papel y artesanías. *Acta Universitaria*, [En línea]. 20(3), pp. 77-83, 2010. Disponible en: <http://www.redalyc.org/articulo.oa?id=41618860011>.
- [3] Ruiz-Corral, J.A., Requerimientos agroecológicos y potencial productivo del agave *Agave tequilana* Weber en México. Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias. [En línea]. Centro de Investigación Regional del Pacífico Centro. 2007, pp. 11-36. Disponible en: http://www.inifapcirne.gob.mx/Revistas/Archivos/agave_final_baja%20resolucion.pdf.
- [4] Vargas, C.G. Obtención de insumos de interés industrial a partir de las fructanas del agave mezcalero potosino (*A. salmiana*). Tesis, Instituto Politécnico Nacional. Centro Interdisciplinario de Investigaciones para el Desarrollo Integral Regional. CIIDIR-IPN-Michoacán, México, 2009.
- [5] Avila-Fernández, A., Galicia-Lagunas, N., Rodríguez-Alegria, M.E., Olvera, C. y López-Munguía, A., Production of functional oligosaccharides through limited acid hydrolysis of agave fructans. *Food Chem*, 129(2), pp. 380-386, 2011. DOI: 10.1016/J.FOODCHEM.2011.04.088.
- [6] Rojas-Urbe, E.H. y Narváez, P.C. Método de análisis de calidad del aceite durante el freído por inmersión para pequeñas y medianas empresas. *Ingeniería e Investigación* [En línea]. 31(1), pp. 83-92, 2011. Disponible en: <http://www.redalyc.org/articulo.oa?id=64321170010>.
- [7] Rosas-Mendoza, M.E. y Fernández-Muñoz, J.L., FTIR aplicada durante la deshidratación osmótica de mango Ataulfo (*Mangifera indica* L.). *Superficies y Vacío*, [En línea]. 25(1), pp. 8-13, 2012. Disponible en: <http://www.scielo.org.mx/pdf/sv/v25n1/v25n1a2.pdf>.
- [8] Montañez, J.L., Victoria, J.C., Flores, R. and Vivar, M.A., Fermentation of *Agave tequilana* Weber Azul fructans by *Zymomonas mobilis* and *Saccharomyces cerevisiae* in the production of bioethanol. *Información Tecnológica*, 22(6), pp. 3-14, 2011. DOI: 10.4067/S0718-07642012000300009.
- [9] Aguilar, R.N., Efecto del almacenamiento de bagazo de caña en las propiedades físicas de celulosa grado papel. *Ingeniería Investigación y Tecnología*, [En línea]. 12(1), pp. 189-197, 2011. Disponible en: <http://www.redalyc.org/html/404/40419907008/>.
- [10] Mora-López, J.L., Reyes-Agüero, J.A., Flores-Flores, J.L., Peña-Valdivia, C.B. y Aguirre-Rivera J.R., Variación morfológica y humanización de la sección *salmiana* del género *Agave*. *Agrociencia*, [En línea]. 45(4), 465-467, 2011. Disponible en: <http://www.colpos.mx/agrociencia/Bimestral/2011/may-jun/art-6.pdf>.
- [11] Silos-Espino, H., Tovar-Robles, C.-L., González-Cortés, N., Méndez-Gallegos, S.J. y Rossel-Kipping, D., Estudio integral del maguey (*Agave salmiana*): Propagación y valor nutricional. XI Simposium-Taller Nacional y II Internacional de producción del nopal y maguey, [En línea]. pp. 75-82, 2011. Disponible en: <http://cvu.colpos.mx/cvu.aspx?idinv=120378>.
- [12] Rodríguez-Macias, R., Alcantar-González, E.G., Iñiguez-Covarrubias, G., Zamora-Natera, F., García-López, P.M., Ruiz-López, M.A. y Salcedo-Pérez, E., Caracterización física y química de sustratos agrícolas a partir de bagazo de agave tequilero. *Interciencia*. [En línea]. 35(7), pp.515-520, 2010. Disponible en: http://www.interciencia.org/v35_07/515.pdf.
- [13] Carrier, M., Loppinet-Serani, A., Denux, D., Lasnier, J.M., Ham-Pichavant, F., Cansell, F. and Monier, C.A., Thermogravimetric analysis as a new method to determine the lignocellulosic composition of biomass. *Biomass and Bioenergy*, 35 (1), pp. 298-307, 2011. DOI: 10.1016/j.biombioe.2010.08.067.

- [14] Manals, E., Pinedo, M. y Giralt, G., Análisis termogravimétrico y térmico diferencial de diferentes biomásas vegetales. *Tecnología Química*, XXXI(2), pp. 36-43, 2011.
- [15] Bolio, G.I., Valadez, A., Valeva, L. and Andreeva, A., Cellulose whiskers from agro-industrial banana wastes: Isolation and characterization. *Revista Mexicana de Ingeniería Química*, [Online]. 10(2), pp. 291-299, 2011. Available at: <http://www.scielo.org.mx/pdf/rmiq/v10n2/v10n2a13.pdf>
- [16] Muñoz, M.F., Hidalgo, M.A. and Mina, J.H., Fique fiber an alternative for reinforced plastics. Influence of surface modification. *Biotechnología en el Sector Agropecuario y Agroindustrial*, [Online]. 12(2), pp. 60-70, 2014. Available at: <http://www.scielo.org.co/pdf/bsaa/v12n2/v12n2a07.pdf>
- [17] Quesada-González, O., Torres-García, E. y Alfonso-Martínez, F.E., Estudio de la degradación térmica del residuo de la naranja. *Revista Cubana de Química*, 23(1), pp. 25-33, 2011. DOI: 10.1016/j.riit.2015.05.001
- [18] Tomás, M.E., Bioetanol de paja de trigo: Estrategias de integración de las etapas del proceso. Tesis, [En línea]. Departamento de Microbiología II, Facultad de Ciencias Biológicas, Universidad Complutense de Madrid. Madrid, España, 2010. Disponible en: <http://eprints.ucm.es/10802/1/T31774.pdf>
- [19] Ayrlimis, N., Jarusombuti, S., Fueangvivat, V. and Bauchongkol, P., Effect of thermal-treatment of wood fibres on properties of flat-pressed wood plastic composites. *Polymer Degradation and Stability*, 96(5), pp. 818-822, 2011. DOI: 10.1016/j.polyimdegstab.2011.02.005
- [20] Rojas-León A., Propuesta sustentable para la utilización de residuos agrícolas y urbanos: Fabricación de un eco-material compuesto. Tesis Instituto de Ciencias Básicas e Ingeniería. Universidad Autónoma de Hidalgo, Pachuca, México. 2014.
- [21] Vinnet-Serrano, E. y Fajardo-Gutiérrez, O., Estrategia para el mejoramiento genético de Agaves en Cuba. Instituto de Investigaciones Hortícolas "Liliana Dimitrova" (IIHLD), [En línea]. 2009, pp. 37-43. Disponible en: http://www.utm.mx/edi_anteriores/temas037/N3.pdf
- [22] Betancourt, S., Gañán, P., Jiménez, A. y Cruz, L.J. Degradación térmica de fibras naturales procedentes de la calceta de plátano (estudio cinético). *Revista Latinoamericana de Metalurgia y Materiales*, [En línea]. 1(1), pp. 215-219, 2009. Disponible en: <http://www.rlmm.org/archivos/S01N1/RLMMArt-09S01N1-p215.pdf>
- [23] Moya-Roque, R., Información comparativa de las 10 especies maderables evaluadas provenientes de plantación en Costa Rica. Ficha técnica 11. En: R. Moya R., et al., eds., *Tecnología de madera de plantaciones forestales: Fichas técnicas*. [En línea]. Revista Forestal Mesoamericana Kurú 7(18-19), pp. 181-219. 2010. Editorial Corporación Garro y Moya, (ISBN: 978-9968-9643-3-3). Disponible en: www.tec.ac.cr/revistaforestal.
- [24] Iñiguez, G., Acosta, N., Martínez, L. Parra, J. y González, O., Utilización de subproductos de la industria tequilera. Parte 7: Compostaje de bagazo de agave y vinazas tequileras. *Revista Internacional de Contaminación Ambiental*, [En línea]. 21(1), pp. 37-50, 2005. Disponible en: <http://www.redalyc.org/pdf/370/37021105.pdf>
- [25] Narváez, Z.J. and Sánchez, T.F., Agaves as a raw material, recent technologies and applications. *Recent Patents on Biotechnology*, 3(3), pp. 1-7, 2009. DOI: 10.2174/187220809789389144
- [26] Chacón, A., Perspectivas agroindustriales actuales de los Oligofructosacáridos (FOS). *Agronomía Mesoamericana*, [En línea]. 17(2), pp. 265-286, 2006. Disponible en: <http://www.redalyc.org/pdf/437/43717212.pdf>
- [27] Peña-Valdivia, C.B., Sánchez-Urdaneta, A.B., Aguirre, J.R.R., Trejo, C., Cárdenas, E. and Villegas, A., Temperature and mechanical scarification on seed germination of "maguey" (*Agave salmiana* Otto ex Salm-Dyck). *Seed Sci. Technol*, 34(1), pp. 47-56, 2006. DOI: 10.15258/sst.2006.34.1.06
- [28] Molina-Guerrero, J.A., Botello-Álvarez, J.E., Estrada-Baltazar, A., Navarrete-Bolaños, J.L., Jiménez-Islas, H., Cárdenas-Manríquez, M. y Rico-Martínez, R., Compuestos volátiles en el mezcal. *Revista Mexicana de Ingeniería Química*, [En línea]. 6(1), pp. 41-50, 2007. Disponible en: <http://www.redalyc.org/pdf/620/62060106.pdf>
- [29] Saengthongpinit, W. and Sajjaanantakul, T., Influence of harvest time and storage temperature on characteristics of inulin from Jerusalem artichoke (*Helianthus tuberosus* L.) tubers. *Postharvest biology and Technology*, 37(1), pp. 93-100, 2005. DOI: 10.1016/j.postharvbio.2005.03.004
- [30] Ramírez-Cortina, C.R., Alonso-Gutiérrez, M.S. y Rigal, L., Valorización de residuos agroindustriales del tequila para alimentación de Rumiantes. *Revista Chapingo, Serie Ciencias Forestales y del Ambiente*, [En línea]. 18(3), pp. 449-457, 2012. Disponible en: <http://www.redalyc.org/pdf/629/62926234015.pdf>
- [31] Bermello, A., Díaz, D., Martínez, R., Quintana, G., Mieres, G. y Leal, J.A., Estudio de un material compuesto de partículas de bagazo de caña y matriz de polietileno mediante espectrometría FTIR. *ICIDCA. Sobre los Derivados de la Caña de Azúcar*, [En línea]. 42(3), pp. 106-111, 2008. Disponible en: <http://www.redalyc.org/pdf/2231/223120667015.pdf>
- [32] Contreras, J.Q.H., Trujillo, P.H.A., Arias, O.G., Pérez, C.J.L. y Delgado, F.E., Espectroscopia ATR-FTIR de celulosa: Aspecto instrumental y tratamiento matemático de espectros. e-Gnosis, [En línea]. 8(9), pp. 1-13, 2010. Disponible en: <http://www.redalyc.org/pdf/730/73013006008.pdf>
- [33] Gutiérrez-Urbe, A.J., Santos-Zea, L., Leal-Díaz, A.M. and Cortes-Ceballos, E., Agave (*Agave spp.*) and its Traditional Products as a Source of Bioactive Compounds. *Current Bioactive Compounds*, 8(3), pp. 218-231, 2012. DOI: 10.2174/157340712802762410
- [34] Vega-Baudrit J., Sibaja, M.B. Nikolaeva, S.N. y Rivera, A.A., Síntesis y caracterización de celulosa amorfa a partir de triacetato de celulosa. *Rev. Soc. Quím. Perú*. [En línea]. 80(1), pp. 45-50, 2014. Disponible en: www.redalyc.org/pdf/896/89640816003.pdf.

E. Jiménez-Muñoz, Estudiante del Doctorado en Ciencias Ambientales por la Universidad Autónoma del Estado de Hidalgo, México.
ORCID: 0000-0001-7883-0600.

F. Prieto-García, Dr. en Química por la Universidad de Valladolid, España.
ORCID: 0000-0001-8878-3908.

J. Prieto-Méndez, Dra. en Ciencias Ambientales por la Universidad Autónoma del Estado de Hidalgo, México.
ORCID: 0000-0001-5705-1704.

OA. Acevedo-Sandoval, Dr. en Química por la Universidad Autónoma del Estado de Hidalgo, México
ORCID: 0000-0003-0475-7003.

R. Rodríguez-Laguna, Dr. en Ciencias. Manejo de Recursos Naturales por la Universidad Autónoma de Nuevo León, México.
ORCID: 0000-0002-0867-008X.

Entregando lo mejor de los **colombianos**



Línea de atención al Cliente Nacional: **01 8000 111 210**

Línea de atención al Cliente Bogotá: **(57-1) 472 2000**

► www.4-72.com.co

DYNA

83 (197), June 2016

is an edition consisting of 250 printed issues
which was finished printing in the month of June of 2016
in Todograficas Ltda. Medellín - Colombia

The cover was printed on Propalcote C1S 250 g,
the interior pages on Hanno Mate 90 g.
The fonts used are Times New Roman, Imprint MT Shadow

- Risk management of occupational exposure to nanoparticles during a development project: A case study
- Prevalence and symptomatology of musculoskeletal problems reported by home care service workers caring for the elderly
- Current state of the art and enduring issues in anthropometric data collection
- Understanding finger postures when touching targets on the touchscreen of mobile devices
- Tackling the challenges of an aging workforce with the use of wearable technologies and the quantified-self
- Analysis and characterization of starchy and cellulosic materials after enzymatic modification
- Robust sample size for weibull demonstration test plan
- A new hardening rule to model the effect of density on soil behavior
- Monitoring and evaluation of energy use in oil treatment facilities
- Differences between the use of ferric sulphate and ferric chloride on biodesulfurization of a large coal particle
- Human error and response to alarms in process safety
- Changes in EEG amplitude (Alpha and Beta waves) with Thermal environment
- Effects of joints and their waterstops on pressures spread over a slab subject to turbulent flow on a horizontal apron
- Assessing the loss-of-insulation life of power transformers by estimating their historical loads and ambient temperature profiles using ANNs and Monte Carlo simulations
- "Thermal assessment of ecological tiles in physical models of poultry houses"
- Statistical performance of control charts with variable parameters for autocorrelated processes
- Symbolic modeling of the Pareto-Optimal sets of two unity gain cells
- Thermodynamic properties of moisture sorption in cassava flour
- Cr/CrN bilayer coating behavior under erosion-corrosion degradation by jet impingement testing
- Validation of an inertial sensor-based platform to acquire kinematic information for human joint angle estimation
- Study of the evolution of the residual stress state in thermal barrier coatings sprayed on AISI 304 stainless steel
- Burst erasure correction using LDPC codes constructed on base matrices generated by matched groups, nested polygons and superposed circulant matrices
- Heat flow assessment in an underground mine: An approach to improve the environmental conditions
- Secure point-to-point communication using chaos
- Las TIC navegando hacia la sostenibilidad económica y ambiental en el transporte marítimo
- Prioritization of TQM practices in Brazilian medical device SMEs using Analytical Hierarchy Process (AHP)
- Strategic guidelines for supply chain coordination in healthcare and a mathematical model as a proposed mechanism for the measurement of coordination effects
- Variables relevantes para la medición de la calidad percibida del servicio bancario
- Coordinated recharge of electric vehicles in real time
- Physicochemical characterization of four species of agaves with potential in obtaining pulp for paper making
- Gestión de riesgos de exposición ocupacional a las nanopartículas en un proyecto en desarrollo: Estudio de caso
- Prevalencia y sintomatología músculo-esquelética auto-referida por los trabajadores del servicio de apoyo domiciliario en la prestación de atención a personas mayores
- Estado actual de la técnica y cuestiones perdurables en la recogida de datos antropométricos
- Comprensión de las posturas de dedos al tocar objetivos en la pantalla táctil de dispositivos móviles
- Hacer frente a los desafíos de una fuerza laboral que envejece con el uso de tecnologías usables y la auto-cuantificación
- Análisis y caracterización de materiales amiláceos y celulósicos después de modificación enzimática
- Tamaño de muestra robusta para planes de demostración weibull
- Una nueva regla endurecimiento para modelar el efecto de la densidad en el comportamiento de los suelos
- Monitoreo y evaluación del uso de la energía en estaciones de tratamiento de crudo
- Diferencias entre el uso de sulfato férrico y cloruro férrico en la biodesulfurización de un carbón con tamaño de partícula grueso
- Error humano y reacción a alarmas en seguridad de procesos
- Cambios en la amplitud del EEG (ondas Alfa y Beta) con el medio ambiente térmico
- Efectos de las juntas y sus sellos en las presiones distribuidas en una losa sujeta a flujo turbulento en un canal horizontal
- Evaluación de la pérdida de vida del aislamiento sólido en transformadores de potencia, estimando la historia de carga y los perfiles de temperatura ambiente por medio de redes neuronales artificiales y simulaciones de Monte Carlo
- Evaluación térmica de tejas ecológicas en modelos físicos de galpones avícolas
- Desempeño estadístico de cartas de control con parámetros variables para procesos autocorrelacionados
- Modelado simbólico del conjunto Óptimo de Pareto de dos celdas de ganancia unitaria
- Propiedades termodinámicas de humedad de sorción en harina de yuca
- Comportamiento de un recubrimiento bicapa de Cr/CrN bajo degradación por erosión-corrosión por ensayo de impacto con chorro
- Validación de una plataforma basada en sensores inerciales para adquirir información cinemática de estimación del ángulo articular humano
- Estudio de la evolución del perfil de esfuerzos residuales en recubrimientos barrera térmica depositados sobre acero inoxidable AISI 304
- Corrección de borrado en ráfaga utilizando códigos LDPC construidos sobre matrices generadas por grupos combinados, polígonos anidados y matrices circulantes superpuestas
- Evaluación de los flujos de calor en una mina subterránea y enfoque para mejorar sus condiciones ambientales
- Comunicación punto a punto segura usando caos
- ICT sailing to economic and environmental sustainability in the maritime transport
- Priorización de las prácticas de GCT en empresas de dispositivos médicos brasileños utilizando el Analytical Hierarchy Process (AHP)
- Lineamientos estratégicos para coordinación en la cadena de suministro de medicamentos y propuesta de un modelo matemático para medir los efectos de la coordinación
- Relevant variables in the measuring of the perceived quality in bank service
- Coordinación de la recarga de vehículos eléctricos en tiempo real
- Caracterización fisicoquímica de cuatro especies de agaves con potencialidad en la obtención de pulpa de celulosa para elaboración de papel