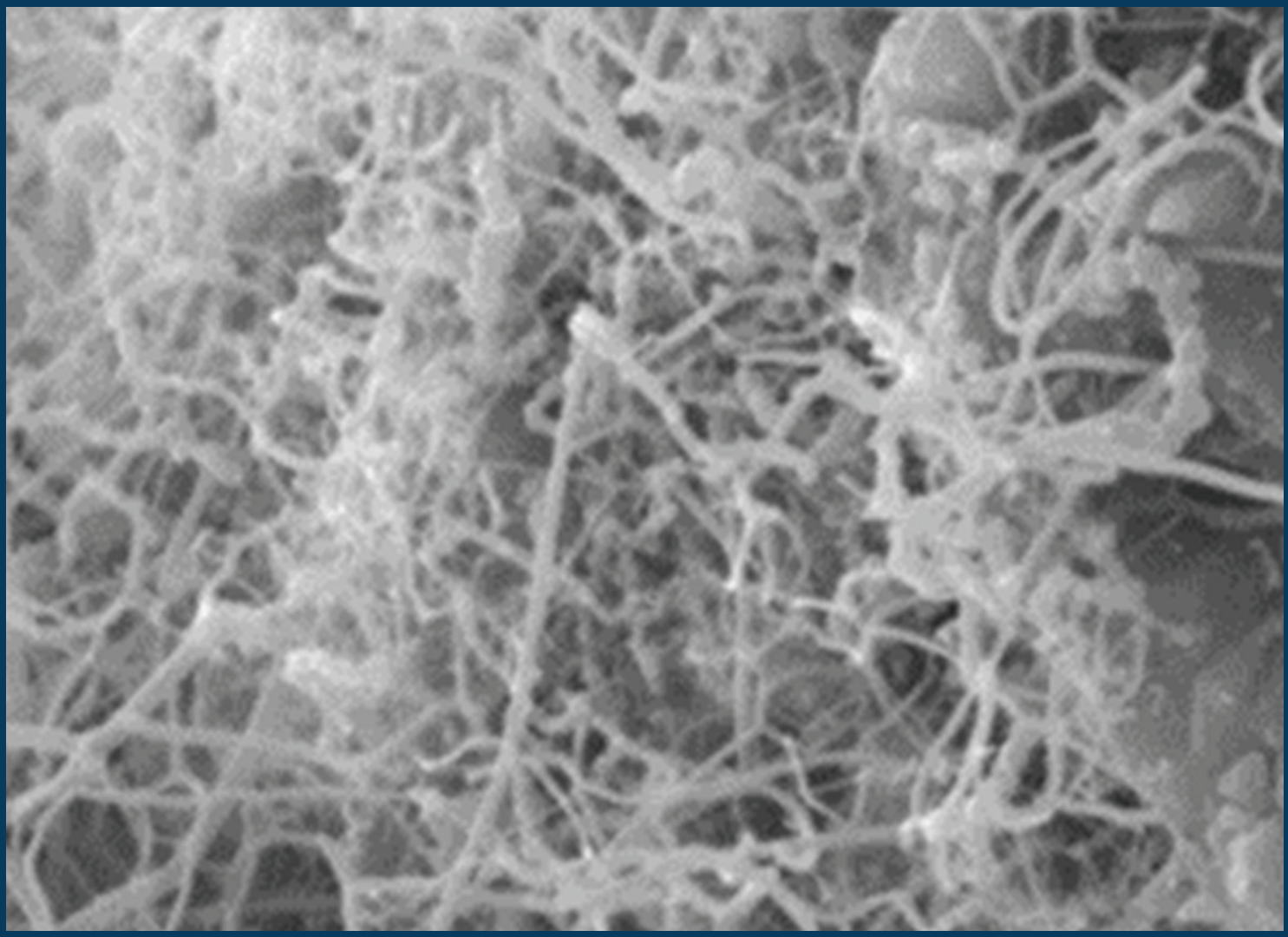
DYNA

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



DYNA 83 (195), February, 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

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS

DYNA
http://dyna.medellin.unal.edu.co/

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

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 Iberoaméricana
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

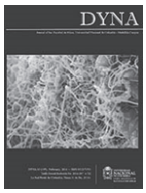
SEDE MEDELLÍN

FACULTAD DE MINAS

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS

DYNA
http://dyna.medellin.unal.edu.co/

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS

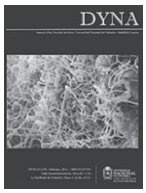


DYNA
http://dyna.medellin.unal.edu.co/

DYNA 83 (195), February, 2016. Medellín. ISSN 0012-7353 Printed, ISSN 2346-2183 Online

CONTENTS

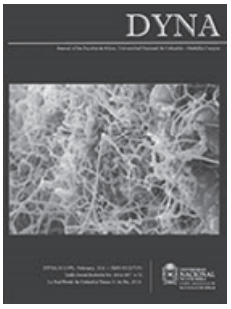
UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS



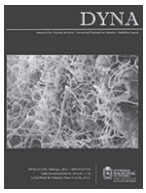
DYNA
http://dyna.medellin.unal.edu.co/

Our cover
Image alluding to Article:
In vitro behavior of the dentin and enamel calcium hydroxyapatite in human premolars subjected to high temperatures

Authors:
Sebastián Medina, Liliana Salazar, Carlos Mejía & Freddy Moreno



UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS

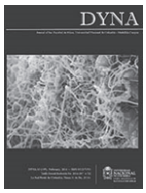


DYNA
http://dyna.medellin.unal.edu.co/

DYNA 83 (195), February, 2016. Medellín. ISSN 0012-7353 Printed, ISSN 2346-2183 Online

CONTENIDO

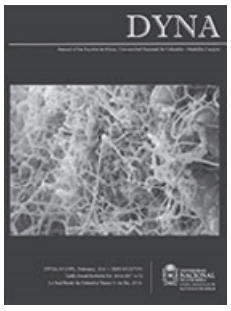
UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS



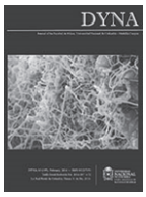
DYNA
http://dyna.medellin.unal.edu.co/

Nuestra carátula
Imagen alusiva al artículo:
Comportamiento in vitro de la hidroxiapatita de calcio de la dentina y del esmalte en premolares humanos sometidos a altas temperaturas

Autores:
Sebastián Medina, Liliana Salazar, Carlos Mejía & Freddy Moreno



UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS



DYNA
http://dyna.medellin.unal.edu.co/

A multiobjective approach for non-discretionary variables in data envelopment analysis

Lidia Angulo-Meza [a], João Carlos Correia Baptista Soares-de Mello [b]& Silvio Figueiredo-Gomes-Júnior [c]

[a] *Departamento de Engenharia de Produção, Universidade Federal Fluminense, Fluminense, Brasil. lidia_a_meza@pq.cnpq.br*
[b] *Departamento de Engenharia de Produção, Universidade Federal Fluminense, Fluminense, Brasil. jcsmello@producao.uff.br*
[c] *Departamento de Engenharia de Produção, Universidade Estadual da Zona Oeste, Rio de Janeiro, Brasil. silviogomes@uezo.rj.gov.br*

Received: September 3rd, 2014. Received in revised form: July 29th, 2015. Accepted: August 18th, 2015.

Abstract
Data Envelopment Analysis is a non-parametrical approach for efficiency evaluation of so-called DMUs (Decision Making Units) and takes into account multiple inputs and outputs. For each inefficient DMU, a target is provided which is constituted by the inputs or outputs levels that are to be attained for the inefficient DMU to become efficient. However, multiobjective models, known as MORO (Multiobjective Model for Ratio Optimization) provide a set of targets for inefficient DMU, which provides alternatives among which the decision-maker can choose. In this paper, we proposed an extension of the MORO models to take into account non-discretionary variables, i.e., variables that cannot be controlled. We present a numerical example to illustrate the proposed multiobjective model. We also discuss the characteristics of this model, as well as the advantages of offering a set of targets for the inefficient DMUs when there are non-discretionary variables in the data set.

Keywords: Data Envelopment Analysis; Multi-objective model; Non-discretionary variables; non-radial models.

Un enfoque multiobjetivo para variables no discrecionales en análisis envolvente de datos

Resumen
El Análisis Envolvente de Datos es un enfoque no paramétrico para la evaluación de la eficiencia de las llamadas DMUs – Unidades Tomadores de Decisión, teniendo en cuenta múltiples entradas y salidas. Para cada DMU ineficiente, se proporciona una meta que está constituida por los niveles de entradas o salidas que deben ser alcanzados para que la DMU ineficiente se vuelva eficiente. Sin embargo, los modelos multiobjetivo, llamados modelos MORO (del ingles *Multiobjective model for Ratio Optimization*) proporcionan un conjunto de metas para DMU ineficientes, lo que ofrece alternativas de las cuales el decisor puede elegir. En este trabajo se ha propuesto una extensión de los modelos MORO para tener en cuenta las variables no-discrecionales, es decir, variables que no se pueden controlar. Se presenta un ejemplo numérico para ilustrar el modelo multiobjetivo propuesto. También se discuten las características de este modelo, así como las ventajas de ofrecer un conjunto de metas para la DMU ineficientes cuando hay variables no discrecionales en el conjunto de datos.

Palabras clave: Análisis Envolvente de Datos; Modelos Multiobjetivo; Variables no-discrecionales; modelos no-radiales.

1. Introduction

Data Envelopment Analysis – DEA [1] is a non-parametric approach for efficiency evaluation of units called DMUs (Decision Making Unit) that use the same inputs to produce the same outputs. An efficiency index is provided for every DMU and for inefficient DMUs, a target is also provided among other information. This target gives the input or output levels for an inefficient DMU to become efficient. In contrast to this non parametric technique, there are a number of parametric techniques used for instance by [2] to measure efficiency.

However, there may be problems in reaching the one target provided for the DEA model as it may not be viable from a managerial or operational point of view. Therefore, models have been presented that seek other targets taking into

account decision-makers' preferences [3,4]. The MORO models [5,6] were presented as a way of proving a set of targets instead of a single target for each inefficient, giving the decision-maker flexibility to choose the viable target.

On the other hand, in classical DEA models, analyses can be output oriented or input oriented. That is, we assume that all outputs or all inputs can be modified, or controlled. In real cases, this may not be possible. Some variables are non-controllable due to external or internal factors, yet at the same time, the DEA analysis has to be performed dealing with this fact. The variables that cannot be changed at discretion are called non-discretionary variables. Some authors have presented methods that deal with these situations: first, Banker and Morey [7], then Golany and Roll [8], Camanho et al. [9] and Estelle et al. [10], among others. All these models provide one target for each inefficient DMU.

In this paper, we present an extension of the MORO models to determine a set of targets for inefficient DMUs when there are non-discretionary variables in the analysis. We therefore take advantage of the results of the multiobjective model providing the decision-maker with flexibility, and we take into account non-discretionary variables that are part of the evaluation. To help the decision-maker to choose a target, we also provide a non-radial efficiency index for each target.

We applied this model to a simple and didactic numerical example to show the simplicity of the approach.

This paper is organized as follows. In Section 2, we present the models for non-discretionary variables and the multiobjective models for determining a set of targets in DEA, the MORO models. The proposed multiobjective model with non-discretionary variables is presented in Section 3. After that, in Section 4, we present a numerical example. We present some final comments in Section 5.

2. DEA non-discretionary and MOLP-DEA models

DEA classical models, CCR [1] and BCC [11] use multiple input and multiple outputs to evaluate DMUs. This is carried out by reducing inputs or increasing outputs equiproportionally, known as radial efficiency.

Nonetheless, the radial efficiency, suitable in many cases, may not be appropriate for many real cases. As a result, models that deal with different situations have been presented. The one situation relevant for this study is the existence of non-discretionary variables. They cannot vary at the discretion of the decision-maker. As part of the analysis they have to be taken into account.

On the other hand, for inefficient DMUs, one target is radially determined. In some cases, the unique target determined for inefficient DMUs may not be feasible due to operational or managerial characteristics. Therefore, models have been developed to present alternative targets [3,4]. The multiobjective DEA model called the MORO model [5,6] was introduced as an alternative to determine a set of targets instead of a single target determined by the DEA mono objective models.

In this section, we present a brief review of non-discretionary DEA models as well as one of the models. We also present the MORO model that will be used in this study.

2.1. Non-Discretionary models in DEA

As mentioned previously, in some real cases, DEA classical models do not take into account variables that cannot be controlled or modified due to fixed production factors or external ones. There are some uncontrollable factors that affect efficiency but do not belong to the production process itself. These variables are commonly called environmental variables and they are not included explicitly in the DEA model. To deal with those variables often requires the use of the so-called multistage DEA model.

In this paper, we are not concerned with those kinds of variables. Rather, we are concerned with the variable that belongs to the production process but cannot be modified according to the decision-maker discretion. In other words, they are fixed for each DMU. For example: when evaluating public schools or public hospitals in Brazil, the number of employees can be one of the inputs. What happens is that by law, they cannot be fired. So, in an input oriented approach, the number of employees variable must be treated as fixed, i.e., a non-discretionary variable as defined in the previous paragraph. We will review papers that deal with the kind of non-discretionary variables as we have defined and we will briefly mention papers dealing with other kinds of non-discretionary or environmental variables. For a more detailed explanation about differences between these two kinds of non-discretionary variables see Camanho et al. [9].

The first model was introduced by Banker and Morey [7] and the input oriented variable returns to scale model is presented in (1).

$$
\begin{aligned}
&\text{Min } \theta \\
&\text{subject } to \\
&\sum_{j=1}^{n} y_{rj}\lambda_j \geq y_{rj_0},\ r=1,\ldots,s \\
&\sum_{j=1}^{n} x_{ij}\lambda_j \leq \theta x_{ij_0},\ i \in D \\
&\sum_{j=1}^{n} x_{ij}\lambda_j \leq x_{ij_0},\ i \in ND \\
&\sum_{j=1}^{n} \lambda_j = 1 \\
&\lambda_j \geq 0
\end{aligned}
\qquad (1)
$$

In this model, we can see that inputs are divided into two sets: discretionary (D), that can be controlled, and non-discretionary (ND), or non-controllable. The equiproportional reduction applies only to the discretionary inputs (set D). Therefore, the only difference between this model and the standard variable returns to scale DEA model [11] is the removal of factor θ from the right-hand side of the non-discretionary inputs. This model provides an efficiency index and a unique target for each inefficient DMU, and it has been widely used, for instance, in Fonseca et al. [12] and Soares de Mello et al. [13].

Banker and Morey [7] also provided the output oriented variable returns to scale DEA model for non-discretionary

variables. Analogous to model (1), in the output oriented version, outputs are divided into two groups and then the factor is only multiplied to the controllable outputs.

The constant returns to scale version of this model can be easily formulated with the exclusion of the convexity constraint [14,15]. Another version of this model was introduced by the same authors, Banker and Morey [7].

Golany and Roll [8] extended Banker and Morey's constant returns to scale model to account for non-discretionary variables in both inputs and outputs. Cooper et al. [14] also presented a model for the same problem, which takes into account the existence of non-controllable variables in both the input and output sets.

Concerning the non-discretionary external factors that affect the production process, also known as environmental variables, (see for instance Fried [16,17]), Daraio and Simar [18,19], Muniz [20], Gomes et al. [21], Lima et al. [22] among others.

2.2. The MORO models

In many cases, the single target provided by the DEA classical models may not be feasible, due to operational or managerial problems, or simply because we have additional information about the variables. Some alternative models have been presented by Thanassoulis and Dyson [3] and Zhu [4]. These models require the decision-maker to establish the importance of each input (output) reduction (increase) factor. For each set of value judgments, a single target is determined.

On the other hand, the MORO models [5,6] provide a non-singular set of targets for each DMU and the decision-maker's information is required a posteriori. So, it is not necessary to run the model with every change in the decision-maker's value judgments. Moreover, as we will see in section 3, adapting the MORO models to account for non-discretionary variables requires only a modification in one subset of restrictions.

The most common MORO model is the MORO-D-CRS, presented in (2).

$$
\begin{aligned}
&\text{Max } \phi_1 \\
&\quad \ldots \\
&\text{Max } \phi_s \\
&\text{Min } \varphi_1 \\
&\quad \ldots \\
&\text{Min } \varphi_m \\
&\text{subject } to \\
&\sum_{j=1}^{n} y_{rj}\lambda_j = \phi_r y_{ro}, \; r=1,\ldots,s \\
&\sum_{j=1}^{n} x_{ij}\lambda_j = \varphi_i x_{io}, \; i=1,\ldots,m \\
&\quad \phi_r \geq 1 \\
&\quad \varphi_i \leq 1 \\
&\phi_r, \varphi_i, \lambda_j \geq 0
\end{aligned}
\qquad (2)
$$

This model is very similar to the envelopment version of the CCR model. This model allows each variable to change independently, and not in a radial way as the classical DEA models. The ϕ_r factor is the variation for the output r, φ_i factor is the variation for the input i. We have one objective function for each factor, and we try to maximize the factor for the outputs and minimize the factors for the inputs. The restrictions ensure that we will find projections in the efficient frontier, since the variations of inputs and outputs are independent we replace the inequalities by equalities [5,6]. The last two restrictions guarantee that the outputs will maintain their levels or increase and the inputs will maintain their levels or decrease. In this way we will obtain targets that dominate the DMU o under evaluation, in an approach similar to the Thanassoulis and Dyson mono-objective model [3]. Therefore this model is called MORO with dominance, constant returns to scale, or MORO-D-CRS.

In an approach similar to the Zhu mono-objective model [4], the last two restrictions could be removed. In doing that, any point in the efficient frontier can be a feasible target for the inefficient DMU,. This model is called MORO CRS without dominance, or MORO-CRS.

To illustrate these situations, we present Figs. 1 and 2 from Soares de Mello et al. [23]. In both figures a variable returns to scale frontier is shown. In Fig. 1, possible targets for DMU o using the MORO-VRS model are shows. In Fig. 2, the MORO-D-VRS is used to determine the targets for DMU o. To obtain the variable returns to scale we introduce the convexity restriction (3) in the model. Such a model would be called MORO-VRS or MORO-D-VRS depending whether we consider dominance or not, as in the aforementioned Figures.

$$\sum_{j=1}^{n} \lambda_j = 1 \qquad (3)$$

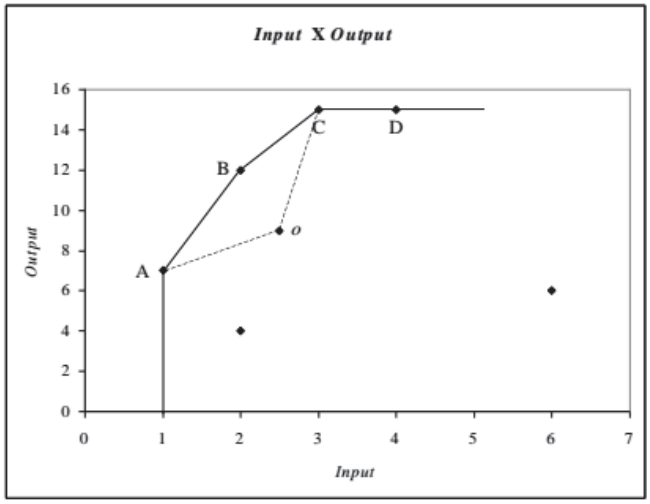


Figure 1.Targets for DMU o using the MORO-VRS model.
Source: [23].

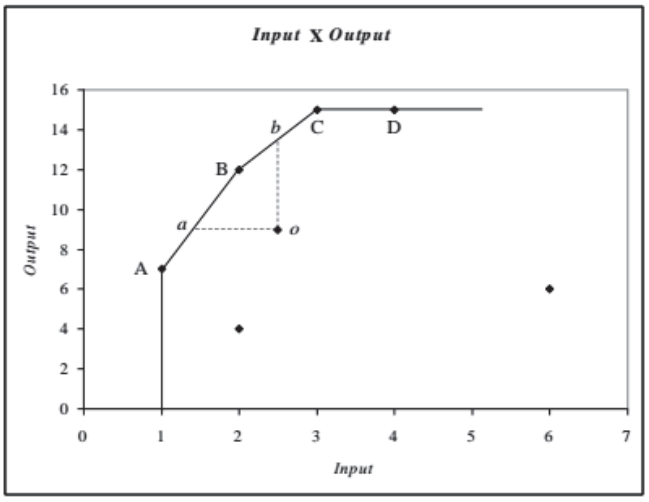


Figure 2.Targets for DMU *o* using the MORO-D-VRS model.
Source: [23].

We have to point out that it is not only the extreme points that are targets for the inefficient DMU, but also the linear combinations of these points that lie in the efficient frontier can be possible targets. This will happene depending on the method used for solving the multiobjective problem. For example, in Fig. 2, for DMU o, the extreme points or targets, are *a*, *B* and *b*. Also, any point in segments *aB* and *Bb* are possible targets for DMU *o*. Therefore, in theory, we will have an infinite set of targets depending on the method used for solving the multiobjective problem [23].

According to Clímaco et al. [24], the MORO models can be classified into the group that uses multiobjective models to solve problems in DEA. Also, according to the same authors the MORO models may be seen as a formalization of the Golany [25] algorithm.

An efficient DMU is on the Pareto efficient frontier and thus $\phi_r^* = \varphi_i^* = 1$, $\forall$ *r, i*, as the equality restrictions of the model require nil value slacks. If this is not the case, the targets for the outputs are given by (4) and the targets for the inputs are given by (5).

$$y_{rj_0}^* = \phi_r^* \, y_{rj_0}, \ \forall r \qquad (4)$$

$$x_{ij_0}^* = \varphi_i^* \, x_{ij_0}, \ \forall i \qquad (5)$$

Therefore, the final value $y_{rj_0}^*$ and $x_{ij_0}^*$ depends on the target chosen by the decision-maker and thus we define the values for ϕ_r^* e φ_i^* among the solutions of the MORO model chosen. In this way, alternative targets can be obtained based on the preferences of the decision-maker.

3. MORO model with non-discretionary variables

As seen previously, the MORO models determine a set of targets for each inefficient DMU. We assume that all variables may change their levels in order to be efficient. In some cases, one of the set's targets may change the level of one variable at a time. Unfortunately, there is no guarantee that the set of targets will always contain a target for any specific non-discretionary variable. Also, the MORO models allow different degrees of change in input and output levels. Thus, to account for non-discretionary variables, we present an extension of the MORO models. The resulting model is in (6) and is called MORO-D-ND, the MORO model with dominance and inequality restrictions with non-discretionary variables, or simply MORO-ND.

$$\begin{aligned}
&\text{Max } \phi_r \qquad \forall r \in D_o \\
&\text{Min } \varphi_i \qquad \forall i \in D_i \\
&\text{subject to} \\
&\sum_{j=1}^{n} y_{rj}^{D_o} \lambda_j = \phi_r y_{rj_0}^{D_o}, \ \forall r \in D_o \\
&\sum_{j=1}^{n} y_{rj}^{ND_o} \lambda_j \geq y_{rj_0}^{ND_o}, \ \forall r \in ND_o \\
&\sum_{j=1}^{n} x_{ij}^{D_i} \lambda_j = \varphi_i x_{ij_0}^{D_i}, \ \forall i \in D_i \\
&\sum_{j=1}^{n} x_{ij}^{ND_i} \lambda_j \leq x_{ij_0}^{ND_i}, \ \forall i \in ND_i \\
&\phi_r \geq 1, \qquad \forall r \in D_o \\
&\varphi_i \leq 1, \qquad \forall i \in D_i \\
&\phi_r, \varphi_i, \lambda_j \geq 0
\end{aligned} \qquad (6)$$

In this model, we have a factor for every discretionary input (D_i) and output (D_o), so the number of objective functions is the number of discretionary variables. We have divided the restrictions of the inputs and outputs in two groups, one that deals with discretionary variables (D_o for outputs and D_i for inputs) and one that deal with non-discretionary variables (ND_o for outputs and D_i *for* inputs). For the first group, as the variables are allowed to change independently, we have a set of equalities, in a similar approach as the MORO models (2). For the second one, in which variables cannot be changed, we set inequalities similar to the envelopment model, in an approach similar to the Banker and Morey's model presented in (2). The last two restrictions of this model are the dominance restrictions, so for the output, we can increase or maintain the level and for the input we can reduce or maintain the level.

As for the other MORO models, we can obtain a set of targets taking into account the variables that are fixed, for any reason, in the analysis. Obviously, the added advantage is that we do not have to specify an orientation (input or output) for the model, because it is a non-radial model.

We can also account for the variable returns to scale introducing the convexity restriction (7).

$$\sum_{j=1}^{n} \lambda_j = 1 \qquad (7)$$

Also, we can find targets without dominance by eliminating the two last restrictions in model (6).

We can also identify an efficient DMU when $\phi_r^* = \varphi_i^* = 1$, $\forall$ *r, i*, as the equality restrictions of the model require nil value slacks. If this is not the case, the targets for the variables are given by equations (8) and (9).

$$y^{*}_{rj_0} = \phi^{*}_{r}\ y_{rj_0},\ \forall r \qquad (8)$$

$$x^{*}_{ij_0} = \varphi^{*}_{i}\ x_{ij_0},\ \forall i \qquad (9)$$

In this case, the non-discretionary variables will maintain their levels. Once again, the alternative targets can be obtained based on the preferences of the decision-maker.

4. Numerical example

In this section, we present an illustrative example with 2 inputs and 1 output. Table 1 presents the data set for the numerical example and the standard BCC efficiency index calculated by the SIAD software [26].

Using the BCC model, DMUs A, B, C, F and H are efficient. Now we turn our attention to the targets for inefficient DMUs. To obtain a set of targets for each inefficient DMU, instead of only one target, we use the MORO-D-ND model in (6) with restriction (7).

For DMU D the model is formulated as in model (10).

$$\begin{aligned}
&\text{Max } \phi \\
&\text{Min } \varphi \\
&\text{subject to} \\
&10\lambda_A + 8\lambda_B + 9\lambda_C + 12\lambda_D + 7\lambda_E + 11\lambda_F + 8\lambda_G + 13\lambda_H = 12\phi \\
&3\lambda_A + 5\lambda_B + 1\lambda_C + 8\lambda_D + 3\lambda_E + 4\lambda_F + 6\lambda_G + 7\lambda_H = 8\varphi \\
&3\lambda_A + 1\lambda_B + 5\lambda_C + 7\lambda_D + 5\lambda_E + 8\lambda_F + 7\lambda_G + 6\lambda_H \geq 7 \\
&\lambda_A + \lambda_B + \lambda_C + \lambda_D + \lambda_E + \lambda_F + \lambda_G + \lambda_H = 1 \\
&\phi \geq 1, \\
&\varphi \leq 1, \\
&\phi_1, \phi_2, \varphi, \lambda_j \geq 0
\end{aligned} \qquad (10)$$

Table 1
Data set for the numerical example

DMU	Input 1	Input 2	Output	Standard BCC Efficiency
A	3	3	10	1,000000
B	1	5	8	1,000000
C	5	1	9	1,000000
D	7	8	12	0,714286
E	5	3	7	0,750000
F	8	4	11	1,000000
G	7	6	8	0,461538
H	6	7	13	1,000000

Source: the authors.

Table 2
Targets and non-radial efficiency indexes for inefficient DMUs.

DMU	Factors		Targets		Non radial Efficiency Index
	φ	ϕ	I2*	O*	
D	**0,687500**	**1,00000**	**5,50000**	**12,00000**	**0,7790**
	0,875000	1,08333	7,00000	12,99996	0,8962
E	0,333333	1,28571	1,00000	8,99997	0,5031
	1,000000	1,46939	3,00000	10,28573	0,7741
G	0,166667	1,12500	1,00000	9,00000	0,4055
	1,000000	1,54167	6,00000	12,33336	0,7516

Source: the authors.

In Table 2 we show set of targets for all inefficient DMUs, Input 1 being the non-discretionary variable. We use TRIMAP [27, 28] to solve the multiobjective model, so we have obtained only basic solutions. We also show an efficiency index using the model developed by Gomes Junior et al. [29] .

The targets obtained by $\phi = 1$ are the same targets provided by the Banker and Morey model (1). Such targets are shown in bold. We shall note that for DMUs E and G, the Banker and Morey target is not present in the set of targets. This is due to the fact that the set of targets found by TRIMAP are limited to the extreme points. We can use other types of software to find other targets in the efficient frontier. For DMU D, the TRIMAP graphic solution is shown in Fig. 3. For a complete interpretation of TRIMAP graphic results with DEA models see Soares de Mello et al. [30].

This Fig. 3 shows the graphic representation of the indifference regions for each basic solution. All solutions contain λ_1 vertex, due to the fact that this vertex corresponds to the non-discretionary variable, which cannot be changed.

We have to point out that the set of targets found by TRIMAP are limited to the extreme point. To obtain Banker and Morey's model results, we have solved their model for DMUs E and G. For DMU E, the model is presented in (11).

$$\begin{aligned}
&\text{Min } \varphi \\
&\text{subject to} \\
&10\lambda_A + 8\lambda_B + 9\lambda_C + 12\lambda_D + 7\lambda_E + 11\lambda_F + 8\lambda_G + 13\lambda_H \geq 12 \\
&3\lambda_A + 5\lambda_B + 1\lambda_C + 8\lambda_D + 3\lambda_E + 4\lambda_F + 6\lambda_G + 7\lambda_H \leq 8\varphi \\
&3\lambda_A + 1\lambda_B + 5\lambda_C + 7\lambda_D + 5\lambda_E + 8\lambda_F + 7\lambda_G + 6\lambda_H \geq 7 \\
&\lambda_A + \lambda_B + \lambda_C + \lambda_D + \lambda_E + \lambda_F + \lambda_G + \lambda_H = 1 \\
&\lambda_j \geq 0
\end{aligned} \qquad (11)$$

For DMUs E and G, the results are in Table 3.

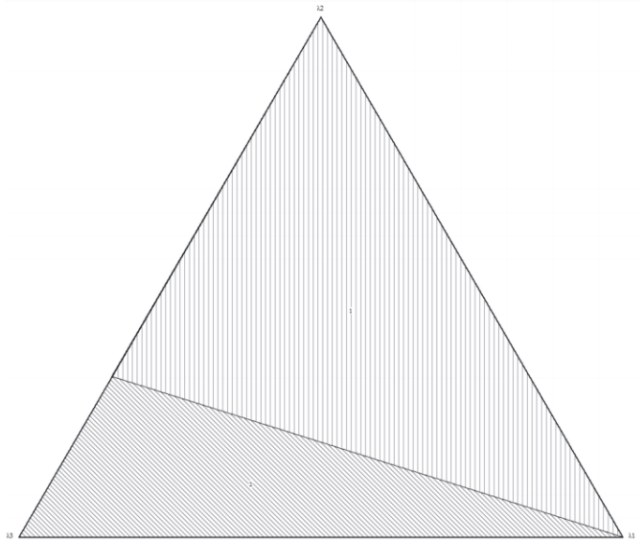


Figure 3.Graphic representation of the solutions for DMU D.
Source: the authors.

Table 3
Results for the Banker and Morey's model

DMU	Factors	Targets	
	φ	I2*	O*
E	0,333333	1,00000	7,00000
G	0,166667	1,00000	8,00000

Source: the authors.

To obtain the same results using model MORO-D-ND (model (6) with restriction (7)), we would have to introduce the restriction $\phi=1$. This clearly shows that the Morey and Banker solutions are dominated by one of the solutions obtained using TRIMAP.

5. Final comments

In this paper, we have presented a multiobjective approach to DEA models with non-discretionary variables. For efficient DMUs, all objective functions equal 1. For inefficient DMUs, we have obtained a set of targets instead of a single target.

We used the TRIMAP to find the results for the multiobjective models, as this method only finds extreme-efficient solutions. For other targets, other methods can be used, such as interactive methods [31,32] to find a suitable solution for the DMU.

However, the use of non-discretionary variables provides more realistic targets for the inefficient DMUs. As each inefficient DMU has more than one target, decision-makers may choose the the most suitable target among them.

Another advantage of the model proposed in this paper when compared with the Banker and Morey model is that, due to the multiobjective approach of the formulation, in our model the targets are also Pareto efficient.

Acknowledgements

We would like to thank CNPq and FAPERJ for their financial support.

References

[1] Charnes, A., Cooper, W.W. and Rhodes, E., Measuring the efficiency of decision-making units, European Journal of Operational Research, 2, pp. 429-444, 1978. DOI: 10.1016/0377-2217(78)90138-8

[2] Patricia, G. and Pellicer, E., Organizational eficiency of consulting engineering firms: Proposal of a performance indicator, DYNA, 76(160), pp. 17-26, 2009.

[3] Thanassoulis, E. and Dyson, R.G., Estimating preferred target input-output levels using Data Envelopment Analysis, European Journal of Operational Research, 56, pp. 80-97, 1992. DOI: 10.1016/0377-2217(92)90294-J

[4] Zhu, J., Data envelopment analysis with preference structure, Journal of the Operational Research Society, 47, pp. 136-150, 1996. DOI: 10.2307/2584258, DOI: 10.1057/jors.1996.12

[5] Lins, M.P.E., Angulo-Meza, L., and Moreira-da Silva, A.C., A multi-objective approach to determine alternative targets in data envelopment analysis, Journal of the Operational Research Society, 55, pp. 1090-1101, 2004. DOI: 10.1057/palgrave.jors.2601788

[6] Quariguasi-Frota Neto, J. and Angulo-Meza, L., Alternative targets for data envelopment analysis through multi-objective linear programming: Rio de Janeiro Odontological Public Health System Case Study, Journal of the Operational Research Society, 58, pp. 865-873, 2007. DOI: 10.1057/palgrave.jors.2602216

[7] Banker, R.D. and Morey, R.C., Efficiency analysis for exogenously fixed inputs and outputs, Operations Research, 34, pp. 513-521, 1986. DOI: 10.1287/opre.34.4.513

[8] Golany, B. and Roll, Y., Some extensions of techniques to handle non-discretionary factors in data envelopment analysis, Journal of Productivity Analysis, 4, pp. 419-432, 1993. DOI: 10.1007/BF01073549

[9] Camanho, A.S., Portela, M.C. and Vaz, C.B., Efficiency analysis accounting for internal and external non-discretionary factors, Computers and Operations Research, 36, pp. 1591-1601, 2009. DOI: 10.1016/j.cor.2008.03.002

[10] Estelle, S.M., Johnson, A.L. and Ruggiero, J., Three-stage DEA models for incorporating exogenous inputs, Computers and Operations Research, 37, pp. 1087-1090, 2010. DOI: 10.1016/j.cor.2009.09.015

[11] Banker, R.D., Charnes, A. and Cooper, W.W., Some models for estimating technical scale inefficiencies in data envelopment analysis, Management Science, 30, pp. 1078-1092, 1984. DOI: 10.1287/mnsc.30.9.1078

[12] Fonseca, A.B.d.M., Soares-de Mello, J.C.C.B., Gomes, E.G. and Angulo-Meza, L., Uniformization of frontiers in non-radial ZSG-DEA models: An application to airport revenues, Pesquisa Operacional, 30, pp. 175-193, 2010. DOI: 10.1590/S0101-74382010000100009

[13] Soares-de Mello, J.C.C.B., Angulo-Meza, L. and Lacerda, F.G., A DEA model with a non discritionary variable for Olympic evaluation, Pesquisa Operacional, 32, pp. 21-29, 2012. DOI: 10.1590/S0101-74382012005000004

[14] Cooper, W. W., Seiford, L., and Tone, K., Introduction to data envelopment analysis and its uses: with DEA-solver software and references. USA: Springer Science, 2006.

[15] Syrjänen, M.J., Non-discretionary and discretionary factors and scale in data envelopment analysis, European Journal of Operational Research, 158, pp. 20-33, 2004. DOI: 10.1016/S0377-2217(03)00362-X

[16] Fried, H.O., Schmidt, S.L. and Yaisawarng, S., Incorporating the operating environment into a nonparametric measure of technical efficiency, Journal of Productivity Analysis, 12, pp. 249-267, 1999. DOI: 10.1023/A:1007800306752

[17] Fried, H.O., Lovell, C.A.K., Schmidt, S.L. and Yaisawarng, S., Accounting for environmental effects and statistical noise in data envelopment analysis, Journal of Productivity Analysis, 17, pp. 157-174, 2002. DOI: 10.1023/A:1013548723393

[18] Daraio, C. and Simar, L., Introducing environmental variables in nonparametric frontier models: A probabilistic approach, Journal of Productivity Analysis, 24, pp. 93-121, 2005. DOI: 10.1007/s11123-005-3042-8

[19] Daraio, C. and Simar, L., Advanced robust and nonparametric methods in efficiency analysis. Methodologies and Applications. New York: Springer, 2007.

[20] Muñiz, M., Paradi, J., Ruggiero, J. and Yang, Z., Evaluating alternative DEA models used to control for non-discretionary inputs, Computers and Operations Research, 33, pp. 1173-1183, 2006. DOI: 10.1016/j.cor.2004.09.007

[21] Gomes, E.G., Soares-de Mello, J.C.C.B., E Souza, G.D.S., Angulo Meza, L. and Mangabeira, J.A.D.C., Efficiency and sustainability assessment for a group of farmers in the Brazilian Amazon, Annals of Operations Research, 169, pp. 167-181, 2009. DOI: 10.1007/s10479-008-0390-6

[22] Lima, V.S., Soares-de Mello, J.C.C.B. and Angulo-Meza, L., Cost-benefit analysis in selected air trips using a non parametric method, African Journal of Business Management, 5, pp. 9678-9685, 2011.

[23] Soares-de Mello, J.C.C.B., Angulo-Meza, L., Gomes, E.G., Serapião, B.P. and Lins, M.P.E., Análise de envoltória de dados no estudo da eficiência e dos benchmarks para companhias aéreas brasileiras, Pesquisa Operacional, 23, pp. 325-345, 2003.

[24] Clímaco, J.C.N., Soares-de Mello, J.C.C.B. and Angulo-Meza, L., Performance measurement - from DEA to MOLP, in Encyclopedia of

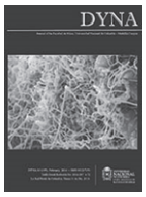
Decision Making and Decision Support Technologies, Adam, F. and Humphreys, P., Eds., ed Pennsylvania: IGI Global, 2008.
[25] Golany, B., An interactive MOLP procedure for the extension of DEA to Effectiveness Analysis, Journal of the Operational Research Society, 39, pp. 725-734, 1988. DOI: 10.2307/2583767, DOI: 10.1057/jors.1988.127
[26] Angulo-Meza, L., Biondi-Neto, L., Soares-de Mello, J.C.C.B., Gomes, E.G. and Coelho, P.H.G., Free software for decision analysis a software package for data envelopment models, in ICEIS 2005 - Proceedings of the 7th International Conference on Enterprise Information Systems, Miami, 2005, pp. 207-212.
[27] Clímaco, J.C.N. and Antunes, C.H., TRIMAP - an interactive tricriteria linear programming package, Foundations of Control Engineering, 12, pp. 101-119, 1987.
[28] Clímaco, J.C.N. and Antunes, C.H., Implementation of an user friendly software package - a guided tour of TRIMAP, Mathematical and Computer Modelling, 12, pp. 1299-1309, 1989. DOI: 10.1016/0895-7177(89)90369-5
[29] Gomes Junior, S. F., Soares de Mello, J. C. C. B., and Angulo-Meza, L., "DEA nonradial efficiency based on vector properties," International Transactions in Operational Research, vol. 20, pp. 341-364, 2013".. DOI: 10.1111/itor.12000
[30] Soares-de Mello, J.C.C.B., Clímaco, J.C.N. and Angulo-Meza, L., Efficiency evaluation of a small number of DMUS: An approach based on Li and Reeves's model, Pesquisa Operacional, 29, pp. 97-110, 2009. DOI: 10.1590/S0101-74382009000100005
[31] Korhonen, P. and Wallenius, J., A Pareto race, Naval Research Logistics, 35, pp. 615-623, 1988. DOI: 10.1002/1520-6750(198812)35:6<615::AID-NAV3220350608>3.0.CO;2-K
[32] Alves, M.J. and Climaco, J.C.N., A review of interactive methods for multiobjective integer and mixed-integer programming, European Journal of Operational Research, 180, pp. 99-115, 2007. DOI: 10.1016/j.ejor.2006.02.033

L. Angulo-Meza, holds a degree in Operations Research Operational Investigación - Universidad Nacional Mayor de San Marcos (1996), a MSc. degree in Production Engineering from COPPE, Federal University of Rio de Janeiro, COPPE/UFRJ (1998) and a PhD. in Industrial Engineering from COPPE Federal University of Rio de Janeiro COPPE/UFRJ (2002). She is currently an associate professor at the Federal Fluminense University (UFF), Department of Production Engineering, Metallurgical Industrial School in Volta Redonda. She has a fellowship from the CNPq and FAPERJ's Young Scientist Program. She has conducted a post-doctoral internship at the Instituto Superior Técnico, Technical University of Lisbon. She has served as the head of the Production Engineering Department of Volta Redonda (VEP), and has experience in Production Engineering, with emphasis in Operations Research, acting on the following topics: Data envelopment analysis (DEA - Data Envelopment Analysis), Multiobjective Linear Programming, Neural Networks, Computational Implementations and Educational Evaluation.
ORCID: 0000-0003-4557-0210

J.C.C.B. Soares-de Mello, is graduated in Mechanical Engineering from the Federal Fluminense University, Brazil (1981), MSc. in Mathematics from the Federal Fluminense University, Brazil (1987), and PhD. in Production Engineering from the Federal University of Rio de Janeiro, Brazil (2002). He has conducted post-doctoral internship at the Faculty of Economics, University of Coimbra in 2006 and senior internship at Instituto Superior Técnico, Technical University of Lisbon in 2012. He is currently an associate professor at the Universidade Federal Fluminense, Brazil. In the administrative area, he has served as department head at the UFF of Applied Mathematics, coordinator of the basic cycle of courses in engineering, and operations coordinator of the vestibular commission department. He is currently the deputy coordinator of the post-graduate program in Production Engineering. He has experience in Production Engineering, with emphasis in Data Envelopment Analysis, acting on the following topics: data envelopment analysis (DEA), multicriteria, educational assessment, air transport and sports evaluation.
ORCID: 0000-0002-6507-2721

S.F. Gomes-Jr. holds a PhD. (2010) and Msc. (2006) in Production Engineering from Universidade Federal Fluminense, Brazil graduated in Civil Engineering from Universidade Federal Fluminense, Brazil (1995) and a degree in Mathematics from the Center for Distance Education of the State of Rio de Janeiro, Brazil (2005). He is currently an associate professor, State University Center-West Zone UEZO and a professor at Redeemer College. He has experience in production engineering, with emphasis in operations research, acting on the following subjects: mathematics, multicriteria, data envelopment analysis and education.
ORCID: 0000-0002-8159-3965

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS



DYNA
http://dyna.medellin.unal.edu.co/

Kinetic aspects of a dried thin layer carrot in a heat pump dryer

Juan Carlos Gómez-Daza [a] & Claudia Isabel Ochoa-Martínez [a]

[a]*Escuela de Ingeniería de Alimentos, Universidad del Valle, Cali, Colombia. juan.gomez.d@correounivalle.edu.co*
[a]*Escuela de Ingeniería de Alimentos, Universidad del Valle, Colombia, claudia.ochoa@correounivalle.edu.co*

Received: November 6th, 2014. Received in revised form: August 20th, 2015. Accepted: December 18th, 2015

Abstract
This article presents a mathematical model for drying thin layer carrot slices (*Daucus carota*) using a heat pump dryer (HPD). To select the equation that best describes the drying curve, 10 semi-theoretical and/or empirical models were evaluated. The parameters were determined using the Sigma-Plot® program, and their goodness of fit was compared using the correlation coefficient, R^2; Chi-squared, χ^2; standard error of the estimate (*SEE)* and root mean square error (*RMSE)*. Additionally, the effect of the relative moisture, sample thickness and air velocity on the effective diffusivity of the process was evaluated using a response surface tool. Although all the models correctly fit the experimental data, based on the statistical tests, the Wang-Singh model was selected as the best.

Keywords: drying; thin layer; modeling; diffusivity.

Aspectos cinéticos del secado de capa delgada de zanahoria en un secador de bomba de calor

Resumen
Este artículo presenta un modelo matemático de capa delgada para secado de rodajas de zanahoria (*Daucus carota*) usando un secador de bomba de calor (HPD). Para seleccionar la mejor ecuación que describe la curva de secado, se evaluaron 10 modelos semi-teóricos y/o empíricos. Los parámetros se determinaron usando el programa Sigma-Plot® y la bondad de su ajuste se comparó usando el coeficiente de correlación R^2; Chi-cuadrado, χ^2; error estándar del estimado (*SEE)* y raíz del error cuadrado medio (*RMSE)*. Adicionalmente, se evaluó el efecto de la humedad relativa, el espesor de la muestra y la velocidad del aire sobre la difusividad efectiva del proceso usando la herramienta de superficie de respuesta. Aunque todos los modelos ajustaron correctamente los datos experimentales, se seleccionó el modelo de Wang-Singh como el mejor, basado en las pruebas estadísticas.

Palabras clave: secado; capa delgada; modelación; difusividad.

1. Introduction

Regarding the drying process, the term "thin-layer" applies to either a particle suspended freely in the drying air, or one layer of particle or a polylayer of many particles' thicknesses; the temperature and relative moisture of the drying air can be considered to be in the same thermodynamic state during the drying period [1]. Using this definition, any mathematical model for a particle also models the particles drying in a thin layer using any drying method, and the thin layer thickness may change with the velocity, temperature and relative moisture of the drying air.

The thickness of a thin layer increases if the drying air speed increases and when the thermodynamic state of the drying air reaches equilibrium with dry particles in the layer [1]. Due to the thin sample structure, a uniform temperature distribution can be assumed and may be modeled using lumped parameters [2].

The thin layer equations can be theoretical, semi-theoretical or empirical. The first only accounts for internal resistance to the moisture transfer between the product and air, while the others only consider external resistances to this moisture transfer. Theoretical models explain the product's behavior during drying and can be used for all process conditions despite including many assumptions that cause considerable error. The most widely used theoretical models

are derived from Fick's second law of diffusion. Similarly, semi-theoretical models are generally derived from Fick's second law and modified to a simplified form. However, using experimental data requires making assumptions, and these theories are only valid within the applied process conditions. Empirical models have similar features to semi-theoretical models that strongly depend on the experimental conditions and provide limited information on the product behavior during drying [2]. The carrot is frequently used in studies of different preservation techniques due to its physical characteristics and available modeling and simulation data [3-12].

This article evaluates the experimental data fit for 10 thin layer models and explores the effect that relative moisture, sample thickness and air speed have on the kinetic behavior of dried carrot slices. The kinetic expression was established from analyzing the response surfaces and kinetic parameters controlling the heat pump drying process. Additionally, we determined the effective moisture diffusivity for each experiment and the activation energy of the processes.

2. Materials and methods

Nantes carrots (*Daucus carota L.*) were washed and cut into slices 3 cm in diameter and with a thickness of 2, 3 and 4 mm, according to the experimental design. The average initial moisture content of a fresh carrot was 0.8935 ± 0.024 (bh). The heat pump dryer (HPD) (Dártico brand) consisted of a cooling circuit containing a condenser, evaporator, compressor and expansion valve. The drying chamber contained 5 trays (0.36 × 0.36 m). The relative humidity (*RH*) and air velocity in the HPD were fixed. The drying temperature was directly linked to the *RH* (Table 1). A three-factor face-centered central composite design (FCCCD) was used with three repetitions at the central point, as shown in Table 1. The total weight (trays plus samples) was recorded for 5 hours with 2 min intervals for the first 10 minutes, 5 min intervals for half an hour, 10 min intervals for an hour and a half and 30 minute intervals for the last two hours.

Table 1.
A three-factor face-centered central composite design (FCCCD)

	level			*value*		
Experiment	x_1	x_2	x_3	***RH* (%)**	***e* (mm)**	***V* (m/s)**
1	-1	-1	-1	20 (50°C)	2	0.8
2	1	-1	-1	50 (35°C)	2	0.8
3	-1	1	-1	20 (50°C)	4	0.8
4	1	1	-1	50 (35°C)	4	0.8
5	-1	-1	1	20 (50°C)	2	1.2
6	1	-1	1	50 (35°C)	2	1.2
7	-1	1	1	20 (50°C)	4	1.2
8	1	1	1	50 (35°C)	4	1.2
9	-1	0	0	20 (50°C)	3	1.0
10	1	0	0	50 (35°C)	3	1.0
11	0	-1	0	35 (40°C)	2	1.0
12	0	1	0	35 (40°C)	4	1.0
13	0	0	-1	35 (40°C)	3	0.8
14	0	0	1	35 (40°C)	3	1.2
15	0	0	0	35 (40°C)	3	1.0
16	0	0	0	35 (40°C)	3	1.0
17	0	0	0	35 (40°C)	3	1.0

Source: The authors

Table 2 shows the semi-theoretical models evaluated during the kinetic study [13]. The parameters for each model were estimated using Sigma-Plot® software. The moisture ratio (*MR*) value was determined according to the external conditions. If the relative humidity of the drying air is constant during the process, the equilibrium moisture content is also constant. The *MR* value was calculated using eq. (1).

$$MR = \frac{M_t - M_e}{M_i - M_e} \quad , \tag{1}$$

where *Mi* is the initial moisture content (dry basis) obtained by drying in a vacuum oven at a constant weight and at 60ºC, M_t is the moisture content (dry basis) at time *t* based on the recorded weight, and M_e is the equilibrium moisture content obtained from the carrot sorption isotherms (dry basis) [14, 15]: 0.055 [20% *RH*; 50 °C], 0.070 [35% *RH*; 40 °C] and 0.080 [50% *RH*; 35 °C].

The model validities were checked using statistical parameters: correlation coefficient (R^2), reduced Chi-squared test (χ^2), standard error of the estimate (*SEE*) and root mean square error (*RMSE*). The highest R^2 and lowest χ^2, *SEE* and *RMSE* values determined the goodness of fit. The aforementioned criteria were calculated using eq. (2) – (5) [2,16].

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(MR_{pred,i} - MR_{exp,i}\right)^2}{\sum_{i=1}^{N}\left(\overline{MR_{pred}} - MR_{pred,i}\right)^2} \tag{2}$$

$$\chi^2 = \frac{\sum_{i=1}^{N}\left(MR_{exp,i} - MR_{pred,i}\right)^2}{N - n} \tag{3}$$

$$SEE = \left[\frac{\sum_{i=1}^{N}\left(MR_{pred,i} - MR_{exp,i}\right)^2}{N - n}\right]^{1/2} \tag{4}$$

$$RMSE = \left[\frac{1}{N}\sum_{i=1}^{N}\left(MR_{pred,i} - MR_{exp,i}\right)^2\right]^{1/2} , \tag{5}$$

where *N* is the number of observations, *n* is the number of constants, $MR_{pred,I}$ is the *i*th predicted moisture ratio and $MR_{exp,I}$ is the *i*th experimental moisture ratio.

The Root Mean Squared Error has the advantage that it retains the units of the forecast variable and is thus more easily interpretable as a typical error magnitude. The Chi-squared Test is easy to implement (with multivariable data for example) [17] and also quite flexible..

The Henderson-Pabis model was rearranged to determine the diffusivity coefficient based on eq. (6) [2]:

$$ln(MR) = ln(a) - kt \, , \tag{6}$$

where a is a shape index and, k is the drying constant defined by eq. (7):

$$k = -\pi^2 D_{eff}/A_2 \ , \tag{7}$$

Table 2.
Kinetic parameters for thin layer models

Model	Page		Henderson-Pabis		Logarithm			Two terms				Weibull	
	$MR = exp(-kt^n)$		$MR = a\ exp(-kt)$		$MR = a\ exp(-kt) + c$			$MR = a\ exp(-k_0 t) + b\ exp(-k_1 t)$				$MR = exp\left[-\left(\frac{t}{k}\right)^n\right]$	
EXP	k	n	k	n	k	a	c	k_o	a	k_1	b	k	n
1	0.00315	1.109	0.00550	1.010	0.00451	1.124	-0.1215	0.0503	-0.123	0.0061	1.092	180.4	1.109
2	0.00053	1.234	0.00194	1.023	0.00041	4.029	-3.0140	0.0193	-0.071	0.0022	1.080	451.4	1.234
3	0.00177	1.277	0.00719	1.050	0.00548	1.204	-0.1676	0.0414	-0.257	0.0087	1.247	142.9	1.277
4	0.00137	1.059	0.00187	1.002	0.00084	1.969	-0.9724	0.0030	2.796	0.0038	-1.800	505.9	1.059
5	0.00154	1.365	0.00906	1.055	0.00598	1.286	-0.2559	0.0163	7.524	0.0184	-6.535	115.2	1.365
6	0.00373	0.983	0.00331	0.986	0.00218	1.330	-0.3523	0.0026	1.070	-0.0017	-0.092	295.3	0.983
7	0.00132	1.388	0.0088	1.062	0.00645	1.228	-0.1845	0.0129	1.822	0.0278	-0.836	118.5	1.388
8	0.00286	1.023	0.00315	0.992	0.00193	1.421	-0.4379	0.0078	-0.473	0.00423	1.453	306.4	1.023
9	0.00054	1.393	0.00410	1.029	0.00106	2.929	-1.9230	0.0263	-0.395	0.0058	1.346	223.5	1.393
10	0.00006	1.627	0.00210	1.089	0.00069	2.816	-1.7360	0.4419	-0.103	0.0022	1.105	401.0	1.627
11	0.00015	1.493	0.00266	1.109	0.00280	1.070	0.0394	1.0620	-0.123	0.0027	1.123	361.3	1.493
12	0.00222	1.109	0.00396	1.008	0.00219	1.509	-0.5141	0.0021	1.615	0.0002	-0.621	246.6	1.109
13	0.00130	1.165	0.00319	1.023	0.00194	1.477	-0.4623	0.0036	1.095	0.0219	-0.089	299.3	1.165
14	0.00081	1.295	0.00405	1.057	0.00222	1.600	-0.5574	0.2265	-0.096	0.0042	1.083	243.8	1.295
15	0.00305	1.049	0.00389	0.999	0.00277	1.255	-0.2649	0.0071	-1.572	0.0056	2.559	250.2	1.049
16	0.00061	1.300	0.00341	1.083	0.00355	1.055	0.0288	84.230	-0.097	0.0035	1.093	297.8	1.300
17	0.00457	0.967	0.00371	0.980	0.00302	1.124	-0.1483	0.0039	1.011	0.0369	-0.045	263.2	0.967

Source: The authors

Table 2.
Kinetic parameters for thin layer models (cont.)

Model	Newton		Modified Henderson-Pabis					Wang-Singh		Vega-Lemus		Proposed model	
	$MR = exp(-Kt)$		$MR = a\ exp(-kt) + b\ exp(-gt) + c\ exp(-ht)$					$MR = 1 + at + bt^2$		$MR = (n + kt)^n$		$MR = ab^t$	
EXP	K	k	a	g	b	h	c	a	b	k	n	a	b
1	0.00537	0.2234	0.385	0.1420	-0.457	0.0060	1.075	-0.0047	6.74E-06	-0.0021	0.992	1.0202	0.9944
2	0.00179	0.0083	-0.179	0.4070	-0.017	0.0025	1.195	-0.0014	-3.5E-07	-0.0009	1.009	1.0267	0.9980
3	0.00667	0.0085	1.212	0.0602	-0.291	0.1712	0.080	-0.0055	8.16E-06	0.0027	-1.009	1.0897	0.9921
4	0.00186	0.0016	1.000	-0.0153	-1.248	-0.0153	1.244	-0.0017	7.9E-07	-0.0008	0.999	1.0039	0.9981
5	0.00845	0.0105	6.951	0.0074	-2.276	0.0142	-3.676	-0.0065	1.04E-05	-0.0034	1.010	1.2592	0.9873
6	0.00342	0.0066	-1.751	0.2636	0.040	0.0051	2.712	-0.0032	3.4E-06	-0.0014	0.987	0.9925	0.9966
7	0.00809	0.0034	-0.641	0.0124	1.605	0.3702	0.037	-0.0064	0.000011	0.0033	-1.015	1.1645	0.9892
8	0.00322	0.0038	1.155	884.10	0.026	0.0122	-0.181	-0.0029	2.9E-06	0.0013	-0.990	0.9982	0.9968
9	0.00388	0.0526	-0.485	0.0055	1.249	0.1703	0.239	-0.003	1.3E-06	0.0017	-1.008	1.0560	0.9955
10	0.00154	0.7902	-0.086	0.0023	1.119	0.0426	-0.032	-0.0007	-3E-06	0.0010	-1.041	1.0923	0.9979
11	0.00195	0.8034	-0.132	-0.0004	0.158	0.0036	0.974	-0.0012	-1.9E-06	-0.0012	1.048	1.1071	0.9974
12	0.00385	0.0060	2.179	0.0088	-1.202	0.1764	0.025	-0.0033	0.000003	-0.0016	0.996	1.0229	0.9959
13	0.00303	0.0037	1.143	0.0125	-0.122	0.1195	-0.022	-0.0026	1.6E-06	-0.0014	1.006	1.0299	0.9967
14	0.00361	0.0027	1.491	0.3389	-0.073	0.0003	-0.430	-0.0028	1.05E-06	-0.0017	1.020	1.0758	0.9957
15	0.00390	0.0051	1.496	53.980	0.015	0.0089	-0.512	-0.0035	3.8E-06	0.0016	-0.992	1.0086	0.9960
16	0.00282	0.0204	0.061	0.0032	1.046	1.0140	-0.111	-0.0023	4.8E-07	0.0015	-1.033	1.0815	0.9967
17	0.00387	0.2358	1.229	0.2124	-1.230	0.0039	1.003	-0.0037	4.7E-06	0.0015	-0.983	0.9847	0.9962

Source: The authors

Table 3.
Statistics parameters

Model	R^2		*SEE*		χ^2		*RMSE*	
Newton	0.9771	±0.0233	0.0352	±0.0216	0.0017	±0.0017	0.0345	±0.0211
Page	0.9913	±0.0119	0.0219	±0.0162	0.0007	±0.0010	0.0211	±0.0156
Henderson-Pabis	0.9916	±0.0104	0.0239	±0.0142	0.0008	±0.0009	0.0229	±0.0137
Logarithm	0.9950	±0.0067	0.0184	±0.0117	0.0005	±0.0006	0.0173	±0.0110
Two terms	0.9979	±0.0031	0.0116	±0.0075	0.0002	±0.0003	0.0106	±0.0069
Wang-Singh	0.9986	±0.0023	0.0087	±0.0065	0.0001	±0.0002	0.0083	±0.0062
M Henderson-Pabis	0.9986	±0.0023	0.0095	±0.0071	0.0001	±0.0002	0.0084	±0.0062
Vega-Lemus	0.9947	±0.0070	0.0183	±0.0116	0.0005	±0.0006	0.0176	±0.0111
Weibull	0.9913	±0.0120	0.0218	±0.0163	0.0007	±0.0010	0.0209	±0.0157
This paper	0.9866	±0.0196	0.0311	±0.0287	0.0017	±0.0035	0.0305	±0.0281

Source: The authors

where A_2 is a geometric constant ($4e^2$ for infinite slices), and e is half the slice thickness if drying occurs on both sides and the full thickness if drying occurs on only one side. Eq. (6) indicates *ln(MR)* varies linearly with t and the slope is equal to k.

Eq. (8) was used to calculate the activation energy (E_a):

$$ln(D_{eff}) = ln(D_0) - 10^3\left(\frac{E_a}{R}\right)\left(\frac{1}{T + 273.15}\right) \quad , \qquad (8)$$

where D_0 (m^2/s) is the Arrhenius factor that is generally defined as the reference diffusion coefficient at an infinitely high temperature, E_a (kJ/mol) is the diffusion activation energy, and R (kJ/kmolK) is the universal gas constant. Eq. (8) is linear and allows E_a to be calculated from the slope.

3. Results and discussion

Table 2 shows the parameters for each model that fit the experimental data and Table 3 shows the fit for each model using the statistical parameters from each experiment. In general, all models had high correlation coefficients and low values for the other statistics. The Wang-Singh and modified Henderson-Pabis models exhibited the best values; however, the Wang-Singh model has fewer adjustable parameters. A statistical discrimination study based on nonlinear regression ensures the Wang-Singh equation best fits the data [18].

Thin layer equations may be theoretical, semi-theoretical, and empirical models. Semi-theoretical models are generally derived from Fick´s second law of diffusion and are modifications of its simplified forms [2].

Empirical models used are: The Lewis (Newton) model, which is analogous with Newton´s law of cooling. The Page model modifies the Newton model to get a more accurate model by adding a dimensionless empirical constant (n) [2].

The models derived from Fick´s second law of diffusion are: the Henderson-Pabis model, and the Logarithmic model, which is formed by adding an empirical term. The Two-Term model uses the first two terms of the general series solution of Fick´s second law of diffusion to correct the shortcomings of the Henderson-Pabis model. The Modified Henderson-Pabis model improves previous models by adding the third term of the general solution from Fick´s second law of diffusion [2].

The Weibull, Wang-Singh, Vega-Lemus and Proposed models, are all empirical models.

Fig. 1 shows the drying curves and their fit to the Wang-Singh model for 5 of the 17 experiments. The first number in the code corresponds to the relative humidity (20 and 50%), the second is the thickness (2 and 4 mm), and the last is the air velocity (0.8 and 1.2 m/s).

The relative humidity and air velocity affect the drying time; a lower relative humidity and higher velocity provide faster drying times regardless of the sample thickness. Krokida [3] found temperature to be the most important factor in the drying rate, while the velocity and humidity have lesser effects. In that study, the evaluated temperatures were higher (65, 75 and 85 °C).

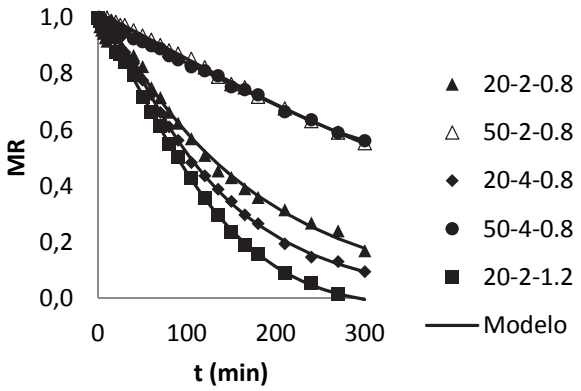


Figure 1. Drying curves modeling by using the Wang-Singh model
Source: The authors

The effective diffusivity (D_{eff}) was determined using eq. (6) - (7) and the least squares analysis for each experiment. Fig. 2 shows the response surfaces for the effective diffusivity, D_{eff}. Larger effective diffusivities were obtained at lower relative humidities, higher material thicknesses and high air velocities. The desirability function confirms these results. The D_{eff} values were between 2.01E-10 and 4.38E-9 m^2/s. Doymaz [19] determined the effective diffusivity values for the convective drying of carrots ranged from 7.76E-10 to 9.34E-10 m^2/s and argues that this variable increases when air flow and temperature are increased ; these are similar to this study's observations (Fig. 2), which account for the decreased *RH* at an increased temperature (Table 1). Kaya et al. [14] also found that decreasing the *RH* (temperature increase for a closed system) increased the effective diffusivity values. However, Phoungchandang et al. [20] determined that the effective diffusivity values ranged from 8.34E-11 to 2.77E-10 m^2/s for dried carrots in a heat pump dryer at 40, 50 and 60 °C with an air velocity of 0.5 m/s. Panagiotou et al. [21] reported D_{eff} values (m^2/s) ranging from 2.20E-12 to 7.46E-9 for moisture contents between 0.10 and 15.0 on a dry basis and temperatures between 20 and 100 °C. Torres et al. [22] also obtained similar results working with yam in a convective drier.

The p value obtained from the ANOVA for the effective diffusivity, D_{eff}, was less than 0.05 for the regression term. This means that there is a statistically significant relationship between the variables on a 95% confidence level, specifically between the quadratic and interaction terms. The linear terms can be omitted from the model because they would have no significance at this confidence level. This is corroborated by the p-values for the regression coefficients in which only the

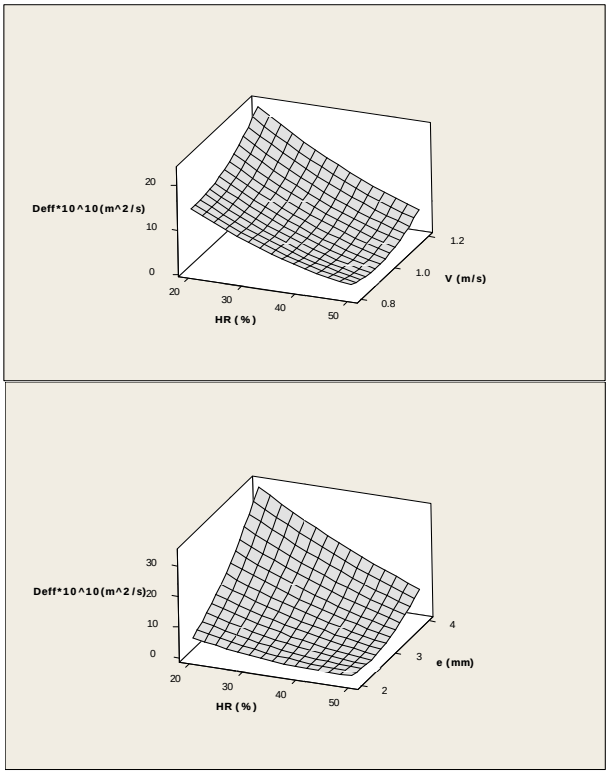


Figure 2. Effect of *RH* (%), sample thickness and air velocity on effective diffusivity D_{eff} (m^2/s)
Source: The authors

second order term (*RH* * *e*) was significant and is effectively an interaction. The *p* value is the specific probability that the observed value of the test statistic, together with all other possible values of the test statistic that are at least as unfavorable as the null hypothesis, will occur [17].

Validation of the model via a residual analysis only happens when the probability curve that shows normal error behavior is observed. Residuals are not far from the line; the variance is homogeneous, and the largest deviation occurred in experiment 9.

The activation energy obtained was 35.50 kJ/mol, and the literature reports activation energies for carrot drying using different equipment and different processing conditions.

According to the energy levels involved and collision theory for reactive molecules, enough energy must be generated to provide the required activation energy and facilitate the reaction. The activation energy itself does not provide information on the reactivity of a given system, only on the temperature dependence of the reaction. Activation Energy activation is also related to the moisture content. The diffusion activation energy increased at lower moisture contents because the interaction between the moisture and solid is generally stronger at lower moisture contents [23].

4. Conclusions

All of the thin layer models evaluated could fit the experimental data. The Wang-Singh model exhibited the best fit for the experimental data, based on the criteria of having the highest correlation coefficient (R^2) and lowest Chi-squared statistic (χ^2), standard error of the estimate (*SEE*) and root mean square error (*RMSE*) values. A significant effect from the *HR*e* interaction was found during the ANOVA for effective diffusivity. The conditions that provide the best evaluated parameters (low *MR* and high D_{eff}) are: 20% *HR*, 4 mm thickness and 1.2 m/s air velocity.

References

[1] Ayas, D.S., Cenkowski, S., Pabis, S. and Muir, W.E., Review of thin-layer drying and wetting equations. Drying Technology, 9(3), pp. 551-588, 1991.

[2] Erbay, Z. and Icier, F., A review of thin layer drying of foods: Theory, modeling, and experimental results. Critical Review in Food Science and Nutrition, 50(5), pp. 441-464, 2010. DOI: 10.1080/10408390802437063

[3] Krokida, M.K., Karathanos, V.T., Maroulis, Z.B. and Marinos-Kouris, D., Drying kinetics of some vegetables. Journal of Food Engineering, 59, pp. 391-403, 2003. DOI: 10.1016/S0260-8774(02)00498-3

[4] Srikiatden, J. and Roberts, J.S., Measuring moisture diffusivity of potato and carrot (core and cortex) during convective hot air and isothermal drying. Journal of Food Engineering, 74(1), pp. 143-152, 2006. DOI: 10.1016/j.jfoodeng.2005.02.026

[5] Togrul, H., Suitable drying model for infrared drying of carrot. Journal of Food Engineering, 77(3), pp. 610-619, 2006. DOI: 10.1016/j.jfoodeng.2005.07.020

[6] Erenturk, S. and Erenturk, K., Comparison of genetic algorithm and neural network approaches for the drying process of carrot. Journal of Food Engineering, 78(3), pp. 905-912, 2007. DOI: 10.1016/j.jfoodeng.2005.11.031

[7] Singh, B. and Gupta, A.K., Mass transfer kinetics and determination of effective diffusivity during convective dehydration of pre-osmosed carrot cubes. Journal of Food Engineering, 79(2), pp. 459-470, 2007. DOI: 10.1016/j.jfoodeng.2006.01.073

[8] Mihoubi, D., Timoumi, S. and Zagrouba, F., Modelling of convective drying of carrot slices with IR heat source. Chemical Engineering and Processing: Process Intensification, 48(3), pp. 808-815, 2009. DOI: 10.1016/j.cep.2008.10.004

[9] Phongsomboon, P. and Intipunya, P., Comparative study on drying of osmotic treated carrot slices. Asian Journal of Food and Agro-Industry, 2(04), pp. 448-456, 2009.

[10] Hiranvarachat, B., Devahastin, S. and Chiewchan, N., Effects of acid pretreatments on some physicochemical properties of carrot undergoing hot air drying. Food and Bioproducts Processing, 89(2), pp. 116-127, 2011. DOI: 10.1016/j.fbp.2010.03.010

[11] Zielinska, M. and Markowski, M., Air drying characteristics and moisture diffusivity of carrot. Chemical Engineering and Processing: Process Intensification, 49(2), pp. 212-218, 2010. DOI: 10.1016/j.cep.2009.12.005

[12] Nazghelichi, T., Kianmehr, M.H. and Aghbashlo, M., Thermodynamic analysis of fluidized bed drying of carrot cubes. Energy, 35(12), pp. 4679-4684, 2010. DOI: 10.1016/j.energy.2010.09.036

[13] Xanthopoulos, G., Oikonomou, N. and Lambrinos, G., Applicability of a single-layer drying model to predict the drying rate of whole figs. Journal of Food Engineering, 81(3), pp. 553-559, 2007. DOI: 10.1016/j.jfoodeng.2006.11.033

[14] Kaya, A., Aydin, O. and Demirtas, C., Experimental and theoretical analysis of drying carrots. Desalination, 237(1-3), pp. 285-295, 2009. DOI: 10.1016/j.desal.2008.01.022

[15] Eim, V.S., Roselló, C., Femenia, A. and Simal, S., Moisture sorption isotherms and thermodynamic properties of carrot. International Journal of Food Engineering, 7(3), Article 13, pp. 1-16, 2011. DOI: 10.2202/1556-3758.1804

[16] Vega-Gálvez, A., Andrés, A., Gonzalez, E., Notte-Cuello, E., Chacana, M. and Lemus-Moncada, R., Mathematical modeling on the drying process of yellow squat lobster (Cervimunida jhoni) fishery waste for animal feed. Animal Feed Science and Technology, 151(3-4), pp. 268-279, 2009. DOI: 10.1016/j.anifeedsci.2009.01.003

[17] Wilks, D.S., Statistical methods in the atmospheric sciences. Second Edition. Amsterdam: Academic Press, 2006.

[18] Barrozo, M.A.S., Sartori, D.J.M. and Freire, J.T., A study of the statistical discrimination of the drying kinetics equations. Food and Bioproducts Processing, 82(C3), pp. 219-225, 2004. DOI: 10.1205/fbio.82.3.219.44176

[19] Doymaz, I., Convective air drying characteristics of thin layer carrots. Journal of Food Engineering, 61(3), pp. 359-364, 2004. DOI: 10.1016/S0260-8774(03)00142-0

[20] Phoungchandang, S. and Wongwatanyoo, J., Desorption isotherms and drying characteristics of carrot using tray and heat pump-assisted dehumidified drying. KKU Research Journal, 15(3), pp. 171-186, 2010.

[21] Panagiotou, N.M., Krokida, M.K., Maroulis, Z.B. and Saravacos, G.D., Moisture diffusivity: Literature data compilation for foodstuffs. International Journal of Food Properties, 7(2), pp. 273-299, 2004. DOI: 10.1081/JFP-120030038

[22] Torres, R., Montes, E.J., Andrade, R.D., Perez, O.A. and Toscano, H., Drying kinetics of two yam (Dioscorea alata) varieties. DYNA, 79(171), pp. 175-182, 2012.

[23] Kahveci, K. and Cihan, A., Drying of food materials: Transport Phenomena. New York: Nova Science Publishers, Inc., 2008.

J.C. Gómez-Daza, received his BSc. in Chemical Engineering in 1996, his MSc. in Chemical Engineering in 1999, and his PhD in Engineering in 2014, all from the Universidad del Valle, Cali, Colombia. From 1987 to 1991, he worked for the Industria de Licores del Valle and Lloreda Grasas and since 2000 he has worked at the Universidad del Valle. Currently, he is a full professor in the Food Engineering School, Facultad de Ingeniería, at the Universidad del Valle. He works as an occasional lecturer at the Universidad Nacional de Colombia, in Palmira, in the areas of process dynamic physicochemical and biological, mathematical and numeric methods and drying. His research interests include: modeling, simulation and drying; process engineering; process dynamics.
ORCID: 0000-0001-7464-0519

C.I. Ochoa-Martínez, received his BSc. Chemical Engineering in 1989, his MSc. in Chemical Engineering in 2001, and his PhD in Engineering in 2006. She has worked in programs and projects in the food area since 2006 at the Universidad del Valle. She is currently a full professor in the Food Engineering School, Facultad de Ingeniería, Universidad del Valle. Her research interests include: modeling, simulation and drying; process engineering. She has several publications in scientific journals and is currently coordinator of the post-graduate program at the Food Engineering School.
ORCID:0000-0002-2666-1726

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS

DYNA
http://dyna.medellin.unal.edu.co/

Factors influencing the occurrence of traffic accidents in urban roads: A combined GIS-Empirical Bayesian approach

Víctor Cantillo [a], Patricia Garcés [b] & Luis Márquez [c]

[a] *Departamento de Ingeniería Civil y Ambiental, Universidad del Norte, Barranquilla, Colombia. vcantill@uninorte.edu.co*
[b] *Facultad de Ingeniería, Programa de Ingeniería Civil, Universidad de Cartagena, Cartagena, Colombia. pgarcesd@unicartagena.edu.co*
[c] *Programa de Ingeniería de Transporte y Vías, Universidad Pedagógica y Tecnológica de Colombia, Tunja, Colombia. luis.marquez@uptc.edu.co*

Received: November 14th, 2014. Received in revised form: August 26th, 2015. Accepted: December 5th, 2015.

Abstract
The problem of urban road accidents in Colombia is remarkable and has a significant magnitude. For this reason, a technical study of this important public health scourge is important. The quantitative techniques employed are usually highly aggregated and will not correctly identify the determinant variables of the problem. This paper examines the relationship between urban road accidents and variables related to road infrastructure, environment, traffic volumes and traffic control. Some accident-prone sections in the city of Cartagena (Colombia) are specifically identified by the empirical Bayesian method based on GIS. A total of 69 accident-prone sections were identified in the city. It was evident that the marginal effect on the accident rate for motorcycles is well above that for cars and buses. Empirical evidence also showed that the sections located in commercial areas tend to have higher frequency of accidents due to the high presence of pedestrians.

Keywords: urban road accident; accident-prone sections; empirical Bayesian approach; Geographic Information System.

Factores que influyen en la ocurrencia de accidentes de tránsito en vías urbanas: Un enfoque combinado GIS-Bayesiano empírico

Resumen
La magnitud del problema de la accidental vial urbana en Colombia es notable y por tal razón interesa estudiar técnicamente este importante flagelo de salud pública. Normalmente las técnicas cuantitativas empleadas son muy agregadas y no permiten identificar correctamente las variables determinantes del problema. El presente artículo estudia la relación existente entre la accidentalidad vial urbana y variables de la vía, el entorno, el tránsito y el control. Específicamente son identificados sectores críticos de accidentalidad en la ciudad de Cartagena (Colombia) mediante el método bayesiano empírico basado en SIG. Se encontró un total de 69 tramos críticos de accidentalidad en la ciudad y se evidenció que el efecto marginal sobre la accidentalidad de las motocicletas es muy superior al de autos y buses. También se encontró evidencia empírica que los tramos ubicados en zonas comerciales tienden a presentar mayor frecuencia de accidentes debido a la alta presencia de peatones.

Palabras clave: accidentalidad vial urbana; tramos críticos de accidentalidad; enfoque bayesiano empírico; Sistema de Información Geográfico.

1. Rationale

Knowing the exact number of people who are affected by traffic accidents is virtually impossible because minor injuries usually do not count as part of the official statistics. The same applies to the case of deaths caused by traffic accidents, as some countries are not governed by the international convention of counting those deaths that occurred up to 30 days after the event [1].

However, some estimates from the World Health Organization indicate that approximately 1.2 million deaths and between 20 and 50 million injuries result from traffic accidents on a global scale [2]. In examining the relationship of deaths and injuries per cause, traffic accidents rank first, causing seven times more deaths than wars and two times more than other forms of violence [2].

In Colombia the problem of road accidents is also of a great magnitude. Records indicate that during the period 2002-2012 accidents in the country resulted in approximately 62,000 deaths and over 443,000 injuries. Traffic accidents in the country have become the second leading cause of violent death and the leading cause of death among young Colombians less than 30 years of age [3].

According to official data from the sector, during 2012 there were a total of 45,592 road traffic injuries (13.5% fatal) in Colombia. Historically it can be observed that the number of deaths from traffic accidents has peaked over the last decade and has exceeded the 9.39% average for the previous nine years, which was 5,625 [4].

The impact of road accidents highlights the need to technically study this scourge of global public health problems that negatively impact the national panorama [5]. Sometimes analyses of traffic accidents are performed using quantitative statistical techniques, which place great importance on the evolution of accidents [6]. When the problem is approached in this way, the commonly used indicators such as morbidity or mortality appear to be highly aggregated and for that reason conceal the problem's determinant variables.

Worldwide, several studies have been developed that seek to explain the occurrence of traffic accidents from different perspectives. The analyses have addressed legal and judiciary perspectives, technical attributes of vehicles and infrastructure as well as the psychological, behavioral and socio-economic components of the road system users [7]. Recently, some analysis techniques based on the use of Geographic Information Systems (GIS) have been used, which allow the generation of maps, models and risk estimates from a spatial perspective.

Theoretically it is possible to reduce the number of accidents through specific actions at the highest accident sites; however, it is possible that the complexity of the problem requires a more elaborate analysis that considers the relationship between the variables involved [8].

Studies such as those in references [9-11] have shown that conventional linear regression models are sometimes inadequate to model the frequency of traffic accidents, as they may provide erroneous inferences. As such, recent research on the road accidents has based their methodologies on the use of Poisson and Negative Binomial regression models [12].

Currently, some researchers have conducted spatial analyses of accidents in order to establish relationships between the occurrence of accidents and the factors that contribute to their generation, using Poisson regression models [13,14]. Studies undertaken in the field have developed prediction models for high accident rates or potentially risky areas by using analysis techniques that can be grouped into four general categories: multivariate analysis, fuzzy logic, neural networks and empirical Bayes [15], which is an approach widely used in transport engineering [16].

The positive benefit of using GIS systems in safety analysis is evident. For example, a spatial data validation system based on GIS has been used to check the accuracy of the crash records [17]. This approach has also been used to explore different spatial and temporal visualization technologies to reveal patterns and significant factors relating to vehicle crashes, [18] or as a management system for accident analysis and the determination of hot spots [19]. Along similar lines to the present work, [20] shows that accidents can be related to infrastructure characteristics, while [21] relates accidents to socioeconomic properties such as income level, presence of children and crime rates.

In this context, using modeling techniques, this paper aims to establish the relationship between urban road accidents and the road, environment, traffic and control variables. Specifically, the model is applied to the city of Cartagena, identifying accident-prone locations using the empirical Bayesian method supported by GIS.

2. Theoretical framework

2.1. Statistical analysis procedure

The method of statistical accident analysis can be performed in four stages: building the model, calculation of the posterior distribution, the posterior distribution analysis, and inference and obtaining of final conclusions on the problem being considered.

In Poisson regression models the likelihood that n accidents occur in the i road section during j period is given by the following expression:

$$P\left(n_{ij}\right) = \frac{exp^{-\lambda_{ij}}\lambda_{ij}^{n_{ij}}}{n_{ij}!} \quad (1)$$

In turn, λ_{ij} is the expected value of n_{ij}, which can be expressed in terms of a vector of relevant variables describing the geometry of the road, the environment and other characteristics that affect the frequency of accidents in the sections studied. An exposure variable V_i is added to this set of variables, which relates to the traffic volume of the particular road section or to the vehicle-kilometers travelled. The expected value of the number of accidents can be estimated as follows:

$$E(n_{ij}) = \lambda_{ij} = \exp(\beta X_{ij} + LnV_{ij}^{\gamma} + \delta) \quad (2)$$

One of the main problems that models based on the Poisson regression can have is the over-dispersion of residuals, i.e., the model underestimates the degree of dispersion of the result because the Poisson distribution assumes that the counts variability of a covariant group is equal to the average. Failure to observe this relationship might result in the estimated coefficients being biased and inefficient, so a Negative Binomial distribution based on a Gamma distributed error term should be used. In this case, the expected value for the number of accidents is rewritten:

$$E(n_{ij}) = \lambda_{ij} = \exp(\beta X_{ij} + LnV_{ij}^{\gamma} + \delta + \varepsilon_{ij}) \quad (3)$$

Where $\exp(\varepsilon_{ij})$ is an error term that distributes Gamma and allows the variance of the mean to be differentiated as follows:

$$Var\left[n_{ij}\right] = E\left[n_{ij}\right] + \alpha E\left[n_{ij}^2\right] \quad (4)$$

In this case, α is known as an over-dispersion parameter, noting that when $\alpha = 0$ the Poisson distribution is obtained, for which the Negative Binomial probability is represented by the following mathematical expression:

$$P(n_i) = \frac{\Gamma(n_{ij}+\alpha^{-1})}{\Gamma(\alpha^{-1})n_{ij}!}\left[\frac{\alpha^{-1}}{\alpha^{-1}+\lambda_{ij}}\right]^{\alpha^{-1}}\left[\frac{\lambda_{ij}}{\alpha^{-1}+\lambda_{ij}}\right]^{n_{ij}} \quad (5)$$

Where $\lambda_{ij} > 0$ and $\alpha \geq 0$. This widely used model can be estimated by maximum likelihood [22-25].

2.2. *Identification of critical areas using the empirical Bayes method*

The empirical Bayesian approach for the analysis of road safety has gradually developed over the last thirty years and today it is a widely recommended method for the estimation of traffic accidents [26,27].

The empirical Bayesian method was originally developed to control the effect called "regression to the mean" in road safety studies "before and after". It is now also used to identify accident-prone locations in road safety analysis [28-32]. This method assumes that λ varies among different analysed sections and that the exact value for any of these is an unknown variable with a Gamma probability function density, expressed as follows:

$$f(\lambda) = \frac{\alpha^{\beta}\lambda^{\beta-1}e^{-\alpha\lambda}}{\Gamma(\beta)} \quad (6)$$

In these equations α is the shape parameter and β is the scale parameter of the Gamma function, with mean and variance given by eq. (7) and (8):

$$E(\lambda) = \frac{\alpha}{\beta} \quad (7)$$

$$Var(\lambda) = \frac{\alpha}{\beta^2} \quad (8)$$

The function parameters are determined from the data available about accidents in the reference group, and are expressed as follows:

$$\hat{\alpha} = \frac{\bar{n}}{\sigma^2 - \bar{n}} \quad (9)$$

$$\hat{\beta} = \frac{\bar{n}^2}{\sigma^2 - \bar{n}} \quad (10)$$

Where n is the number of accidents in a given section. The parameters of the posterior distribution, which is also of Gamma type, will be given by:

$$\alpha' = 1 + \hat{\alpha} \quad (11)$$

$$\beta' = n + \hat{\beta} \quad (12)$$

Where the expected value and the variance will be given by:

$$E(\lambda\backslash n) = \frac{\beta'}{\alpha'} \quad (13)$$

$$Var(\lambda\backslash n) = \frac{E(\lambda\backslash n)^2}{\beta'} \quad (14)$$

To identify the potentially critical sites, the probability that E(λ|n) exceeds E(λ) is:

$$P(\lambda|n > \lambda) = 1 - \int_0^{E(\lambda)} \frac{\alpha^{\beta}\,(\lambda\,|n)^{\beta-1}e^{-\alpha(\lambda\,|n)}}{\Gamma(\beta)} d(\lambda\,|n) \quad (15)$$

Where:

E(λ): Expected number of accidents for the reference group during the study period.

E(λ|n): Expected number of accidents adjusted by the empirical Bayesian method for a section of the reference group in the same study period.

Then, if the previously calculated probability exceeds threshold δ, which defines the confidence level chosen, the section is considered critical.

The determination of the parameters was done through maximum likelihood and, the statistical analysis was done using tests such as the t-student test.

3. Settings

The study was conducted in Cartagena, Colombia, which has a population of 923,219 inhabitants in an area of 616 km^2. The city has an urban road network of 656 km and its main economic activities are: tourism, petrochemicals and port industries.

The data used in this study corresponds to the information on accidents that was gathered in Cartagena during 2007 by the Administrative Department of Traffic and Transportation. The accidents classified as "only property damage" were not included in the study, so the analysis was focused on accidents involving deaths and injuries. Although we had more up-to-date information, we preferred to work with the data from 2007 considering that the other available data was affected by the construction of Cartagena's BRT system.

Local statistics from recent years show a significant increase in the number of fatalities and injuries and, a close relationship with the volumes of motorcycles circulating in the city, as shown in Fig. 1, 2.

It is possible that the phenomenon known as "mototaxismo" has a correlation with the number of road accidents in Cartagena. In this regard, the report on the status of road safety in the Americas region established that over the last ten years motorcycle-related deaths have dramatically increased [33].

One of the details that intrigued this study was the fact that Colombia is ranked as the country with the highest mortality rate for motorcyclists in Latin America, with 3.6 deaths per 100,000, followed by Brazil with 2.9, Paraguay 2.5 and Suriname 2.2 [33]. Fig. 1, 2 confirm that the percentage of accidents is high with some degree of severity in Cartagena when there is at least one motorcycle involved.

Fig. 3 shows that about 43% of those killed in traffic accidents correspond to motorcyclists, followed by pedestrians with 29%. This is consistent with what is shown in Fig. 4, which indicates that the type of accident with the most severe consequences is when a motorcycle collides with another vehicle.

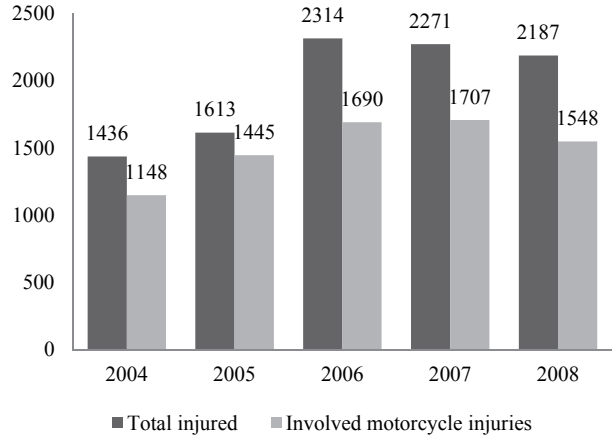


Figure 1. Motorcycle injuries and total injuries
Source: The authors.

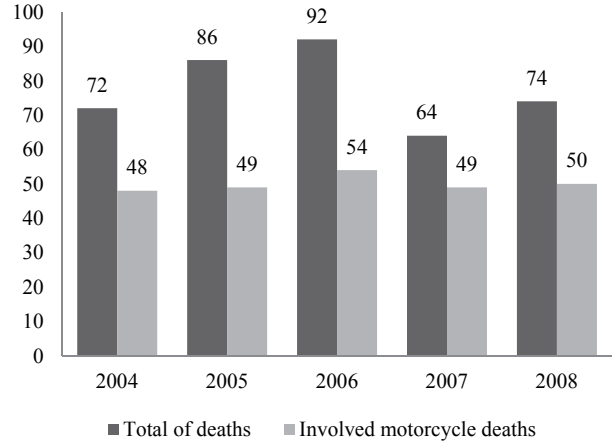


Figure 2. Motorcycle deaths and total deaths
Source: The authors.

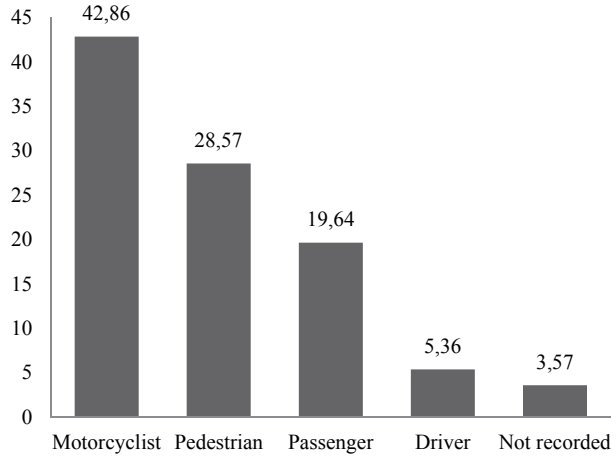


Figure 3. Percentage of deaths in traffic accidents for each condition
Source: The authors.

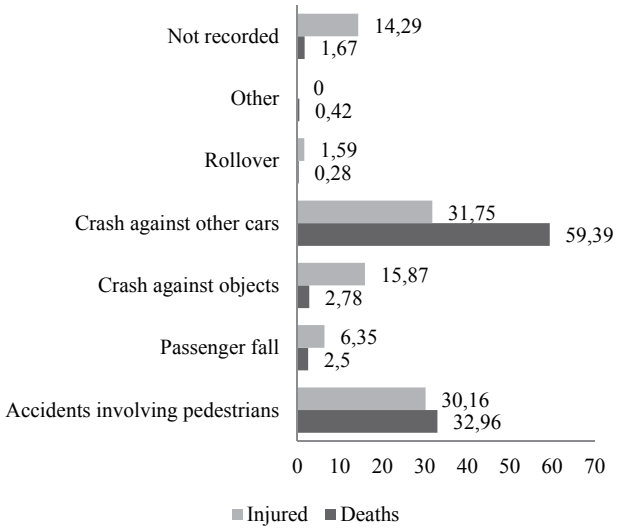


Figure 4. Percentage of injured and killed in traffic accidents by type of accident
Source: The authors.

Taking into consideration these behavioral patterns, we decided to include vehicular composition as one of the variables to be studied, in order to compare not only the effect of the motorcycle accident frequency, but also that of buses and cars. This is in response to referents that specifically demonstrate the impact of buses on accidents, both directly and indirectly [34-39]

4. Methodology

Exploratory and quantitative research was conducted, involving the processing of statistical data related to road accidents in the city of Cartagena in 2007.

The accident analysis considered the two usual classifications: accidents by severity and type. In terms of severity, accidents with injuries and deaths were considered, and only accidents with damage were excluded from the analysis. The second classification took into account if the accident happened as a result of a crash, if it was an accident involving pedestrians, a rollover, if an occupant fell, or any other type.

Each accident was classified by severity and type, and other aspects were analyzed such as its location on the road network of the city as well as the type and model of the vehicles involved. With respect to the individual attributes, age and gender were also considered, and the victim's status was specified as a: driver, passenger, driver's passenger or pedestrian. Additionally, some road geometric characteristics, road conditions and traffic flows were observed.

The methods used in the investigation primarily considered the Transport Research Laboratory's methodology (TRL), used in Great Britain, Sweden and other members of the European Road Assessment Programme (EuroRAP) and, secondarily, the Institute of Highways and Transportation's methodology (IHT) [40].

The TRL methodology does not take into account the possible causes of accidents, and the random component thereof; the statistical analysis simply determines the level of risk through the assessment of accidents and their severity. A risk index is calculated, which is defined as the number of fatal and serious accidents for 1,000 million vehicles per kilometer. Conversely, the IHT methodology proposes a clear and detailed procedure for the identification, diagnosis and selection of the sites to be studied. In this paper, combining both methods, the steps described below were developed.

4.1. Collection and processing of relevant information

Initially, we used the information contained in the accident reports of the Administrative Department of Traffic and Transport of Cartagena in 2007. The information was organized and digitized so that site of the event, date and time, age of people involved, type of vehicles involved and other important attributes could be identified.

In Colombia, as in other countries, the use of accident data from police reports has certain disadvantages [41, 42], also not all traffic accidents with victims are reported by the police [43]. For this reason, we compared the available information provided by the Road Prevention Fund (FPV) for the same period. The results from this verification were consolidated into a database of 1,367 traffic accidents with injuries or deaths.

Complementary, topographic surveys were conducted to define the geometric characteristic of each case study, such as road widths and turning radii. Additionally, vehicle volumes and vehicle composition studies were also applied, and were supplemented with information from previous studies conducted by the city agents. We also had cartographic information from various sources, which was used to quantitatively identify possible factors involved in the occurrence of accidents.

4.2. Spatial analysis

From the spatial distribution of accidents, a database to classify the road network in areas with similar characteristics in terms of traffic, land use, occupation of public space, and geometry was built. This classification allowed 241 sections of the road network to be identified and, for each of them, the number of accidents was recorded.

The spatial analysis also accessed maps by using kernel density tools and calculated magnitudes per area unit based on the entities of interest, in this case the location of accidents. This type of analysis has been used to analyze accidents in various urban settings, located in different countries such as Canada [44], India [45] and Turkey [19].

4.3. Generation of statistical models

The identification of explanatory variables was performed by Principal Component Analysis (PCA), which is widely used in the analysis of data that are not dimensionally homogeneous or when the order of magnitude of the random variables measured is not the same as with the data studied.

Figure 5. Spatial analysis of the concentration of accidents involving injuries using kernel type tools
Source: The authors.

The PCA identified the following variables that explain the frequency of accident occurrence: land use (1: Commercial, 0: otherwise), number of intersections, road width, section length, traffic directions (1: Two-way circulation, 0: otherwise) and average daily traffic.

Poisson and Negative Binomial regression models were estimated to describe the variation of the observed frequency of accidents on the section i during period j. Then, using an empirical Bayesian analysis, the identification of accident-prone locations was undertaken.

5. Results

Accident spatial analysis based on GIS allowed, besides the identification of homogeneous sections, the existing relationship between infrastructure, land use and accidents to be established, in order to consolidate the database that was to be used in the model generation phase. Fig. 5, which contains the spatial analysis of the concentration of accidents resulting in injuries using kernel type tools, clearly illustrates the relationship between road infrastructure and accidents. Other GIS data layers, such as layer urban land use for the city of Cartagena, was used for the spatial analysis prior to model calibration.

The statistical summary of the considered explanatory variables is shown in Table 1, where $AADT_B$, $AADT_M$ and $AADT_A$, represent the average daily transit for buses, motorcycles and cars, respectively.

Table 1.
Statistical summary of independent variables.

Variable	Notation	Mean	Mode	Deviation
Length of section (m)	L	533.39	500	370.40
No. of intersections		5	1	4.54
$Ln\ (AADT \cdot L)$	X_1	7.64	9.24	1.98
Road width	X_2	7.82	7.30	2.00
$AADT_B \cdot L/1000$	X_3	1.04	2.69	1,65
$AADT_M \cdot L/1000$	X_4	2.53	3.31	4,15
$AADT_A \cdot L/1000$	X_5	3.62	4.35	5.09
No. of intersections/L	X_6	0.01	0.01	0.0068

Source: The authors.

Although the model calibration was undertaken with and without constant, a better fit in those cases where the constant was specified was always found. This behavior was introduced in the Poisson model as well as in the negative binomial type models.

The results are presented in Table 2. This contains estimates and the statistical t-student that are in parentheses, which allows significance to be measured. It may be noted that the over-dispersion parameter is small and statistically insignificant, so the estimates of the obtained parameters for the two proposed models do not differ substantially.

All parameters have the expected signs and the important effect of the AADTM and AADTB is highlighted. This is much smaller than the effect of cars. It can be seen that the marginal effect of a motorcycle on the frequency of accidents with deaths and injuries is 2.5 times greater than the effect of an additional bus and 6 times larger than that of an automobile.

Moreover, *ceteris paribus* roads with two-way traffic have higher accident frequency. The same analysis is true for the sections located in areas with commercial land use, characterized by the strong presence of pedestrians. Intersection density, i.e. the number of intersections divided by the length of the section, is also a factor that tends to increase the occurrence of accidents, but the statistical significance of this variable is not high.

Table 2.
Estimated models.

Parameter	Notation	Poisson model	Negative Binomial
Constant		-0.1481 (-0.66)	-0.1481 (-0.69)
$Ln\ (TPDA \cdot L)$	X_1	0.1813 (7.74)	0.1813 (8.08)
Road wide	X_2	-0.0240 (-1.30)	-0.0240 (-1.42)
$AADT_B \cdot L/1000$	X_3	0.0817 (3.37)	0.0817 (3.83)
$AADT_M \cdot L/1000$	X_4	0.0317 (3.87)	0.0317 (4.74)
$AADT_A \cdot L/1000$	X_5	0.0117 (1.36)	0.0117 (1.66)
No. of intersections/L	X_6	6.8561 (1.45)	6.856 (1.55)
Two-way	X_7	0.1376 (1.96)	0.1375 (2.11)
Commercial use	X_8	0.1046 (1.38)	0.1046 (1.50)
Over-dispersion	α		0.0250 (1.61)
R^2		69.8	81.9

Source: The authors.

Table 3.
Classification of critical sections identified.

Type of road	Total of sections	Identified critical sections	Percentage (%)
Arterial	71	48	67.6
Collector	52	19	36.5
Local	118	2	1.7
Total	241	69	28.6

Source: The authors.

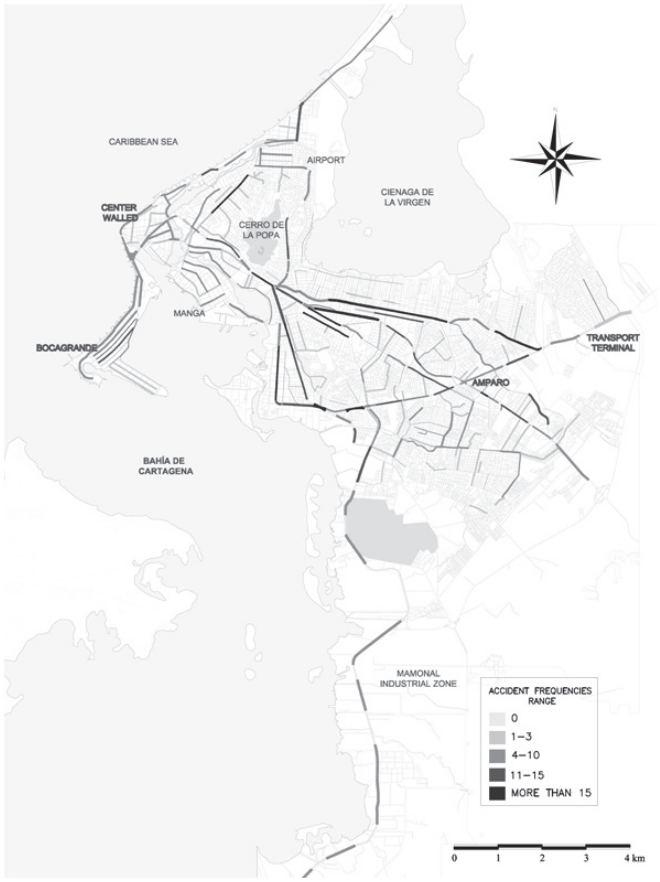


Figure 6. Density of accidents in Cartagena
Source: The authors.

Following the methodology proposed to implement the empirical Bayesian approach, the following parameters were calculated for the analyzed group of sections: S^2=31.23, $\hat{\alpha}$=0.441 and $\hat{\beta}$=2.424. As shown in Table 3 there were a total of69 critical sections, corresponding to 28.6% of sections analyzed, identified.

It is important to note that the largest number of accident-prone locations was located on arterials, on which 67.6% of the sections were considered critical. We also found the same to be true on the collector roads to a lesser degree. Local roads with just 1.7% of accident-prone locations were also identified. Fig. 6 illustrates this by showing the density of accidents resulting in injuries or deaths for the city of Cartagena, which identifies the accident-prone locations studied.

6. Conclusions

In an urban context, such as in the city of Cartagena, we have found empirical evidence demonstrating the relationship between the frequency of accidents and variables as well as traffic density, road width, density of intersections per segment, flow direction and land use environment.

The results found are representative of cities located in less developed countries, with low amount of traffic, which have a high occupancy of public space and important use of public transport and motorcycles in the cities. The latter are even sometimes used as an informal mode of public transport. This analysis enabled us to, from a systemic approach; identify the most potentially accident-prone areas with the help of the empirical Bayesian method.

The study was able to successfully involve spatial analysis of accidents through the use of a GIS to identify homogeneous sections in Cartagena, that were then established by modeling the impact of vehicular volumes of motorcycles, buses and cars on the expected number of traffic accidents for each sector. This showed that the marginal effect on the accident rate for motorcycles is significantly higher than for that of buses and, it is much higher than for that of cars.

Regarding road infrastructure variables, we found that there are more accidents in the sections with two-way traffic than in the single ones. The level of risk also decreases with the width of the road. This is a very important discovery to be taken into consideration when embarking on plans for the operational management of the road network in the city, as well as well as when the urban road sections for planning are defined. Likewise, the sections located in commercial areas, which normally have high presence of pedestrians, tend to have a higher frequency of the occurrence of accidents with serious consequences. It is in these places where authorities should maintain a greater presence to mitigate problems resulting from accidents.

By using the empirical Bayesian approach it was possible to statistically identify critical areas for accidents, which will allow local authorities to focus the efforts of mitigating the problems caused by the accidents. The method employed showed the importance of the spatial location of the accident and the formation of reference groups for data analysis, which were initially assessed on the basis of their similar characteristics, regardless of the frequency or severity of the accidents.

Among a total of 241 sections established, it was found that nearly 30% are identified as high-risk accident sections. This percentage is consistent with the findings of [15] who found that most accidents occurred on curves, which represent 29.7% of the total length of the studied motorway. These findings can also be found in [7] who reported that in the case of regional roads, 26% of basic spatial units (hectometer of road) have at least 1 accident.

The estimation of predictive Poisson type and negative binomial models for the frequency of accidents resulting in either injuries or deaths allows causal relationships to be established between the frequency of accidents and the explanatory variables. In this sense, this research provides some initial insights into directions for actions that the Administrative Department of Traffic and Transportation could take in order to reduce the likelihood of high impact accidents in urban environments.

First, measures should be focused on arterials as they are the areas with the highest number of critical sections, especially those located in areas with commercial land use that are characterized by a heavy pedestrian presence.

Future research can use similar methods to those presented in this paper for "before and after" studies that continue the application of the empirical Bayesian approach in response to the arrival of Cartagena's BRT system.

Acknowledgements

We would like to thank the Administrative Department of Science, Technology and Innovation (COLCIENCIAS) for having partially financed this work through contracts RC No 0791-2013 (Cod. 121562238496) and 470-2014 (Cod. 1215-675-47211).

References

[1] Johansson, R., Vision Zero – Implementing a policy for traffic safety. Safety Science, 47, pp. 826-831, 2009. DOI: 10.1016/j.ssci.2008.10.023

[2] Organización Mundial de la Salud (OMS)., Informe sobre la situación mundial de la seguridad vial 2013. [Online]. Washigton: OMS, Bloomberg Philanthropies. [date of reference July 25th of 2014]. Available at: http://apps.who.int/iris/bitstream/10665/83798/1/WHO_NMH_VIP_13.01_spa.pdf

[3] Contraloría General de la Nación., Seguimiento a los resultados de la política pública de Seguridad Vial en Colombia. Bogotá: Contraloría General de la República, [Online]. 2012. [date of reference July 25th of 2014]. Available at: http://www.contraloria.gov.co/documents/155638087/171673939/Informe_Seguridad_Vial_15.pdf/6b90bdad-2fec-4c20-8a81-6a65a8a5a2ac?version=1.1

[4] Moreno, S.L., Muertes y lesiones por accidentes de transporte, Colombia, 2012. Grupo Centro de Referencia Nacional sobre Violencia, Instituto Nacional de Medicina Legal y Ciencias Forenses, Bogotá, [Online]. 2012. [date of reference July 25th of 2014]. Available at: http://www.medicinalegal.gov.co/documents/10180/34861/7+transporte+forensis+2012.pdf/2fbb18b6-2ae7-4f58-8c25-220cb4d3be37

[5] Ministerio de Transporte. Plan Nacional de Seguridad Vial Colombia 2013 – 2021. PNSV 2013 – 2021, Bogotá, [Online]. 2014. [date of reference July 25th of 2014]. Available at: https://www.mintransporte.gov.co/descargar.php?idFile=11576

[6] Observatorio de movilidad urbana de Bogotá. Líneas de investigación del observatorio de movilidad enmarcado en la seguridad vial. Análisis 1994-2000, 2001.

[7] Flahaut, B., Impact of infrastructure and local environment on road unsafety: Logistic modeling with spatial autocorrelation. Accident Analysis & Prevention, 36, pp. 1055-1066, 2004. DOI: 10.1016/j.aap.2003.12.003

[8] Muhlrad, N. and Lassarre, S., Systems approach to injury control. En: Tiwari, G., Mohan, D. and Muhlrad, N., The way forward: Transportation planning and road safety. Nueva Delhi, McMillan India Ltd., pp. 52-73, 2005.

[9] Jovanis, P. and Chang, H.L., Modeling the relationship of accidents to miles traveled. Transportation Research Record 1068, pp. 42-51, 1986.

[10] Joshua, S. and Garber, N., Estimating truck accident rate and involvements using linear and poisson regression models. Transportation Planning and Technology, 15, pp. 41-58, 1990. DOI: 10.1080/03081069008717439

[11] Miaou, S.P. and Lum, H., Modeling vehicle accidents and highway geometric design relationships. Accident Analysis & Prevention, 25(6), pp. 689-709, 1993. DOI: 10.1016/0001-4575(93)90034-T

[12] Lee, J. and Mannering, F., Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis. Accident Analysis & Prevention, 34(1), pp. 149-161, 2002. DOI: 10.1016/S0001-4575(01)00009-4

[13] Lord, D. and Miranda-Moreno, L., Effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter of Poisson-gamma models for modeling motor vehicle crashes: A Bayesian perspective. Safety Science, 46(5), pp. 751-770. 2008. DOI: 10.1016/j.ssci.2007.03.005

[14] Agüero-Valverde, J. and Jovanis, P., Spatial analysis of fatal and injury crashes in Pennsylvania. Accident Analysis & Prevention, 38(3), pp. 618-625, 2008. DOI: 10.1016/j.aap.2005.12.006

[15] Caliendo, C., Guida, M. and Parisi, A., A crash-prediction model for multilane roads. Accident Analysis & Prevention, 39, pp. 657-670, 2007. DOI: 10.1016/j.aap.2006.10.012

[16] Gutierrez, M. and Cantillo, V., Classic and bayesian estimation of Subjective Value of Time. DYNA 81(187), pp. 158-166, 2014. DOI: 10.15446/dyna.v81n187.40863

[17] Loo, B.P.Y., Validating crash locations for quantitative spatial analysis: A GIS-based approach, Accident Analysis & Prevention, 38(5), pp. 879-886, 2006. DOI: 10.1016/j.aap.2006.02.012

[18] Plug, C., Xia, J. and Caulfield, C., Spatial and temporal visualisation techniques for crash analysis. Accident Analysis & Prevention, 43(6), pp. 1937-1946, 2011. DOI: 10.1016/j.aap.2011.05.007

[19] Erdogan, S., Yilmaz, I., Baybura, T. and Gullu, M., Geographical information systems aided traffic accident analysis system case study: City of Afyonkarahisar. Accident Analysis & Prevention, 40(1), pp. 174-181, 2008. DOI: 10.1016/j.aap.2007.05.004

[20] Vieira, S., The influence of the infrastructure characteristics in urban road accidents occurrence. Accident Analysis & Prevention, 60, pp. 289-297, 2013. DOI: 10.1016/j.aap.2013.02.042

[21] Cottrill, C.D. and Thakuriah, P., Evaluating pedestrian crashes in areas with high low-income or minority populations. Accident Analysis & Prevention, 42(6), pp. 1718-1728, 2010. DOI: 10.1016/j.aap.2010.04.012

[22] Miaou, S.P., The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. Accident Analysis & Prevention, 26(4), pp. 471-482, 1994. DOI: 10.1016/0001-4575(94)90038-8

[23] Shankar, V.N., Mannering, F. and Barfield, W., Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. Accident Analysis & Prevention, 27(3), pp. 371-389, 1995. DOI: 10.1016/0001-4575(94)00078-Z

[24] Shankar, V.N., Mannering, F. and Barfield, W., Statistical analysis of accident severity on rural freeways. Accident Analysis & Prevention, 28(3), pp. 391-401, 1996. DOI: 10.1016/0001-4575(96)00009-7

[25] Carson, J. and Mannering, F., The effect of ice warning signs on ice accident frequency and severity. Accident Analysis & Prevention, 33(1), pp. 99-109, 2001. http://dx.doi.org/10.1016/S0001-4575(00)00020-8

[26] Hauer, E., Harwood, D.W., Council, F.M. and Griffith, M.S., Estimating safety by the empirical Bayes method. A tutorial. Transportation Research Record, 1784, pp. 126-131, 2002. DOI: 10.3141/1784-16

[27] Guerrero-Barbosa, T. and Amarís-Castro, G., Application of Bayesian techniques for the identification of accident-prone road sections. DYNA 81(187), pp. 209-214, 2014. DOI: 10.15446/dyna.v81n187.41333

[28] Persaud, B., Lyon, C. and Nguyen, T., Empirical Bayes procedure for ranking sites for safety investigation by potential for improvement. Transportation Research Record, 1665, pp. 7-12, 1999.

[29] Harwood, D.W., Torbic, D.J., Bauer, K.M., Persaud, B.N., Lyon, C.A. and Hauer, E., Safety analyst: Software tools for safety management of specific highway sites. Federal Highway Administration, Washington, D.C., 2002.

[30] Lord, D., Modeling motor vehicle crashes using Poisson-gamma models: Examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter. Accident Analysis & Prevention, 38, pp. 751-766, 2006. DOI: 10.1016/j.aap.2006.02.001

[31] Miaou, S.P. and Lord, D., Modeling traffic crash-flow relationships for intersections: Dispersion parameter, functional form, and Bayes versus Empirical Bayes. Transportation Research Record, 1840, pp. 31-40, 2003.

[32] Cheng,W. and Washington, S.P., Experimental evaluation of hotspot identification methods. Accident Analysis & Prevention, 37, pp. 870-881, 2005. DOI: 10.1016/j.aap.2005.04.015

[33] Organización Panamericana de la Salud., Informe sobre el estado de la seguridad vial en la región de las Américas. Washington, D.C. [Online]. 2009. [date of reference July 25th of 2014]. Available at: http://www.who.int/violence_injury_prevention/road_safety_status/2009/gsrrs_paho.pdf

[34] Cameron, I.C., Harris, N.J. and Kehoe, N.J.S., Tram-related injury in Sheffield. Injury International Journal of the Care of the Injured, 32, pp. 275-277, 2001.

[35] Hedelin, A., Bunketorp, O. and Björnstig, U., Public transport in metropolitan areas – A danger for unprotected road users. Safety Science, 40, pp. 467-477, 2002. DOI: 10.1016/S0925-7535(01)00014-5

[36] Hamed, M.M., Jaradat, A.S. and Easa, S.M., Analysis of commercial mini-bus accidents. Accident Analysis & Prevention, 30, pp. 555-567, 1998. DOI: 10.1016/S0001-4575(98)00021-9

[37] Vayre, P., The risk of accident for a parisian pedestrian. Comptes-rendus de l'académie des sciences, series III, Sciences de la Vie, [Online]. 324, pp. 1175-1179, 2001. Available at: http://tra2014.traconference.eu/papers/pdfs/TRA2014_Fpaper_17521.pdf

[38] af Wåhlberg, A.E., Characteristics of low speed accidents with buses in public transport: Part II. Accident Analysis & Prevention, 36, pp. 63-7, 2004. DOI: 10.1016/S0001-4575(02)00128-8

[39] Brenac, T. and Clabaux, N., The indirect involvement of buses in traffic accident processes. Safety Science, 43, pp. 835-843, 2005. DOI: 10.1016/j.ssci.2005.04.003

[40] IHT Institution of highways and transportation., Guidelines for Urban Safety Management, London, 1990.

[41] Alsop, J. and Langley J., Under-reporting of motor vehicle traffic crash victims in New Zealand. Accident Analysis & Prevention, 33, pp. 353-359, 2001. DOI: 10.1016/S0001-4575(00)00049-X

[42] Austin, K., The identification of mistakes in road accident records: Part 2, casualty variables. Accident Analysis & Prevention, 27, pp. 277-282, 1995. DOI: 10.1016/0001-4575(94)00066-U

[43] Laumon, B. and Martin, J.L., Analyse des biais dans la connaissance épidémiologique des accidents de la route en France. Revue d'Epidémiologie et de Santé Publique, 50(3), pp. 277-285, 2002.

[44] Young, J. and Park, P.Y., Hotzone identification with GIS-based post-network screening analysis. Journal of Transport Geography, 34, pp. 106-120, 2014. DOI: 10.1016/j.jtrangeo.2013.11.007

[45] Prasannakumar, V., Vijith, H., Charutha, R. and Geetha, N., Spatio-temporal clustering of road accidents: GIS Based analysis and assessment. Procedia Social and Behavorial Sciences, 21, pp. 317-325, 2011. DOI: 10.1016/j.sbspro.2011.07.020

V. Cantillo, received his BSc. Eng. in Civil Engineering in 1987, from the Universidad del Norte, Colombia; his MSc. in Traffic and Transport Engineering in 1990, from the Universidad del Cauca, Colombia and his PhD. in Engineering Sciences (Transport), in 2004, from the Pontificia Universidad Católica de Chile, Chile. He has been a professor at the Universidad del Norte, Barranquilla, Colombia since 1990. He is the head of the Transport Research Group - TRANVIA. His research interests include: transport planning, transport economics, transport modeling and logistics. He has also been working in other topics such as education in engineering, econometrics and construction materials.
ORCID: 0000-0003-1184-2580

P. Garcés, received her BSc. Eng. in Civil Engineering in 1997, from the Universidad de Cartagena, Colombia and her MSc. in Civil Engineering in 2010, from the Universidad del Norte, Barranquilla, Colombia. She has been a professor at the Universidad de Cartagena, Cartagena, Colombia, since 2004.
ORCID: 0000-0001-7561-8049

L. Márquez, received his BSc. Eng. in Transportation and Highways Engineering in 1993 and his MSc. in Transportation Engineering, in 2009, both from Universidad Pedagógica y Tecnológica de Colombia, Tunja, Colombia. Currently he is undertaking a PhD in Civil Engineering Sciences, from the Universidad del Norte, Barranqilla, Colombia. He has been an associate professor at the Universidad Pedagógica y Tecnológica de Colombia, Tunja, Colombia since 2004. His research interests include: transport planning, transport economics, transport modeling and logistics.
ORCID: 0000-0002-5052-1694

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS

DYNA

DYNA
http://dyna.medellin.unal.edu.co/

Crack growth in pyrographite under the conditions of radiation

Pedro Alejandro Tamayo-Meza [a], Viacheslav Alexandrovich Yermishkin [b], Usiel Sandino Silva-Rivera [c], Alejandro Leyva-Díaz [d], Josué Osmar Trejo-Escandón [e], Juan Manuel Sandoval-Pineda [f], & Luis Flores-Herrera [e]

[a] *Escuela Superior de Ingeniería Mecánica y Eléctrica, Instituto Politécnico Nacional, Ciudad de México, México. ptamayom@ipn.mx*
[b] *Institute of Metallurgy, Russian Academy of Science, Moscow, Russian Federation.*
[c] *Escuela Superior de Ingeniería Mecánica y Eléctrica, Instituto Politécnico Nacional, Ciudad de México, México. usiel31@yahoo.com.mx*
[d] *Escuela Superior de Ingeniería Mecánica y Eléctrica, Instituto Politécnico Nacional, Ciudad de México, México. a.leyvadíaz@gmail.com*
[e] *Escuela Superior de Ingeniería Mecánica y Eléctrica, Instituto Politécnico Nacional, Ciudad de México, México. jtescandon@hotmail.com*
[f] *Escuela Superior de Ingeniería Mecánica y Eléctrica, Instituto Politécnico Nacional, Ciudad de México, México. jsandovalp@ipn.mx*
[g] *Escuela Superior de Ingeniería Mecánica y Eléctrica, Instituto Politécnico Nacional, Ciudad de México, México. lafloresh@ipn.mx*

Received: November 27th, 2014. Received in revised form: April 12th, 2015. Accepted: December 10th, 2015.

Abstract
The damage induced by radiation in the pyrographite is accompanied by significant plastic deformation. Microcracks arise in the edges of the radiated areas that develop radially in the direction of the unaffected matrix. The peculiarities in the formation of the tension state of the radiated area and adjacent unaffected areas are analyzed in order to explain the reasons behind the growth of cracks. The analysis is carried out on graphite disks with constant thicknesses that are exposed to radiation with high-energy electrons in an HVTEM-JEOL 1000. A differential equation for specific load conditions is obtained from the analysis of equilibrium conditions of a disc-shaped element.

Keywords: Pyrographite, crack, irradiation, HVTEM, stress state.

Crecimiento de grietas en pirografito bajo condiciones de radiación

Resumen
El daño inducido por radiación en el pirografito va acompañado por una significativa deformación plástica. En los bordes de la zona radiada surgen microgrietas que se desarrollan de manera radial en dirección de la matriz no afectada. Se analizan las particularidades en la formación del estado de tensión de la zona radiada, y en los campos adyacentes no afectados por la radiación para explicar las razones que originan el crecimiento de grietas. El análisis se lleva a cabo sobre discos de grafito con espesor constante, expuestos a radiación con electrones de altas energías dentro de un HVTEM-JEOL 1000. Del análisis de las condiciones de equilibrio de un elemento en forma de disco, se obtiene una ecuación diferencial para condiciones específicas de carga.

Palabras clave: Pirografito, grieta, irradiación, HVTEM, estado de esfuerzos.

1. Introduction

Exposure of pyrographite to radiation with high-energy electrons causes a type K → A phase transition, accompanied by considerable plastic deformation. It was established that microcracks emerge at the edges of the irradiated zone, these microcracks then grow unceasingly in the radial direction within the irradiated matrix, Fig. 1.

To clarify the possible causes of the incubation and growth of cracks, it is necessary to examine the peculiarities in the formation of a stress state in the irradiated zone as well as in the conjugated matrix zone that is not affected by the radiation. This cannot be caused by the change of the temperature field in the irradiated zone, since the heating magnitude in the HVTEM column by means of the electron beam does not exceed 40ºC [1,2]. Besides, prolonged radiation experiments undoubtedly favor the balance of the temperature field gradients in both the sample thickness and the radial direction within the irradiated zone limits.

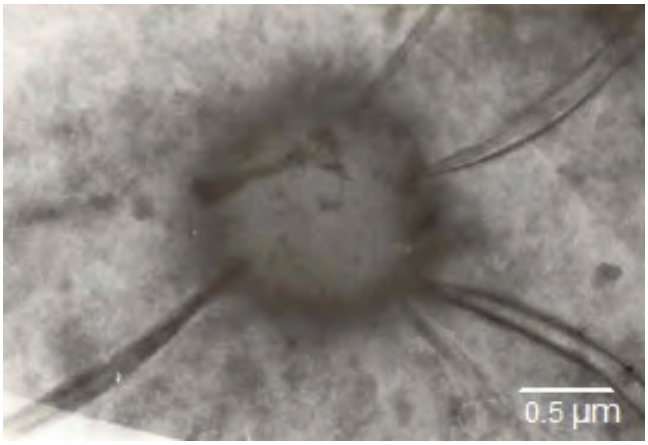


Figure 1. Cracks in pyrographite emerged as a result of irradiation with energy electrons: HVTEM-JEOL 1000.
Source: The authors.

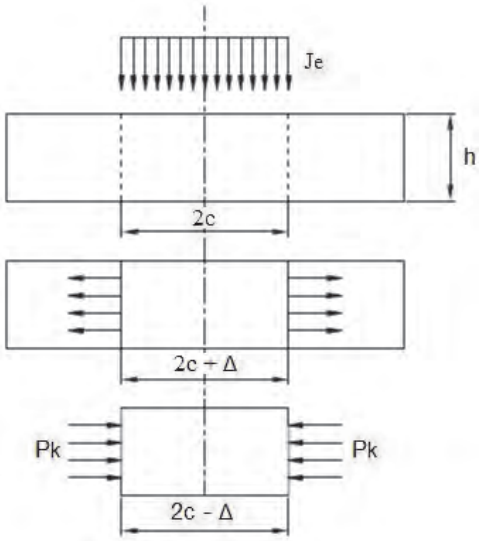


Figure 2. Radiation scheme pyrographite samples: Experiments conducted within the HVTEM, JEOL 1000.
Source: The authors.

2. Theoretical approach

The radiation scheme of pyrographite samples in the form of flat disks performed within the HVTEM is shown in Fig. 2.

As a result of a prolonged high energy electron bombing, the amorphization is developed in the pirografites, which leads to a decrease in the material specific volume in the irradiated zone [3]. Considering that the irradiated zone is surrounded by a matrix that is not affected by radiation, a forced interaction consequently arises between it and the loosened of the irradiated zone's material [4].

From the equilibrium conditions of a circular sheet element of constant thickness analysis, it is possible to obtain a differential equation for a determined loading scheme. Following the plate theory methods [5,6], we obtain the equilibrium equation for the plate and solve it for the initial and boundary conditions that are derived from the supposed scheme of interaction between an irradiated zone wrapped by a sample matrix. The differential equation is obtained as follows:

$$\sigma_r + \frac{d\sigma_r}{dr} - \sigma_\tau = 0$$

or

$$\frac{\alpha(\sigma_r \cdot r)}{dr} - \sigma_\tau = 0 \quad (1)$$

where σ_r is the radial stress, and σ_τ is the tangential stress.

The link between the components of the stress tensor σ_r, σ_τ, σ_z and the deformations ε_r, ε_τ can be expressed in terms of the Generalized Hook's Law:

$$\varepsilon_r = \frac{1}{E}[\sigma_r - \mu(\sigma_\tau + \sigma_z)]$$
$$\varepsilon_\tau = \frac{1}{E}[\sigma_\tau - \mu(\sigma_r + \sigma_z)] \quad (2)$$

where E is the Young modulus, μ is the Poisson modulus.

Supposing that in the conditions of our problem $\sigma_z = 0$, it is possible to obtain the expressions for σ_r and σ_τ from the eq. (2) as follows:

$$\sigma_r = \frac{E}{1-\mu^2} \cdot (\varepsilon_r + \mu\varepsilon_\tau)$$
$$\sigma_\tau = \frac{E}{1-\mu^2} \cdot (\varepsilon_\tau + \mu\varepsilon_r) \quad (3)$$

It is not difficult to prove that

$$\varepsilon_r = \frac{du}{dr}$$

but

$$\varepsilon_\tau = \frac{u}{r} \quad (4)$$

where ε_r is the radial deformation, ε_τ is the tangential deformation, u is the radial displacement.

After substituting eq. (4) in the eq. (3) we obtain:

$$\sigma_r = \frac{E}{1-\mu^2} \cdot \left(\frac{du}{dr} + \mu\frac{u}{r}\right)$$
$$\sigma_\tau = \frac{E}{1-\mu^2} \cdot \left(\frac{u}{r} + \mu\frac{du}{dr}\right) \quad (5)$$

After substituting eq. (5) in the equilibrium differential equation (1), the latter can be expressed as:

$$\frac{d^2u}{dr^2} + \frac{1}{r}\frac{du}{dr} - \frac{u}{r^2} = 0$$

or

$$\frac{d}{dr}\left[\frac{du}{dr} + \frac{u}{r}\right] = 0 \quad (6)$$

Finally, it can be presented in the following form:

$$\frac{d}{dr}\left[\frac{1}{r}\cdot\frac{d(ur)}{dr}\right]=0 \tag{7}$$

Observations of the zone irradiated by HVTEM show an absolute absence of extinction contours, even in the very moment when microcracks start to form. This circumstance reveals the absence of bending during the interaction of forces in the irradiated zone and the surrounding matrix, i.e., the latter has compression characteristics at the edges of the zone. The integration of eq. (6) provides a general solution for radial displacements in the zone affected by radiation:

$$u=c_1 r+\frac{c_2}{r} \tag{8}$$

where c_1 and c_2 are integration constants, the values of which come from the contour conditions.

In our problem, these can be expressed as $r = c$; $\sigma_r = -p_k$; $r=\infty$; $\sigma_r = 0$ for the matrix outside the zone of radiation, and $r = c$; $\sigma_r = p_k$; $r = 0$; $u = 0$ for the irradiated zone.

To determine the integration constants in solution (8), we inserted them in eq. (5) and consequently obtain:

$$\begin{aligned}\sigma_r &= \frac{E}{1-\mu^2}\left[c_1(1+\mu)-c_2(1-\mu)\frac{1}{r^2}\right],\\ \sigma_\tau &= \frac{E}{1-\mu^2}\left[c_1(1+\mu)+c_2(1-\mu)\frac{1}{r^2}\right].\end{aligned} \tag{9}$$

Considering the contour conditions, eq. (9) for the plate placed outside the irradiated zone obtained the following form:

$$\frac{E}{1-\mu^2}\left[c_1(1+\mu)-\frac{c_2}{c^2}(1-\mu)\right]=p_k \tag{10}$$

from which the expression for the integration constants c_1 y c_2 can be presented as

$$\begin{aligned}c_1 &= 0,\\ c_2 &= \frac{p_k\, c^2\,(1+\mu)}{E}\end{aligned} \tag{11}$$

Substituting expression (10) in eq. (8) and (9), we obtain

$$u=\frac{p_k c^2(1+\mu)}{Er} \tag{12}$$

$$\sigma_r=-\frac{p_k c^2}{(1-\mu)r^2} \tag{13}$$

$$\sigma_\tau=\frac{p_k c^2}{(1+\mu)r^2} \tag{14}$$

Similarly, for the irradiated zone, the substitution of the corresponding boundary conditions leads to the following expressions:

$$\begin{aligned}\frac{E}{1-\mu^2}\left[c_1(1+\mu)-\frac{c_2}{c^2}(1-\mu)\right]&=-p_k,\\ u(r=0)=c_1 r+c_2\frac{1}{r}&=0\end{aligned} \tag{15}$$

From which we deduce:

$$\begin{aligned}c_2&=0,\\ c_1&=\frac{p_k(1-\mu)}{E}\end{aligned} \tag{16}$$

After substituting the values of c_1 and c_2 in eq. (8) and (9) we obtain:

$$u=-\frac{p_k(1-\mu)^2}{E} \tag{17}$$

$$\sigma_r = p_k \quad \text{and} \quad \sigma_\tau = p_k \tag{18}$$

3. Discussion of the results

The material in the irradiated area has been under the effect of a biaxial stress state in compression, the magnitude of which can be determined by the degree of amorphization process development. The surrounding matrix is also in a state of biaxial stress. However, the compression forces decrease in the radial direction that is inversely proportional to the square of the radius measured from the center spot of radiation, but the stress forces, which are equal by modulus to compression forces, act in the tangential direction,. A physical approach to the problem is to determine the magnitudes of the contact pressures p_k, which act in the limits of the irradiated area. To find the value of this pressure, we use the solution proposed by Lame [5], which describes the emergence of a contact pressure when a cylindrical piece joins a sheet having a central perforation with a certain tightening Δc:

$$u_2-u_1=\Delta c \tag{19}$$

where u_2 is the sheet displacement in the borders of the irradiated zone described by eq. (12), u_1 is the zone of displacement in the boundary described by eq. (17), Δc is the tightening intensity.

In this case the tightening is a result of the swelling of the irradiated zone due to amorphization development stimulated by radiation. We determined its size from the conditions of substance mass conservation in the irradiated zone during the radiation process applied to it:

$$4\pi c^2\cdot h\gamma_k=4\pi(c+\Delta c)^2\cdot(h+\Delta h)\gamma_{rz} \tag{20}$$

where γ_k is the initial crystalline pyrographite density, γ_{rz} is the present value of the material density in the irradiated zone, h is the sample thickness, *and c* is the irradiated zone radius.

After certain transformations, eq. (20) can be represented in the following form:

$$\gamma_k = \left(1+\frac{\Delta c}{c}\right)^2 \cdot \left(1+\frac{\Delta h}{h}\right) \cdot \gamma_{rz} \quad (21)$$

Recognizing that during the pyrographite amorphization process the deformation develops isotropically, i.e., $\varepsilon_x = \varepsilon_y = \varepsilon_z$, and γ_{rz} is the density of totally amorphized pyrographite γ_a, and the crystalline pyrographite is expressed as γ_k, then according to the mixture law:

$$\gamma_{rz} = \gamma_a \alpha + \gamma_k (1-\alpha) \quad (22)$$

where α is the amorphization degree.

Eq. (21) can be expressed in the following form:

$$\gamma_k = \left(1+3\frac{\Delta c}{c}\right) \cdot [\gamma_a \alpha + \gamma_k (1-\alpha)] \quad (23)$$

Solving eq. (23) with respect to *c* and substituting eq. (22) in it we obtain

$$\Delta c = \frac{c}{3}\left[\frac{\alpha(\gamma_k - \gamma_a)}{\gamma_a \alpha + \gamma_k (1-\alpha)}\right] \quad (24)$$

Substituting eq. (24) in eq. (19), we have that for $r = c$

$$\frac{p_k c \cdot (1+\mu_k)}{E_k} + \frac{p_k c (1-\mu_k)}{E_a} = \frac{c}{3}\left[\frac{\alpha(\gamma_k - \gamma_a)}{\gamma_a \alpha + \gamma_k (1-\alpha)}\right] \quad (25)$$

where eq. (26) is for the contact pressure:

$$p_k = \frac{\dfrac{\alpha(\gamma_k - \gamma_a)}{\gamma_a \alpha + \gamma_k (1-\alpha)}}{3\left[\dfrac{1+\mu_k}{E_k} - \dfrac{1-\mu_a}{E_a}\right]} \quad (26)$$

where μ_k and μ_a are Poisson coefficients for the crystalline and amorphous pyrographite, respectively; and E_k , E_a are Young moduli for the crystalline and amorphous pyrographite, respectively.

As there is no data on the physical properties of pyrographite in the amorphous state, eq. (26) can only be used to estimate the tendency in the change of the localized stress state to the degree of amorphization development. From this it is observed that in the initial state, when $\alpha = 0$, the tightening $\Delta c = 0$. The maximum value of $\Delta c = \frac{c}{3}\left(\frac{\gamma_k - \gamma_a}{\gamma_a}\right)$ is reached when pyrographite is amorphized totally in the irradiated zone. At this moment the contact pressure reaches the value of

$$p_k = \frac{\dfrac{\gamma_k - \gamma_a}{\gamma_a}}{3\left[\dfrac{1+\mu_k}{E_k} + \dfrac{1-\mu_a}{E_a}\right]} \quad (27)$$

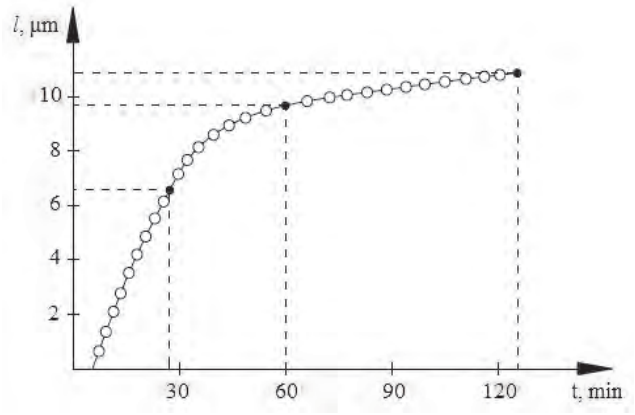


Figure 3. Kinetics of the development of a crack under irradiation conditions with high-energy electrons; experiment within the HVTEM-JEOL 1000; sample material-pyrographite
Source: The authors.

When the character of stress distribution in the matrix zone not exposed to radiation described by formula 11 is analyzed, the reasons for the emergence of cracks at the border of the irradiated area and their propagation in the radial direction become clear.

For $r = c$, the stress forces in the tangential direction reach their maximum value. The resistance under these forces is much weaker in pyrographite compared with the effects of compression forces [7]. The fact that the cracks propagate along the crystallographic directions $[0\bar{1}0], [1\bar{1}0], [100], [010], [\bar{1}10], [\bar{1}00]$ indicates that the fracture anisotropy resulting from the crystallographic structure of pyrographite expresses itself, and the characteristics of the deformation and stress state in an anisotropic approximation can be considered to be approximations. In Fig. 3, the kinetics of crack development in a matrix of pyrographite unaffected by the radiation is observed. It can be seen that the curve consists of two sections:

I. A section in which the crack grows at a rate of 1.10^{-5} cm/sec, and
II. A section in which the crack develops at a rate of 1.1 x 10^{-7} cm/sec.

4. Conclusions

1. We established that when a stream of high-energy electrons impacts pyrographite, amorphization is stimulated in it, which can be described as a transformation of the crystalline phase into the amorphous state. We studied the influence of the radiation parameters on the amorphization kinetics.

2. We proposed a phenomenological model of amorphization by radiation, which considers a possibility of a direct phase transformation of the type K → A, as well as the inverse transformation A → K. Based on this, we obtained a formula that describes the kinetics of amorphization by radiation in the following form:

$$\alpha(t) = \frac{1 - e^{-\frac{K_1}{2}\left(1+\frac{K_1}{K_2}\right)t^2}}{1 + \frac{K_2}{K_1}} \qquad (28)$$

3. We detected the effect of pyrographite amorphization under the influence of a high-energy electron stream inside an HVTEM. We proposed a phenomenological model that describes the kinetics of amorphization in pyrographite due to radiation using reference points put in the material surface. This was performed by means of the implantation of ionized copper atoms.

We studied the deformation and stress state of pyrographite, both in the zone irradiated by electrons as well as in the zones adjacent to the matrix unaffected by radiation.

We proposed a method to determine the energy threshold of radiation damage in pyrographite using data from the kinetics analysis of its amorphization under conditions of radiation by high-energy electrons performed inside the HVTEM column under various stress acceleration values.

Acknowledgment

The authors express their gratitude to all the staff at the Laboratory 20, Baikov Institute of Metallurgy, Russian Academy of Sciences, for their support. We also thank the Secretary of Research of the National Polytechnic Institute of Mexico, SIP, for their support for Project 20131380, 20141031.

References

[1] Snykers, M. and Janssens, C., The use of the JEM-1250 HVTEM of the University of Antwerpen as an instrument for void swelling simulation experiment. BLG.521, 18 P., 1978.

[2] Timofeev, V.N., Izmereniye temperatury nagreva obraztsa puchkom elektronov v prosvechivayushem elektronom mikroskope s pomoshiyu lorentsevoy elektronnoy mikroskopii. Teplofizika kondensirovannyj sred. Selected Papers. M: Nauka, pp. 74-78, 1985.

[3] Abe, H., Naramoto, H. and Kinoshita, C., Amorphization of graphite under ion or electron irradiation. Mater. Res. Soc Proc., pp.373-383, 2011.

[4] Panyukov, S.V., Subbotin, A.V. and Arzhakov, M.V., Irradiation induced dimensional changes in bulk graphite: The theory. Journal of Nuclear Materials, 439, pp. 72-83, 2013. DOI: 10.1016/j.jnucmat.2013.03.070

[5] Ponomariov, S.D., Biderman, V.L. and Lijariov, K.K., Razshety na prochnost v mashinostroyenii. M: MASHGIZ, 974 P., 1958.

[6] Timoschenko, S.P. and Voynovski-Criger, S. Plastin, I., Obolochki, M., FITMAZGIZ, 576 P., 1963.

[7] Birgiliev, Y.S., Makarchenko, V.G. and Chirilov, Y.S., Sootnosheniya mezhdu prochnostnymi jarakteristikami v obluchennom grafite. Problemy Prochnosti, 1. pp. 95-100, 1977.

P.A. Tamayo-Meza, is a Dr. of Technology and Metallurgy. He holds a PhD. in Engineering from the Academy of Sciences of the USSR, Moscow, Russia, and an MSc. in Mechanical Engineering from the Faculty of Engineering, University of Moscow, Russia. His professional experience in research includes the development of fiber reinforced materials, mechanics of materials characterization by quasi-relaxation, transmission electron microscopy, scanning, and atomic force; the study of corrosion phenomena, solid state physics, physics, and fracture mechanics; creep and fatigue, physical mechanics of materials, and theory of heat treatment; and mechanical cryogenic treatment.
ORCID: 0000-0001-8026-8928

V.A. Yermishkin, obtained his Dr. in Physical Sciences and Mathematics in Lomonosov State University. He works as chief of Laboratory of High-Voltage Electron Microscopy in the Institute of Metallurgy and Materials Science at the Russian Academy of Sciences. His professional experience in research includes physics, fracture mechanic, creep and fatigue, transmission electron microscopy, scanning and atomic force and development of fiber reinforced materials. He has published more than 132 works.
ORCID: 0000-0001-7280-4516

U.S. Silva-Rivera, obtained his Dr. of Science in Mechanical Engineering, specializing in Mechanical Design (2015) and his MSc. degree in Manufacturing Engineering (2011) from the School of Mechanical and Electrical Engineering in the National Polytechnic Institute of Mexico, and his BSc. degree in Military Industrial Engineering, specializing in Chemical Engineering (2003), from Military Engineering School in the University of the Army and Air Force. His research interests are experimental ballistics, computational fluid dynamics, materials and mechanical design; he has a knowledge of production processes and quality control of small arms ammunition, as well as in design, development and production of assault rifles.
ORCID: 0000-0001-5597-1638

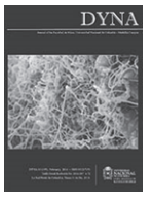
A. Leyva-Díaz, obtained his BSc. degree from the Technological Institute of Tuxtla Gutierrez in 2011, and his MSc. in Manufacturing Engineering from the School of Mechanical and Electrical Engineering in the National Polytechnic Institute of Mexico in 2014. He has a knowledge of mechanical design, static and dynamic analysis in 3D models through finite element method software. His research interests are Mechanical Design, Finite Element Analysis, and the development and characterization of new materials.
ORCID: 0000-0002-7478-9441

J.O. Trejo-Escandón received his BSc. in Mechanical Engineering in 2012, from the Technological Institute of Tuxtla Gutierrez, Mexico and a MSc. in Manufacturing Engineering in 2015, from the National Polytechnic Institute - IPN, Mexico. He currently works at the Development & Innovation Center in Schneider-Electric in Monterrey, Mexico, where he works as a Structural Simulation Specialist. His areas of research interest focus on Mechanical Design, Finite Element analysis and development and characterization of new materials.
ORCID: 0000-0002-5076-7045

J.M. Sandoval-Pineda, is a professor in the National Polytechnic Institute of México. He obtained his Dr. of Science in Mechanical Engineering with honors in 2008 and his MSc. in Mechanical Engineering in 2004, from the National Polytechnic Institute, Mexico. His professional experience in research includes physics, fracture mechanic, creep and fatigue, and transmission electron microscopy.
ORCID: 0000-0002-6529-7920

L.A. Flores-Herrera, is a professor in the National Polytechnic Institute of México. He obtained his Dr. of Science in Mechanical Engineering with honors in 2007 and his MSc. in Mechanical Engineering in 2003 from the National Polytechnic Institute, Mexico. Hi is a specialist in Finite Element Analysis and his research area focuses on Microelectromechanical systems (MEMS).
ORCID: 0000-0003-1081-5193

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS



DYNA
http://dyna.medellin.unal.edu.co/

In vitro behavior of the dentin and enamel calcium hydroxyapatite in human premolars subjected to high temperatures

Sebastián Medina [a], Liliana Salazar [b], Carlos Mejía [c]& Freddy Moreno [d]

[a] *School of Basic Sciences at Universidad del Valle, Cali, Colombia. sebastianmedina.c@gmail.com*
[b] *School of Basic Sciences at Universidad del Valle, Cali, Colombia. liliana.salazar@correounivalle.edu.co*
[c] *School of Dentistry at Universidad del Valle, Cali, Colombia. camejia@emcali.net*
[d] *Faculty of Health of Sciences at Pontificia Universidad Javeriana, Cali, Colombia. fmorenog@javerianacali.edu.co*

Received: March 21th, 2014. Received in revised form: May 20th, 2015. Accepted: December 10th, 2015.

Abstract
This paper details a cross-sectional, descriptive observational in vitro study of a pseudo-experimental nature that analyzes Electron Microscopy (SEM) by scanning the physical behavior of enamel and dentin calcium hydroxyapatite. The purpose is to describe the separation of these two mineralized dental tissues at the dentin-enamel junction (DEJ) when the teeth are subjected to high temperatures. This study provides scientific evidence that may broaden the discussion on the use of separation of the dentin-enamel junction as a constant and repetitive reliable marker for forensic use (use in forensic sciences) that can contribute to the dental identification process and documentation in a legal medical autopsy, given a situation in which bodies or human remains have been burned, carbonizated or incinerated.

Keywords: Calcium hydroxyapatite, enamel, dentin, dentin-enamel junction, high temperatures, odontological identification, forensic dentistry.

Comportamiento *in vitro* de la hidroxiapatita de calcio de la dentina y del esmalte en premolares humanos sometidos a altas temperaturas

Resumen
Estudio observacional descriptivo de corte transversal y de naturaleza pseudo-experimental in Vitro que analiza, a través de microscopía electrónica de barrido (MEB), el comportamiento físico de la hidroxiapatita de calcio del esmalte y la dentina; con el propósito de describir la separación de estos dos tejidos mineralizados dentales a nivel de la unión amelo-dentinaria cuando los dientes son sometidos a altas temperaturas. Con ello, se proporciona evidencia científica que amplía la discusión sobre el empleo de la separación de la interfase esmalte-dentina como un marcador fehaciente constante y repetitivo de uso forense, el cual puede contribuir con los procesos de identificación odontológica y documentación de la necropsia médico legal para el caso de cadáveres o restos humanos que resulten quemados, carbonizados o incinerados.

Palabras clave: Hidroxiapatita de calcio, esmalte, dentina, unión amelo-dentinaria, odontología forense, identificación odontológica.

1. Introduction

In Colombia, deaths involving the burning, carbonization and incineration of human body [1,2] have increased throughout the first decade of the twenty-first century. This state hinders identification procedures that are commonly employed methods in forensic sciences [3], including technically and scientifically supported procedures by forensic dentistry such as restorative dental treatments [4], or electronic devices implanted in teeth. These require a knowledge of the biological behavior of interfaces between dental tissues and different biomaterials [5].

Teeth are the best preserved organs in extreme conditions (including temperatures) [6]. However, there is little research on the changes that occur in dental tissues when they are subjected to high temperatures. Early in vitro studies focused

mainly on the macroscopic description of the structural changes, but did not develop markers with sufficient macroscopic and microscopic scientific support that would allow the application of the results to procedures that help identify a human corpse or human burned, carbonized or incinerated remains [7].

Several studies [8,10,11] have reported that as temperature increases, enamel is detached from dentin at the dental-enamel junction (DEJ). This has been associated with the burning of organic matrix components (non-fibrillar component –glycoproteins, glycosaminoglycans and proteoglycans– and fibrillar component –collagen–), and the physicochemical changes of the inorganic component (calcium hydroxyapatite crystals) of these mineralized dental tissues. It results in decreased tissue volume and loss of DEJ continuity; therefore, the purpose of this study is to describe in vitro behavior of dentin and enamel calcium hydroxyapatite in human premolars that are subjected to high temperatures and, more specifically, to determine whether there is a causal relationship between high temperatures and the separation of the DEJ phenomenon.

2. Dentin

Dentin is a mineralized biological tissue that is composed of 70% inorganic material (hydroxyapatite crystals and trace elements), 18% organic material (type I collagen fiber and proteins such as osteonectin, osteopontin, Osteoclastin-like dentin Gla protein, dentin phosphorine, dentin matrix protein and dentin sialoprotein), and 12% water. Moreover, this tissue has dentinal tubules mainly occupied by odontoblastic processes, and cell specializations extending from pulp odontoblasts, which is why the dentin-pulp complex is described as an integrated structural unit formed by the body of the odontoblast (pulp) and the odontoblast process (dentin) [12-14].

3. Enamel

Enamel is a highly mineralized acellular biological tissue lining like a cap, which is the outer surface of mammals' teeth. Given that this is the hardest and most resistant tissue in the body, the function of enamel is to protect the dentin-pulp complex. It is constituted of 95% inorganic material (calcium hydroxyapatite crystals), 2% organic material (proteins such as amelogenin, enamelin, ameloblastin and tuftelin, among others), and 3% water. As a mineralized tissue, enamel has a functional unit called a "prism", "cane" or "enamel rod", which corresponds to a number of calcium hydroxyapatite crystals parallel to the longitudinal axis of the prism. These are packaged in an organizational pattern, cylindrical in shape and arranged in rows of horizontal alignment. This arrangement occupies the entire thickness of the enamel, except the part located near the dentin, which is not in aprismatic [12-14].

4. Dentin-enamel junction

From when they develop, dentin and enamel are linked by an interface called dentin-enamel junction (DEJ), which is a common starting point for both tissues. This guarantees that dentin and enamel calcium hydroxyapatite crystals never come into contact, as the junction is constituted by amorphous regions the extracellular matrix of which were not completely mineralized: these are respectively called the mantle dentin and aprismatic enamel [15,16].

As described above, the progressive and opposite mineralization of both DEJ tissues begins with the deposition of calcium and phosphate ions on the extracellular matrix. Thus, the DEJ's solution of continuity is due to the hydroxyapatite crystal growth, which results from the synthesis of the organic extracellular matrix and its subsequent mineralization in both tissues, first of the dentin and the enamel immediately after. Therefore, the DEJ corresponds to a transition zone with a scalloped or corrugated appearance between the enamel and the dentin [17,18]. The scallop pattern is represented on the surface of the dentine by a series of peaks that rise towards the enamel, creating a series of concave tunnels with a "honeycomb" aspect that correspond to convexities on the surface of the enamel that belong to groups of enamel prisms [18,19]. The DEJ is therefore a complex mechanism of attachment, the primary function of which involves the formation of an activity surface for the odontoblasts and the ameloblasts to secrete dentin and enamel respectively. Once the tooth is completely formed it contributes to the biomechanical integrity of the structure of which these two tissues are composed [20-22].

5. Calcium hydroxyapatite

Robinson et al [23], expressed in their study undertaken with atomic force microscopy that the development of the ultrastructure of mineralized dental tissue encompasses a series of spherical subunits with a diameter that is similar to a hydroxyapatite crystal (300 and 500 nm). As such, each nanosphere forms an enucleation center that is made of amorphous calcium phosphate and stabilized by the protein components of the extracellular matrix. The longitudinal fusion of nanospheres constitutes a typical hydroxyapatite crystal.

Calcium hydroxyapatite crystal (CHC) corresponds to a biomineral formed by crystalline calcium phosphate that is stored in mineralized tissues, and represents 99% of calcium deposit and 80% of total phosphorus in the body. These calcium phosphates are spherical beads of octacalcium phosphate, arranged in three-dimensional calcium hydroxyapatite crystals, which eventually form enamel prisms [12-14]. Thus, octacalcium phosphate has been recognized as the precursor of hydroxyapatite [24] as it contributes to the formation of the initial mineral phase of the extracellular matrix during biomineralization, and to the subsequent formation of calcium hydroxyapatite [2] by crystallization.

6. Dental tissue mineralization

Biomineralization is the process by which specialized mineralized connective tissues (bone, cartilage, enamel, dentin and cementum) are built by the deposition of a mineral or inorganic phase on the organic extracellular matrix.

Regarding the mineralized tissues in teeth, the biomineralization occurs by deposition of calcium hydroxyapatite crystals on the extracellular matrix during the process of amelogenesis (enamel), dentinogenesis (dentin), cementogenesis (cement) and osteogenesis (alveolar bone) mediated by ameloblasts, odontoblasts, and cementoblast and osteoblasts respectively. It is genetically regulated by a group of extracellular matrix mineralization proteins recognized as SCPP (Secretory Calcium-binding Phosphoprotein) [26-28].

During amelogenesis, which is mineralization of the extracellular matrix of enamel, there is a six-step biological crystal formation process from substitutes: 1. Delimitation (ameloblasts are stimulated and begin the secretion of the extracellular matrix from their Tomes's processes); 2. Existence of a preformed organic matrix (built as a structural frame of proteins secreted by the ameloblasts that put themselves together as nanospheres of amelogenin); 3. Oversaturation of the extracellular matrix (creation of a saturated solution of calcium and phosphate ions secreted by ameloblasts); 4. Control of the crystal cores formation (enucleation or crystal self-assembly to set crystal cores controlled by enamelin, tuftelin, amelogenins, ameloblastins and dentin sialophosphoprotein); 5. Control of growth, morphology and orientation of crystals by the extracellular matrix; and 6. Control of the completion of crystal growth (ripening of CHC and proteolytic degradation of excess organic content in the extracellular matrix). This sequential, longitudinal and progressive growth process, starts at the DEJ and ends at the enamel surface [29-31].

With Regards to dentinogenesis, the dentin mineralization mechanism differs considerably from enamel mineralization. Odontoblasts initiate the biomineralization process, which consists of seven consecutive steps: 1. Fibrillar extracellular matrix synthesis (collagen frame); 2. Capture and storage of intracellular calcium; 3. Local concentration of calcium and phosphate ions; 4. Formation of matrix vesicles for its calcification; 5. Release of vesicles from the odontoblast to the extracellular matrix (within these vesicles, amorphous calcium phosphate dots merge together to form CHC); 6.Vesicles rupture and CHC release; and 7. Orientation of CHC in relation to the collagen fibers [29, 32].

7. Materials and methods

This is a cross-sectional (descriptive observational) in vitro study of a pseudo-experimental nature in which the exposure of the sample was undertaken by convenience. Through scanning electron microscopy (SEM) this paper describes the behavior of calcium hydroxyapatite in dentin and enamel human premolars in order to determine whether there is a causal relationship in the detachment phenomenon that occurs between the dentin and enamel at the DEJ when teeth are subjected to high temperatures.

7.1. Sample collection

Upon obtaining endorsement from the Human Ethics Committee of the Health Faculty at the Universidad del Valle, we proceeded to collect a sample of 60 premolar teeth, healthy under clinical observation (no cavities or fractures).

Table 1.
Classification and distribution of the sample before being subjected to high temperatures.

Groups	**Control**	**Intervention**				
Temperature (**°C**)	0	200	400	600	800	1.000
Teeth	10	10	10	10	10	10

Source: The authors

They had no dental treatment and were obtained from patients who attended the Oral Surgery Clinic in the Dental School at the Universidad del Valle, who required premolar extraction for orthodontic reasons and had previously signed an informed consent.

7.2. Handling and preservation of the sample

Once the teeth were extracted, we proceeded to wash them profusely with tap water to eliminate traces of blood and tissue, and placed them in a dark, tightly sealed container with a Chloramine T fixative solution at 5%. The teeth remained in Chloramine T for a week and were then placed in saline solution at room temperature according to what is stipulated in ISO/DIS 11405:2003 [35].

7.3. Sample distribution

Teeth were Classified and randomized according to the temperature to which they were subjected. A control group of ten teeth, which were not subjected to high temperatures, was established in order to have a control (Table 1).

7.4. Application of high temperatures

Once the sample was collected the intervention groups of teeth were individually placed in small trays made of investment material (Whip mix Serafina®), designed under the protocol established by the Dental Materials Unit of the University of Pavia [8]. We then proceeded to place them in a muffle furnace (Thomas Benchtop 1256®) in groups of 10, starting with 30°C for each temperature range (200°C, 400°C, 600°C, 800°C and 1000°C). For example, ten teeth in the 200°C group were put in the furnace, each one in its corresponding tray, at a temperature range of 30°C to 200°C. The furnace was allowed to cool back to room temperature before we took the trays with the teeth out. The same procedure was repeated for the groups of teeth subjected to 400°C, 600°C, 800°C and 1000°C temperatures. This protocol is a standardized practice at the School of Dentistry, University of Valle [11].

7.5. Management of the teeth after application of high temperatures

An acrylic self-curing resin base (New Stetic®) was made for the control group and for the intervention group that was subjected to 200ºC and 400ºC. These teeth were mounted on a universal testing machine (Tinius Olsen® H50KS®) in a Physical and Mechanical testing Laboratory at the School of Materials Engineering (EIMAT) in the Universidad del

Valle, which has a 50 Kilo-Newton capacity. A constant vertical, compressive force was applied to them at a crosshead speed of 1 millimeter per minute with a rounded, hardened steel tip with a 3 mm diameter located between the buccal and palatal or lingual cusps until the software machine (Tinius Olsen Horizon®) detected a fracture and the crown was fragmented. This process was undertaken with the objective of provoking adhesive failure, which separates enamel from dentin by unevenly distributed tensional forces. For the teeth in the intervention group that were subjected to 600°C, 800°C and 1000°C, the tissues were separated with a scalpel (Brad Parker® No. 15) and a carver (Lecron Medix®), as these work well with the DEJ's structural weakness and natural spontaneous fracture after having been exposed to high temperatures. From this we gained samples of enamel and dentin.

7.6. Observation of samples of enamel and dentin

Fragments of enamel and dentin that corresponded to each other were set on a glass slide with cyanoacrylate, coated with a gold film, and observed and photographed through a SEM using a JEOL® JSM 6490 LV® from the School of Engineering Materials (EIMAT) at the Universidad del Valle Engineering School. They had a voltage acceleration range of 0.3 KV to 30 KV from a 3nm vacuum source of electrons at a high voltage acceleration.

8. Results

When a tooth is exposed to high temperatures, structural changes that depend on the maximum temperature reached may occur. Consequently, it can stay intact (at 200°C), or it can be burned – change color and fissures and cracks may form (at 400°C)–, it may be carbonized –reduced to charcoal by incomplete combustion (at 600°C)–, be incinerated – reduced to ashes (at between 800°C - 1000°C)–, or burst – radicular and coronal outbreak (at 1200°C)–. Based on what has been previously stated the results of this study and the subsequent discussion section will be affirmed by describing the changes mentioned and we will focus on the changes that took place in the DEJ region.

8.1. Macroscopic detachment of enamel and dentin at the DEJ

At 200°C there are no macro-structural changes that affect the DEJ beyond the loss of brightness (opacity of the enamel), which is associated with dehydration and cement (Fig. 1A). This was observed when the teeth were compared before and after being subjected to high temperatures. At 400°C, the enamel became opaque and brown (coronal dentin and showed signs of combustion of the organic extracellular matrix); in the cervical region loss of continuity of the enamel was observed by longitudinal and transverse fractures, as well as dentin detachment, indicating loss of DEJ continuity (Fig. 1B). At 600°C, white chalk-colored enamel was found due to incineration and coronal dentin adopted a grayish appearance by a from the carbonization phase to the incineration phase. In the cervical region fragmentation of enamel was observed due to deepening of surface cracks and crevices; this revealed an absolute loss of DEJ continuity (Fig. 1C). At 800°C incinerated enamel adopted a grayish color, and a greater detachment of the enamel dentine and loss of DEJ continuity (Fig. 1D) was observed in the cervical region. Finally, at 1000°C the crown of the teeth turned white because of the burning of enamel and dentin. In the cervical region the DEJ separation was much more evident due to incineration and removal of enamel in addition to the decreased volume of the root with respect to the crown (Fig. 1E).

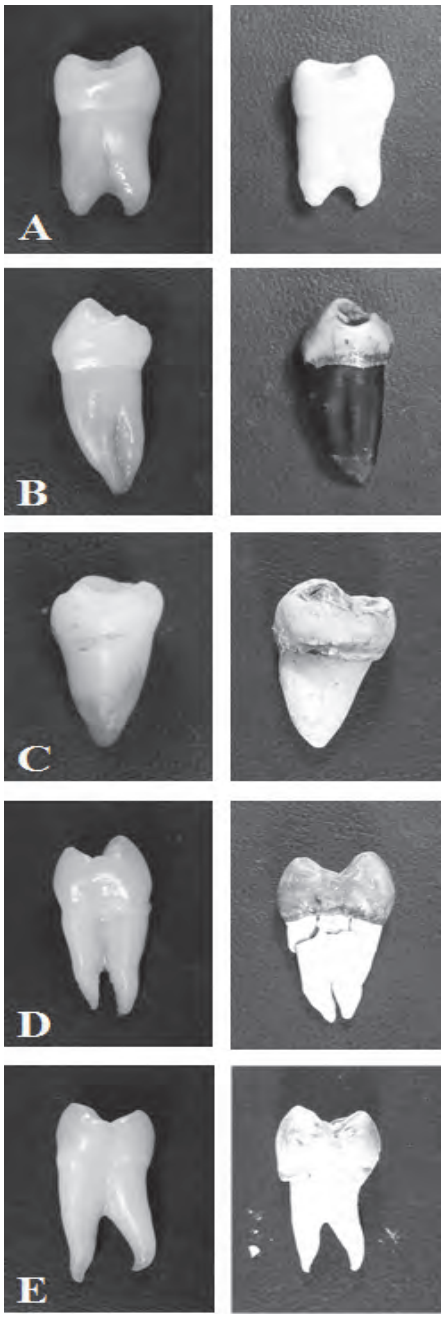


Figure 1. Teeth before (left) and after (right) being subjected to high temperatures. We can observe the progressive separation of the DEJ in the cervical third of the crown. A. 200°C; B. 400°C; C. 600°C, D. 800°C, and E. 1000°C.
Source: The authors.

8.2. CHC of enamel and dentin

SEM was used to observe the microscopic behavior of the inorganic component (mineral phase) of aprismatic enamel and mantle dentin (respective surfaces of the DEJ) represented in the CHC.

In the teeth samples that were not subjected to any change of temperature there was a regular pattern of calcium hydroxyapatite nanospheres comprising octacalcium phosphate, and a homogenous size distribution within enamel was observed. The presence of the pattern was seen in the collagen fiber network (Fig. 2A and 2B) in the dentine.

At 200°C, a more compact pattern of calcium hydroxyapatite and dehydration associated with the start of combustion of the low organic component was observed in the enamel. In the dentin, calcium hydroxyapatite began the process of compaction (Fig. 3A and 3B).

At 400°C the fusion of calcium hydroxyapatite associated with the coalescence of octacalcium phosphate nanospheres that were virtually merged was observed in the enamel. In dentin, compact calcium hydroxyapatite was noticed (Fig. 4A and 4B).

At 600°C, complete melting of the octacalcium phosphate and calcium hydroxyapatite nanospheres was observed in the enamel. In the dentin, a fully fused inorganic component was found and some isolated octacalcium phosphate nanospheres were seen (Fig. 5A and 5B).

At 800°C the loss of enamel micro-morphology and the enlarged size of the octacalcium phosphate nanospheres were observed. In the dentin a homogeneous melting pattern of hydroxyapatite (Fig. 6A and 6B) was observed.

Finally at 1000°C, the integration of octacalcium phosphate nanospheres in separate and larger clusters was noticed in the enamel. Molten dentin continued to have a homogeneous pattern of calcium hydroxyapatite (Fig. 7A and 7B).

9. Discussion

9.1. Dentin and enamel detachment at the DEJ

Merlati et al [7] described DEJ detachment beginning at 400°C. Moreno et al [36], discussed that DEJ detachment begins at 200°C, and that it becomes more evident at 400°C when an outbreak of cervical enamel occurs. At 600°C and at 800°C, full enamel-dentin detachment at the cervical and the middle thirds is observed. When the temperature reaches 1000°C enamel is fragmented and separated from dentin just like the way that a cap is removed. These authors ascribe the detachment phenomenon to the chemical composition of enamel and dentin.

Because enamel has a high inorganic content, which is represented by octacalcium phosphate as CHC, and a low organic content and water (undergoing combustion and evaporation, respectively), the temperature increase alters the organization of such crystals, augmenting their cohesion (thermal contraction). The latter generates the initial appearance of fissures and cracks, which provide a cracked aspect, and ultimately, lead to their fracture.

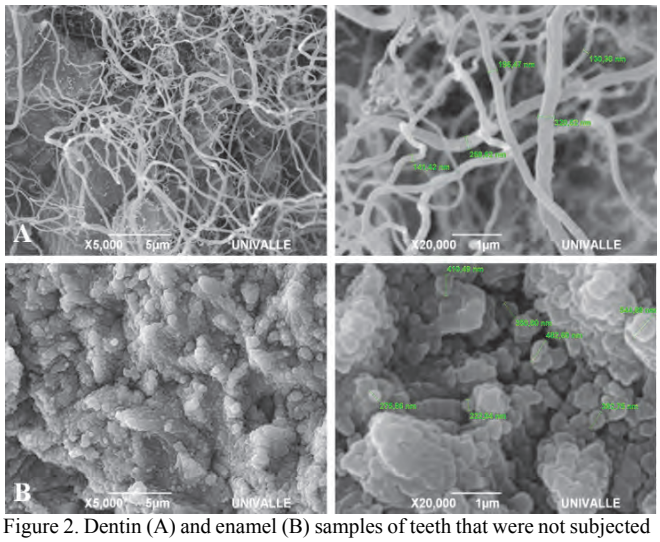


Figure 2. Dentin (A) and enamel (B) samples of teeth that were not subjected to high temperatures.
Source: The authors.

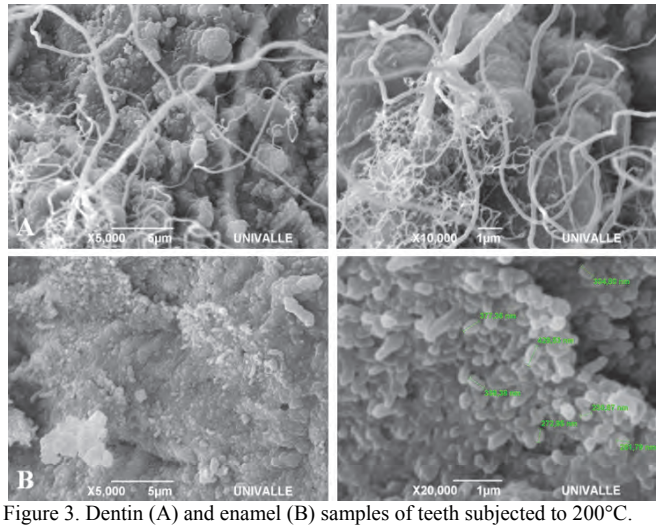


Figure 3. Dentin (A) and enamel (B) samples of teeth subjected to 200°C.
Source: The authors.

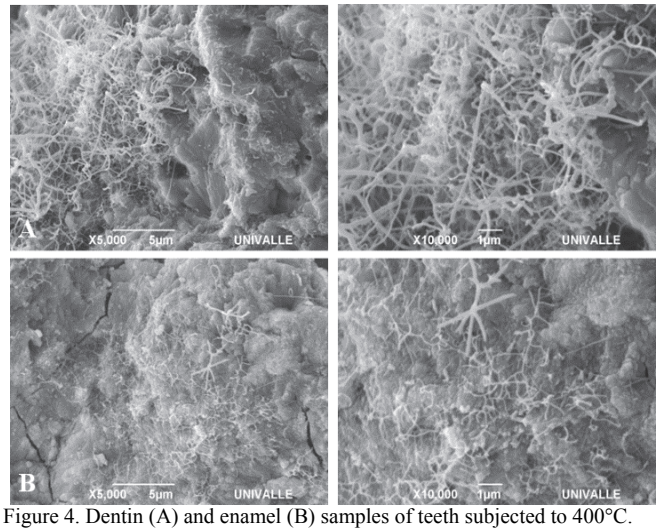


Figure 4. Dentin (A) and enamel (B) samples of teeth subjected to 400°C.
Source: The authors.

Also, a higher water and organic content means that dentin takes longer to dehydrate and, because it is protected by enamel itself, in comparison with enamel, it provides a certain spectrum of thermal contraction. This phenomenon is more evident in the cervical third, where the scalloped pattern is reduced.

In the gross examination of the samples that were subjected to temperatures of 200°C, no visible change was observed in the cervical region of the coronal third, as stated by Merlati et al [7] and Moreno et al. [36] At 400°C and at 600°C, enamel and dentin began their process of carbonization. Aprismatic enamel with less inorganic substance quickly lost its low water content and protein, and underwent coalescence of octacalcium phosphate nanospheres, which gradually increased in size and fused together, conforming the CHC. Exactly the opposite happened to the mantle dentin, which has a much higher organic content, which is the reason why that when water was lost and the protein component was denatured, the mantle dentin lost more volume. This caused a burst of enamel by dimensional shrinkage of the dentin at a DEJ level, first at the cervical third and then gradually, up until the middle third. At 800°C and 1000°C, enamel was fully compacted and dentin had gradually begun the process of incineration of the organic component. Thus, the shrinkage phenomenon became much more obvious, which is the reason why the enamel was completely separated from the coronal dentin (in the manner of a cap), as was described by Moreno et al [36].

This spontaneous separation of enamel and dentin was a gradual process that started between 200°C and 400°C and became more evident as the temperature increased up until its total separation that occurred at 1000°C. The explanation of this macroscopic change can be found in the scalloped pattern of the DEJ, which according to Gallagher et al [16], Marshall et al [18], Habelitz et al [20], Imbeni et al [21] and Radlanski and Renzen [22], is very marked in the cusps zone of the DEJ and gradually fades until it disappears in the cervical region. Hence, the latter is the region of least resistance to thermal shock between enamel and dentin, and is the site at which the DEJ initially loses its continuity solution, even at the lowest temperature.

9.2. Behavior of enamel and dentin CHC

Regarding the organization of CHC, no report in the literature has explained in detail their behavior, both in enamel and dentin when subjected to high temperatures.

Holden et al [37] tackled the subject, but studied bone tissue through SEM analysis. They subjected fragments of has disintegrated in such way that the collagen framework is fully exposed. Similarly, the hydroxyapatite crystals are found to be in a spherical shape. Between 800°C and 1000°C incinerated bone looks like white chalk, morphology of concentric lamellae is lost and CHC may present two different morphologies: hexagonal and spherical. Finally from 1000°C to 1600°C, CHCs are fused together; this finding is associated with the sintering process (coalescence at high temperatures) that is much more evident in the inside of the osteon, which is completely disorganized. This is the reason why the classic pattern of the concentric lamellae around the Havers duct cannot not observed.

Subsequently, Holden et al [38], subjected bone fragments to high temperature and analyzed them by X-ray diffraction. They found that the number of CHCs increase as the temperature rises, beginning at 600°C (termed by the authors as re- crystallization). Above 1000°C, CHCs are broken down to calcium oxide, above 1200°C to calcium triphosphate and above 1400°C to phosphate calcium oxide.

Venkatesan and Kim [39] conducted a study in which fish tuna vertebrae were subjected to high temperatures (from 200°C to 1200°C) in order to characterize the behavior of the hydroxyapatite by infra-red spectroscopy, X-ray, TEM and SEM. The authors concluded that as the temperature increases CHCs become agglomerated, beginning at 600°C: temperature at which the micro-crystals grow in size (from 80 to 300nm). In this study, the behavior of bone tissue in terms of color changes determines carbonization signs from 200°C, complete carbonization by burning of the organic matrix (water and protein components) at 600°C and incineration from 800°C. This is similar to what occurs to dentin.

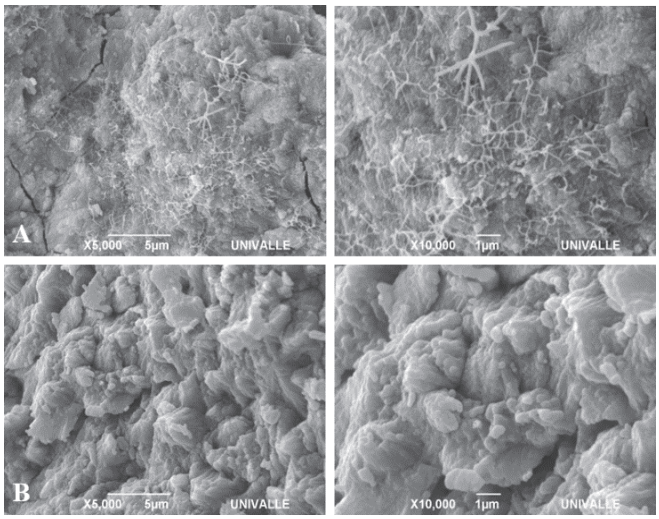


Figure 5. Dentin (A) and enamel (B) samples of teeth subjected to 600°C. Source: The authors.

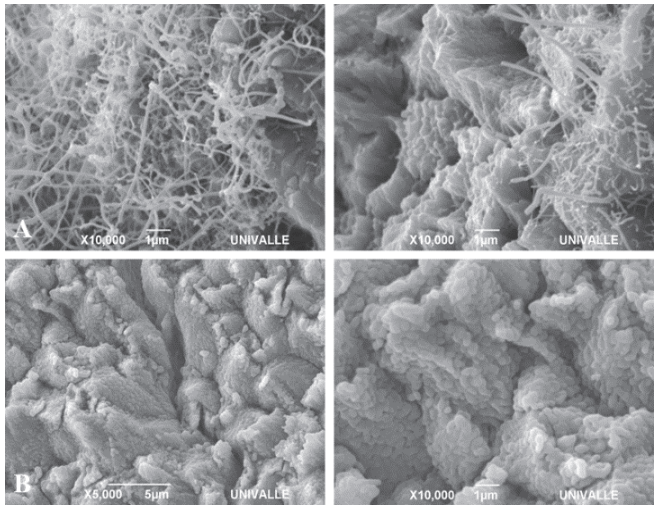


Figure 6. Dentin (A) and enamel (B) samples of teeth subjected to 800°C. Source: The authors.

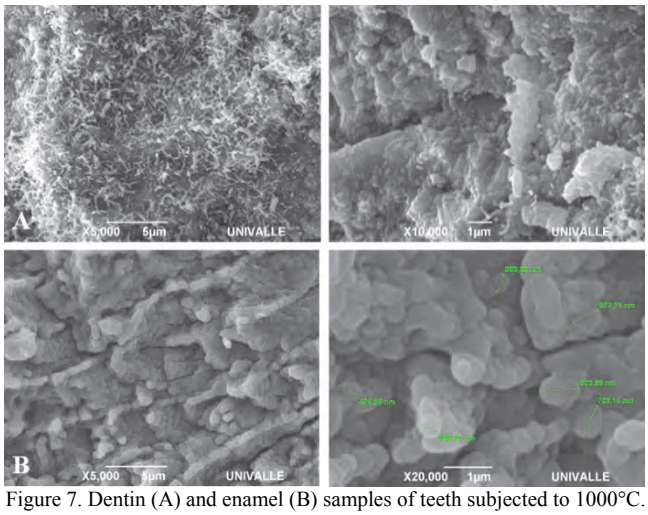


Figure 7. Dentin (A) and enamel (B) samples of teeth subjected to 1000°C. Source: The authors.

In this study, it was evident that CHC aprismatic enamel and mantle dentin, which constitute the enamel-dentine junction, underwent the same changes that were described in bone tissue. Enamel and dentin start their combustion of extracellular organic matrix at 200°C. At 400°C carbonization starts, which leads to incineration between 600°C and 1000°C. With increasing temperature, the CHC adopts a rounded appearance. This process is most evident in aprismatic enamel given its high inorganic content, whereas in dentin, given its high organic content, changes in calcium hydroxyapatite can only be seen starting from a temperature of 800°C.

10. Conclusions

Microstructural changes (fusion of CHC and sintering of octacalcium phosphate nanospheres) in enamel and dentin, when they are subjected to high temperatures, explain the specific macrostructural changes (DEJ separation) in each temperature range: these being constant and repetitive.

Separation of enamel and dentin at the DEJ can be a reliable medicolegal marker to be able to approximate the temperature to which teeth were subjected. This can be undertaken during the forensic dental identification and documentation of the forensic autopsy process, if there is enough scientific support, in burned, carbonized and incinerated individuals.

It is recommended that the analysis of the behavior of CHC at high temperatures be extended in other mineralized dental tissues (cement, alveolar compact bone and spongy bone), both in interfaces and free surfaces. In order to do this we suggest the design of an animal model in which domestic pig teeth that are articulated on their own socket (not isolated) are subjected to high temperatures for the purpose of being closer to the in vivo conditions.

Acknowledgments

This research was funded by the COLCIENCIAS 2012 Young Researchers internship grant (Programa de Jóvenes Investigadores e Innovadores "Virginia Gutierrez de Pineda" COLCIENCIAS), and by the Vice-President of Research at the Universidad del Valle.

References

[1] Ramírez I.A., Castaño A., González J.O. y Hernández H.W., Homicidios Colombia 2005, en Instituto Nacional de Medicina Legal y Ciencias Forenses. Forensis 2005, datos para la vida. Santa Fe de Bogotá, Imprelibros, 2006. pp. 27-75.

[2] González J.O. y Hernández H.W., Muertes accidentales Colombia 2005. En Instituto Nacional de Medicina Legal y Ciencias Forenses. Forensis 2005, datos para la vida. Santa Fe de Bogotá, Imprelibros, 2006. pp. 254-315.

[3] Pretty, I.A. and Sweet, D., A look at forensic dentistry. Part 1: The role of teeth in the determination of human identity, Br Dental J, 190, pp. 359-366, 1995. DOI: 10.1038/sj.bdj.4800972.

[4] Bowers, C.M. and Bell, G.L., Manual of forensic odontology. 3rd edition. Colorado Springs: American Society of Forensic Odontology, 1997.

[5] Aragón, N., Moreno, F. and Salazar, L., In vitro behavior of interfaces in human molars with an implanted passive RFID microchip and subjected to compression forces, DYNA, 80(178), pp. 5-10, 2013.

[6] Delattre, V.F., Burned beyond recognition: Systematic approach to the dental identification of charred human remains, J Forensic Sci, 45, pp. 589-596, 2000.

[7] Merlati, G., Danesino, P., Savio, C., Fassina, G., Osculati, A., Menghini, P., Observations of dental prostheses and restorations subjected to high temperatures: experimental studies to aid identification processes, J Forensic Odontostomatol , 20(2), pp.17-24, 2002.

[8] Merlati, G., Savio, C., Danesino, P., Fassina, G., Menghini, P. Further Study of restored and unrestored teeth subjected to high temperatures, J Forensic Odontostomatol, 22, pp.17-24, 2004.

[9] Rötzscher, K., Grundmann, C., Benthaus, S., The effects of high temperatures on human teeth and dentures, Int Poster J Dent Oral Med, 6, Poster 213, 2004.

[10] Ferreira, J.L., Espina-de Ferreira, A. and Ortega, A.I., Methods for the analysis of hard dental tissues exposed to high temperatures, Forensic Sci Int, 178, pp. 119-124, 2008. DOI: 10.1016/j.forsciint.2007.12.009.

[11] Moreno, S., Merlati, G., Marín, l., Savio, C. and Moreno, F., Effects of high temperatures on different dental restorative systems: Experimental study to aid identification processes, J Forensic Dent Sci, 1(1), pp. 17-23, 2009. DOI: 10.4103/0974-2948.50883.

[12] Gómez-de Ferrais, M.E. y Campos, A., Histología y embriología bucodental. Segunda edición. Buenos Aires: Panamericana, 2002.

[13] Garant, P.R., Oral cells and tissues. Chicago: Quintessence Books; 2003.

[14] Nanci, A., Ten Cate's Oral Histology: Development, Structure, and Function. 8th ed. St. Louis: Elsevier Mosby, 2008.

[15] Whittaker, D.A., The enamel-dentine junction of human and Macaca irus teeth: A light and electron microscopic study, J Anat, 125(2), pp. 323-335, 1978.

[16] Gallagher, R.R, Demos, S.G., Balooch, M., Marshall, G.W. and Marshall, S.J., Optical spectroscopy and imaging of the dentin-enamel junction in human third molars, J Biomed Mater Res, 64A, pp. 372-377, 2003. DOI: 10.1002/jbm.a.10436.

[17] Chan, Y.L., Ngan, A.H. and King, N.M., Nano-scale structure and mechanical properties of the human dentine-enamel junction, J Mech Behav Biomed Mater, 4(5), pp. 785-795, 2011. DOI: 10.1016/j.jmbbm.2010.09.003.

[18] Marshall, S.J., Balooch, M., Habelitz, S., Balooch, G., Gallagher, G. and Marshall, W., The dentin-enamel junction. A natural multilevel interface, J Eur Ceramic Soc, 23, pp. 2897-2904, 2003. DOI: 10.1016/S0955-2219(03)00301-7

[19] Smith, T.M., Olejniczak, A.J., Reid, D.J., Ferrell, R.J. and Hublin, J.J., Modern human molar enamel thickness and enamel-dentine junction shape, Arch Oral Biol, 51, pp. 974-995, 2006. DOI: 10.1016/j.archoralbio.2006.04.012.

[20] Habelitz, S., Marshall, S.J, Marshall, G.W.and Balooch, M., The functional width of the dentino-enamel junction determined by AFM-

based nanoscratching, J Struct Biol, 135, pp. 294-301, 2011. DOI: 10.1006/jsbi.2001.4409.
[21] Imbeni, V., Kruzic, J.J., Marshal, G.W., Marshal, S.J. and Ritchie, R.O., The dentin-enamel junction and the fracture of human teeth, Nature Mater, 4, pp. 229-232, 2005. DOI: 10.1038/nmat1323.
[22] Radlanski, R.J. and Renz, H., Insular dentin formation pattern in human odontogenesis in relation to the scalloped dentino-enamel junction, Ann Anat, 189, pp. 243-250, 2007. DOI: 10.1016/j.aanat.2006.11.007.
[23] Robinson, C., Fuchs, P. and Weatherell, J.A., The appearance of developing rat incisor enamel using a freeze fracturing technique, J Cryst Growth, 53, pp. 160-165, 1981. DOI: 10.1016/0022-0248(81)90062-2.
[24] Ijima, K.H., Wakamatsu, N., Goto, T., Doi, Y. and Moriwaki, Y., Effects of ca addition on the formation of octa-calcium phosphate and apatite in solution at ph 7.4 and at 37°C, J Cryst Growth, 193(1), pp. 182-188, 1998. DOI: 10.1016/S0022-0248(98)00455-2.
[25] Ijima, M. and Moriwaki, Y., Effects of ionic inflow and organic matrix on crystal growth of octacalcium phosphate; relevant to tooth enamel formation, J Cryst Growth, 198/199, pp. 670-676, 1999. DOI: 10.1016/S0022-0248(98)00986-5.
[26] Veis, A., Mineralization in organic matrix frameworks, Rev Mineral Geochem, 54, pp. 249-289, 2003.
[27] Kawasaki, K., Suzuki, T. and Weiss, K.M., Genetic basis for the evolution of vertebrate mineralized tissue, PNAS, 101(31), pp. 11356-11361, 2004. DOI: 10.1073/pnas.0404279101.
[28] Wilt, F.H., Developmental biology meets materials science: Morphogenesis of biomineralized structures, Dev Biol, 280, pp. 15-25, 2005. DOI: 10.1016/j.ydbio.2005.01.019.
[29] Thesleff, I. and Nieminent, P., Tooth morphogenesis and cell differentiation, Curr Op Cell Biol, 8, pp. 844-850, 1996. DOI: 10.1016/S0955-0674(96)80086-X.
[30] Fincham, A.G, Moradian-Oldak, J. and Simmer, J.P., The structural biology of the developing dental enamel matrix, J Struct Biol, 126(3), pp. 270-299, 1999. DOI: 10.1006/jsbi.1999.4130.
[31] Simmer, J.P. and Fincham, A.G., Molecular mechanisms of dental enamel formation, Crit Rev Oral Biol, 6(2), pp. 84-108, 1995. DOI: 10.1177/10454411950060020701.
[32] Ruch, J.V., Lesot, H. and Begue-Kin, C., Odontoblast differentiation, Int J Dev Biol, 39, pp. 51-68, 1995.
[33] MINISTERIO DE LA PROTECCIÓN SOCIAL., Resolución por la cual se establecen las normas científicas, técnicas y administrativas para la investigación en salud. [Online]. Resolución 008430/1993 de 4 de Octubre [Date of reference: January 2014]. Available at: http://www.minproteccionsocial.gov.co/vbecontent/library/documents/DocNewsNo267711.pdf
[34] ASOCIACIÓN MÉDICA MUNDIAL., Principios éticos para las investigaciones médicas en seres humanos, Declaración de Helsinki. [Online]. Finlandia, junio 1964. [Date of reference: January 2014]. Available at: http://www.wma.net/s/policy/b3.htm
[35] ISO 11405., Dental materials: Testing of adhesion to tooth structure. International Organization for Standardization, 2003.
[36] Moreno, S., León, M.E., Marín, L.y Moreno, F., Comportamiento de los tejidos dentales y de algunos materiales de obturación dental sometidos a altas temperaturas con fines forenses, Colomb Med, 39(1), pp. 28-46, 2008.
[37] Holden, J.L., Phakey, P.P. and Clement, J.G., Scanning electron microscope observations of heat-treated human bone, Forensic Sci Int, 74, pp. 29-45, 1995. DOI: 10.1016/0379-0738(95)01735-2.
[38] Holden, J.L., Clement, J.G. and Phakey, P.P., Age and temperature related changes to the ultrastructure and composition of human bone mineral, J Bone Miner Res, 10(9), pp. 1400-1409, 1995. DOI: 10.1002/jbmr.5650100918.
[39] Venkatesan, J. and Kim, S-K., Effect of temperature on isolation and characterization of hydroxyapatite from tuna (Thunnus obesus) bone, Materials, 3(10), pp. 4761-4772, 2010. DOI: 10.3390/ma3104761.

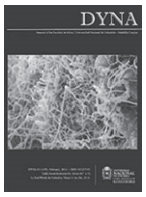
S. Medina, is a D.D.S. in the School of Dentistry at the Universidad del Valle (Cali, Colombia) and a Young Researcher - COLCIENCIAS 2012. Currently he is a Master's student in the School of Basic Sciences at the Universidad del Valle (Cali, Colombia) and a member of the Soft Tissue and Mineralized Research Group at the Universidad del Valle. His research interests include soft and mineralized biological tissues applied to health sciences.
ORCID: http://orcid.org/0000-0002-7000-3499

L. Salazar, has an MSc. in Basic Sciences from the Universidad del Valle, Cali, Colombia. She currently teaches in the School of Basic Sciences at the Universidad del Valle, Cali, Colombia and she is a director of the Soft Tissue and Mineralized Research Group at the Universidad del Valle. She has worked in programs and on projects regarding soft and mineralized biological tissues applied to health sciences.
ORCID: 0000-0002-3087-8493

C. Mejía, has an MSc. in Basic Sciences from the Universidad del Valle, Cali, Colombia. He currently teaches in the School of Dentistry at the Universidad del Valle (Cali, Colombia) and is a member of the Soft Tissue and Mineralized Research Group at the Universidad del Valle. His research interests include soft and mineralized biological tissues applied to health sciences.
ORCID: 0000-0003-2176-5435

F. Moreno, has an MSc. in biomedical basic sciences from the Universidad del Valle, Cali, Colombia. He currently teaches at the Basic Health Sciences Department in the Faculty of Health Sciences at the Pontificia Universidad Javeriana, Cali, Colombia and the School of Dentistry in the Universidad del Valle, Cali, Colombia. His research interests include soft and mineralized biological tissues applied to forensic sciences.
ORCID: 0000-0003-0394-9417

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS



DYNA
http://dyna.medellin.unal.edu.co/

Sparse representations of dynamic scenes for compressive spectral video sensing

Claudia V. Correa-Pugliese [a], Diana F. Galvis-Carreño [b] & Henry Arguello-Fuentes [c]

[a] *Department of Electrical and Computer Engineering, University of Delaware, Newark, DE, USA. clavicop@udel.edu*
[b] *Escuela de Ingeniería Química, Universidad Industrial de Santander, Bucaramanga, Colombia. diana.galvis1@correo.uis.edu.co*
[c] *Escuela de Ingeniería de Sistemas, Universidad Industrial de Santander, Bucaramanga, Colombia. henarfu@uis.edu.co*

Received: December 12th, 2014. Received in revised form: July 29th, 2015. Accepted: August 19th, 2015.

Abstract
The coded aperture snapshot spectral imager (CASSI) is an optical architecture that captures spectral images using compressive sensing. This system improves the sensing speed and reduces the large amount of collected data given by conventional spectral imaging systems. In several applications, it is necessary to analyze changes that occur between short periods of time. This paper first presents a sparsity analysis for spectral video signals, to obtain accurate approximations and better comply compressed sensing theory. The use of the CASSI system in compressive spectral video sensing then is proposed. The main goal of this approach is to capture the spatio-spectral information of dynamic scenes using a 2-dimensional set of projections. This application involves the use of a digital micro-mirror device that implements the traditional coded apertures used by CASSI. Simulations show that accurate reconstructions along the spatial, spectral and temporal axes are attained, with PSNR values of around 30 dB.

Keywords: spectral dynamic scenes, compressive spectral imaging, sparse representations, coded apertures, CASSI.

Representaciones dispersas de escenas dinámicas y reconstrucciones a partir de muestreo compresivo

Resumen
El sistema de adquisición de imágenes espectrales de apertura codificada (CASSI) es una arquitectura óptica que capta imágenes espectrales usando muestreo compresivo. Este sistema acelera la detección y reduce la gran cantidad de datos adquiridos por los sistemas tradicionales. En algunas aplicaciones es necesario analizar la variabilidad de la escena en períodos cortos de tiempo. Este trabajo presenta un análisis de las bases de representación para imágenes espectrales dinámicas, con el fin de obtener aproximaciones correctas a partir de su representación dispersa, y permitir la aplicación de muestreo compresivo. Posteriormente se propone el uso del sistema CASSI captar la información espacial y espectral de escenas dinámicas utilizando un conjunto de proyecciones bidimensionales. Esto implica el uso de un dispositivo de microespejos digitales que implementa las aperturas codificadas utilizadas en CASSI. Resultados muestran que es posible obtener reconstrucciones correctas en las dimensiones espaciales, espectral y temporal, con valores de PSNR alrededor de 30 dB.

Palabras clave: imágenes espectrales dinámicas, muestreo compresivo de imágenes multi-espectrales, representaciones dispersas, aperturas codificadas, CASSI.

1. Introduction

Traditional imaging architectures capture light intensity values on each spatial location and compression techniques are then used for data storage and transmission [1]. In contrast, spectral imaging provides light intensity values across a range of wavelengths. Thus, each spatial point of a spectral image provides a complete spectral signature of the composition of a scene. Conventional spectral imaging systems rely on Nyquist criterion to acquire the spatio-spectral information of an object or scene. These systems experience an extremely low sensing speed and, need to store large amounts of collected data, proportional to the desired resolution [2]. An alternative approach for spectral imaging

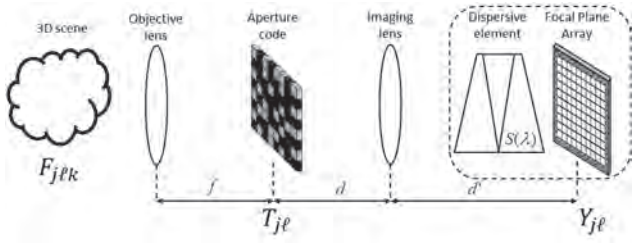


Figure 1. General CASSI Optical architecture.
Source: [2]

acquisition, known as Compressive Spectral Imaging (CSI), has recently emerged. CSI applies compressed sensing (CS) principles to capture and recover the spatial and spectral information of a scene in a single two-dimensional set of projections. In particular, CSI assumes that a spectral image $\mathbf{F} \in \mathbb{R}^{N\times N\times L}$, has a sparse representation in a basis $\mathbf{\Psi}$, such that it can be recovered from $v \ll N^2 L$ random projections [3]. Therefore, the selection of the sparse basis $\mathbf{\Psi}$ is critical to obtain good reconstruction results [4].

The coded aperture snapshot spectral imager (CASSI) shown in Fig. 1 is an optical architecture designed to capture CSI measurements [3,5]. The CASSI architecture comprises a set of lenses, a coded aperture, a dispersive element (commonly a prism), and a focal plane array (FPA) detector. Several variations of the CASSI system have been proposed to improve the quality of the obtained images. For instance, multiple shots can be attained by varying the coded aperture patterns, thus, more information about the scene is extracted [6-8]; an optimal coded aperture design for spectral selectivity has been proposed in [3]; a high resolution coded aperture and a low resolution FPA are used to obtain spatial super resolution in the CASSI system without incurring on expensive detectors [9]. Furthermore, spectral super resolution is attained by adding a second coded aperture [10]. Finally, traditional block-unblock coded apertures have recently been replaced by an array of optical filters [11].

In many applications such as surveillance, or some microscopic biological studies, the scenes under analysis are not completely static; conversely, many changes may occur between short periods of time. Thus, not only the spatial and spectral, but also temporal information is of high interest. For instance, hyperspectral video is used for object or human tracking [12-15], for cancer detection through endoscopy [16], bile duct inspection [17] and several types of surgery [18,19]. The acquisition of this four-dimensional information from a scene is known as spectral video sensing. Furthermore, when CS techniques are used to sense these video signals, it is known as compressive spectral video sensing. Previous works have proposed different video spectral acquisition approaches. For instance, in [20] different sets of spectral bands are measured on each video frame, and then a sparsity assumption is used to reconstruct the data. Since each frame does not contain information from all the spectral bands, this approach is not capable of capturing the variations that may occur on the spectral bands during the acquisition time. Other spectral video sensing approaches include multiple sensors to capture several video streams that are processed to obtain a single high-resolution signal [21], or dispersive elements in conjunction with occlusion masks to capture spectral information in a monochrome camera [22,23]. These approaches, however, do not employ CS theory. Moreover, an architecture named coded aperture compressive temporal imaging (CACTI) captures a single coded measurement by shifting a large coded aperture [24]; this coded measurement is then used to estimate several video frames, but no spectral information is taken into account. Similar spectral video sensing approaches can be found in [25-27]. CS concepts have been recently exploited in spectral video sensing, in particular, a recent variation of CACTI is the coded aperture compressive spectral-temporal imaging (CACSTI) [28, 29], which employs mechanical translation of a coded aperture and spectral dispersion to capture a multi-spectral dynamic scene onto a monochrome detector. Capturing information from all frames in a single snapshot however, leads to an extremely ill-posed reconstruction problem.

This paper presents a sparsity analysis of spectral video signals. These sparse representations can be exploited by using the CASSI system to capture the spatio-spectral information of dynamic scenes. In particular, this approach implements the coded aperture patterns using a digital micro-mirror device (DMD) that switches the patterns to independently encode the information from different frames. More specifically, the compressive spectral video problem can be expressed in the following ways: the input source is a four-dimensional array with two spatial, one spectral and, one temporal dimension. The physical phenomenon is mathematically described in the following way: the $\boldsymbol{m}-$th spectral video frame of the input source, $\boldsymbol{F_m} \in \mathbb{R}^{\boldsymbol{N\times N\times L}}$, is first spatially modulated by the coded aperture $\boldsymbol{T_m} \in \mathbb{R}^{\boldsymbol{N\times N}}$, where $\boldsymbol{m = 0, \ldots, D-1}$ indexes the temporal dimension; thus, a coded aperture pattern remains fixed to capture the information from each frame. Then, the dispersive element decomposes the encoded source frame into its spectral components. Finally, the encoded spatio-spectral information from a specific frame is integrated across the spectral components into the FPA, such that multiplexed spatio-spectral information is captured on each pixel. The output of the system for the $\boldsymbol{m}-$th frame can be modeled as $\boldsymbol{y_m = H_m f_m}$, where $\boldsymbol{f_m}$ is the vector form the video frame $\boldsymbol{F_m}$ and, $\boldsymbol{H_m}$ is the transfer function of the system that contains the effects of the coded aperture $\boldsymbol{T_m}$ and the prism. This procedure is repeated to capture information from a scene in different frames of time.

A variation of the CASSI system allows multiple snapshot acquisition of a spectral scene [2,6,8,30]. This modification results in better reconstruction quality. Using this multiple-shot scheme, several measurement sets are captured for each frame in the spectral video, using different coded aperture patterns. Different patterns can be implemented using DMD [7] or piezo-electric devices [8]. Thus, the measurements from K snapshots and D frames can be arranged as $\mathbf{y} = [(\mathbf{y}^0)^T \ldots (\mathbf{y}^{K-1})^T]^T$, where $\mathbf{y}^i = \left[\left(\mathbf{y}_0^i\right)^T \left(\mathbf{y}_1^i\right)^T \ldots \left(\mathbf{y}_{D-1}^i\right)^T\right]^T$, such that the sensing model can be rewritten as $\mathbf{y} = \mathbf{Hf}$, where $\mathbf{H}$ is the sensing matrix that contains all $\mathbf{H}_m$'s and $\mathbf{f}$ is the vector representation of the complete video data set $\mathbf{f} = \left[\mathbf{f}_0{}^T \mathbf{f}_1{}^T \ldots \mathbf{f}_{D-1}{}^T\right]^T$. In practice, the maximum number of measurements directly depends on both the pattern rate of the DMD and, the integration time of the detector. Most commercial DMDs have pattern rates of around 30 KHz, yet most CCD detectors can integrate 30

frames-per-second. In other words, a high-speed detector is a critical device in these kinds of applications.

The set of projections captured in the FPA, $\mathbf{y}$, is then used to recover the four-dimensional (spatio-spectral-temporal) input scene. The reconstruction is performed by solving an optimization problem that finds a sparse representation of the original data in a given basis. Commonly, the reconstruction problem is expressed as $\hat{\mathbf{f}} = \boldsymbol{\Psi}(\text{argmin}_{\boldsymbol{\theta}}\|\mathbf{y} - \mathbf{H}\boldsymbol{\Psi}\boldsymbol{\theta}\|_2 + \xi\|\boldsymbol{\theta}\|_1)$, where $\boldsymbol{\theta}$ is a sparse representation of $\mathbf{f}$ in the basis $\boldsymbol{\Psi}$, and ξ is a regularization constant.

This paper contains two major contributions; first, a sparsity analysis is developed in order to determine the basis that provides the sparsest representation of spectral video signals. Then, we present a mathematical model for the multi-shot CASSI system that can capture dynamic scenes using a two-dimensional set of projections. This paper is organized as follows: first, an introduction of sparse representation for dynamic scenes is presented; then, the mathematical model for compressive spectral imaging of spectral dynamic scenes is shown; finally, simulations and results to test this approach are included in section 4.

2. Sparse representation of spectral video signals

Compressed sensing exploits the fact that many signals are naturally sparse, or have a sparse representation on a given basis. In other words, this concept establishes that most of the energy from a signal is concentrated in either a small portion of its elements or its coefficients on a representation basis. Let $\mathbf{F} \in \mathbb{R}^{N\times N\times L\times D}$ be a spectral video with $N \times N$ pixels of spatial resolution, L spectral bands and D video frames. The vector form of $\mathbf{F}$, $\mathbf{f} \in \mathbb{R}^n$ with $n = N^2LD$, can be represented on the basis $\boldsymbol{\Psi} \in \mathbb{R}^{n\times n}$ as

$$\mathbf{f} = \boldsymbol{\Psi}\boldsymbol{\theta}\,, \tag{1}$$

where $\boldsymbol{\theta}$ is a sparse vector of coefficients.

In particular, CSI also relies on the sparsity nature of the data. Commonly, one representation basis is used for each dimension of a spectral image. Thus, four representation bases are used for spectral video signals, $\boldsymbol{\Psi}_1$ and $\boldsymbol{\Psi}_2$ for the spatial axes, $\boldsymbol{\Psi}_3$ for the spectral axis and, $\boldsymbol{\Psi}_4$ for the temporal coordinate. In general, if one frame of a video spectral signal is a common spectral image data cube, then it can be expressed as $\mathbf{f} = \boldsymbol{\Psi}_{3\text{D}}\boldsymbol{\theta}$, where $\boldsymbol{\Psi}_{3\text{D}} = \boldsymbol{\Psi}_1 \otimes \boldsymbol{\Psi}_2 \otimes \boldsymbol{\Psi}_3$ and, $\otimes$ denotes the kronecker product. Usually in spectral images, a 2D Wavelet transformation is used for the spatial dimensions $\boldsymbol{\Psi}_1 \otimes \boldsymbol{\Psi}_2$ and, the Discrete Cosine Transform (DCT) is used for the spectral dimension, $\boldsymbol{\Psi}_3$. Fig. 2 shows the sparse representation of one frame from a spectral video using three different Kronecker product bases. Fig. 2 (a) shows the 8 original spectral bands of the single frame, Fig. 2 (b) presents the spectral frame representation using a 1-dimensional Wavelet transformation, Fig. 2 (c) shows the frame representation in a 2-dimensional Wavelet basis and, Fig. 2 (d) shows the spectral frame representation in a three-dimensional basis obtained from the Kronecker product between a 2D Wavelet Symmlet 8 and a DCT bases. It can be noticed in Fig. 2 that the Kronecker product basis provides a sparser representation of the spectral frame. Thus, most of the energy from the signal is concentrated in fewer coefficients $\boldsymbol{\theta}$.

The effect of the different bases is illustrated in Fig. 3, where different approximations of one spectral frame are obtained by retaining only 1% of the sparse representation coefficients in a Wavelet 1D, Wavelet 2D and a Kronecker product bases. These approximations are obtained by expressing the signal in the corresponding representation bases, then the coefficients are sorted according to their magnitude and the smallest coefficients of the video frame in each basis are set to zero, while the 1% largest elements are preserved. A reconstruction is then obtained by applying the correspondent inverse transformation represented as $\boldsymbol{\Psi}$. It can be noticed in Fig. 3 that the approximation images show a great similarity with the original, especially when the Kronecker product basis is employed.

Previous works analyze the sparse representation of a single frame from a spectral video that can be seen as a static spectral image and, can be modeled using a three-dimensional basis, $\boldsymbol{\Psi}_{3\text{D}}$. However, appropriate sparse representations of the whole dynamic spectral scenes have not been yet considered in the literature. It has been previously shown that the three-dimensional basis provides the sparsest representation of the three-dimensional structure of a spectral image. Similarly, a four-dimensional basis ($\boldsymbol{\Psi}_{4\text{D}}$) exploits the sparsity of a dynamic spectral image, given that a single transformation is assumed for each coordinate of the signal. Thus, a dynamic spectral (four-dimensional) video can be mathematically represented as

$$\mathbf{f} = \boldsymbol{\Psi}_{4\text{D}}\boldsymbol{\theta} \tag{2}$$

where $\boldsymbol{\Psi}_{4\text{D}} = \boldsymbol{\Psi}_1 \otimes \boldsymbol{\Psi}_2 \otimes \boldsymbol{\Psi}_3 \otimes \boldsymbol{\Psi}_4$ and, $\{\boldsymbol{\Psi}_i\}_{i=1}^4$ is a set of different 1-dimensional transformations. An analysis of the representation bases applied to spectral video signals is presented in Section 4.

3. Compressive spectral imaging for spectral dynamic scenes

Compressive spectral imaging theory has previously been used to acquire spatial and spectral information of a scene. These previous optical architectures can be extended to the acquisition of dynamic spectral scenes, by exploiting the sparse basis discussed in the preceding section. In particular, the CASSI architecture presented in Fig. 1 can be employed to sense video spectral information. Fig. 4 shows the sensing process for a dynamic spectral scene.

Several measurement shots are usually captured in CSI, such that the captured projections extract most of the details in the scene, and thus the obtained reconstruction is more accurate. Furthermore, increasing the number of captured projections during a particular frame leads to a less ill-posed inverse problem. In particular, each additional measurement shot uses a different coded aperture for each frame, which remains fixed during the integration time of the detector. First, the mathematical model for a single shot is presented, and then a model for the multiple shot scheme is developed.

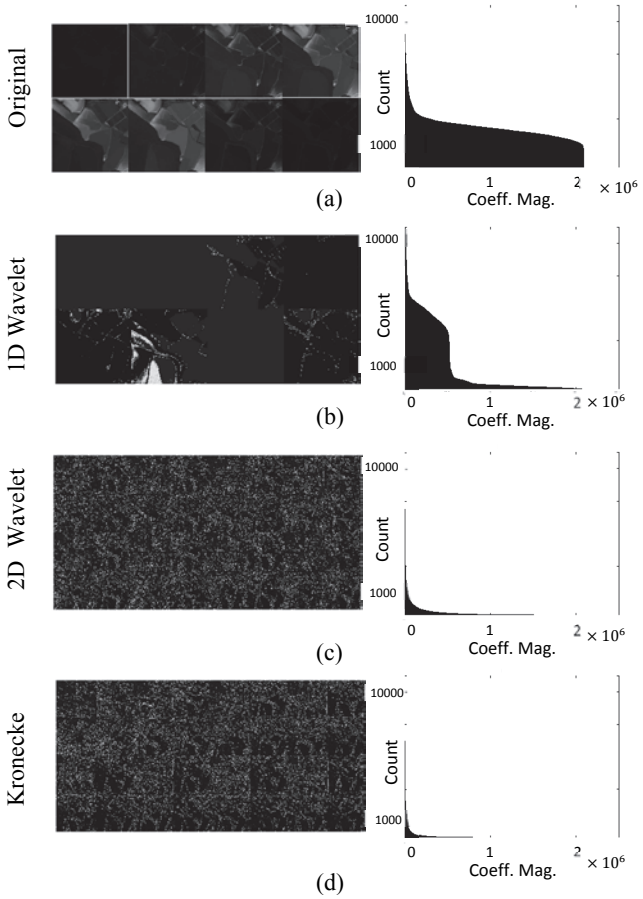


Figure 2. Sparse representation comparison between the (a) original video frame coefficients and its representation on the (b) one-dimensional Wavelet, (c) two-dimensional Wavelet and, (d) the Kronecker product basis between the 2D Wavelet and DCT.
Source: Authors.

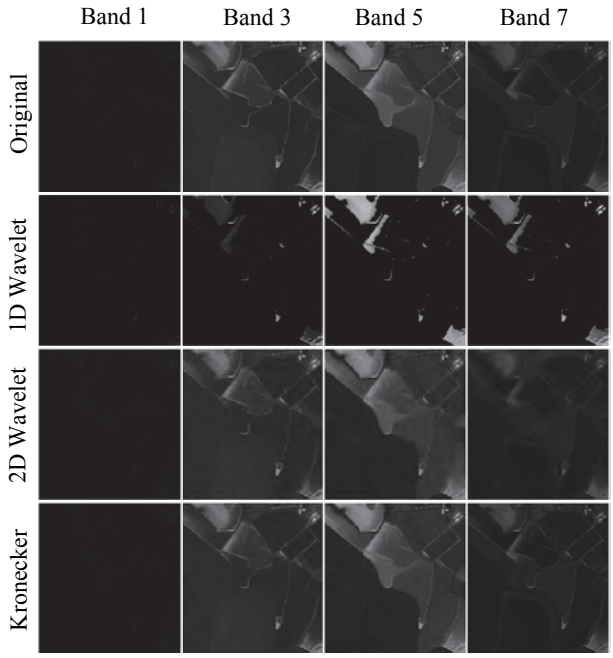


Figure 3. Sparse spectral frame representation using different bases. Selected spectral bands are represented using 1% of their sparse representation coefficients.
Source: Authors.

3.1. Single snapshot mathematical model

Let $f_0(x, y, \lambda, \tau)$ be a dynamic spectral source, where x, y index the spatial axes, λ is the index for the spectral dimension, and τ is the temporal/frame index. Each frame from the source is first spatially modulated by a time-dependent coded aperture $T(x, y, \tau)$. This coded aperture remains fixed for each frame during the integration time of each measurement shot. In other words, every frame from the scene is modulated by a different pattern in the coded aperture.

Then, the coded field correspondent to each frame is dispersed by a prism yielding $f_1(x, y, \lambda, \tau)$, as expressed in eq. (3)

$$f_1(x, y, \lambda, \tau) = \iint f_0(x, y, \lambda, \tau) T(x, y, \tau) h\big(x' - x - S(\lambda)\big) dx' dy' = f_0(x - S(\lambda), y, \lambda, \tau) T(x - S(\lambda), y, \tau) \quad (3)$$

where $S(\lambda)$ represents the dispersion function of the prism and, $h(\cdot)$ is the impulse response of the system. The output for the m-th frame, $\mathbf{Y}_m$ is obtained by integrating the field $f_1(x, y, \lambda, \tau)$ over the spectral range sensitivity of the camera, Λ, during the interval time $[m\Delta_t, (m+1)\Delta_t]$, where Δ_t is the integration time of the detector. Thus, the resulting field $Y_m(x, y)$ can be expressed as

$$\begin{aligned} Y_m(x, y) &= \int_{m\Delta_t}^{(m+1)\Delta_t} \int_{\Lambda} f_1(x, y, \lambda, \tau) d\lambda d\tau \\ &= \int_{m\Delta_t}^{(m+1)\Delta_t} \int_{\Lambda} f_0(x - S(\lambda), y, \lambda, \tau) \\ &\quad \times T(x - S(\lambda), y, \tau) d\lambda \, d\tau \end{aligned} \quad (4)$$

for $m = 0, \ldots, D-1$.

Since the detector is a pixelated array, the energy from the m-th frame that is captured in the $(j, \ell) -$ th pixel can be expressed as

$$(Y_m)_{j\ell} = \iint Y_m(x, y) p(j, \ell; x, y) dx dy \quad (5)$$

where $p(j, \ell; x, y) = \text{rect}\left(\frac{x}{\Delta} - j, \frac{y}{\Delta} - \ell\right)$ represents the rectangular pixel, with pixel size Δ. Similarly, the m-th coded aperture can be also discretized as $T_m(x, y) = \sum_{j,\ell} (T_m)_{j\ell} \text{rect}\left(\frac{x}{\Delta} - j, \frac{y}{\Delta} - \ell\right)$ and the discrete source can be represented as

$$F_{j\ell km} = \int_{m\Delta_\tau}^{(m+1)\Delta_t} \iiint f_0(x, y, \lambda, \tau) dx dy d\lambda d\tau \quad (6)$$

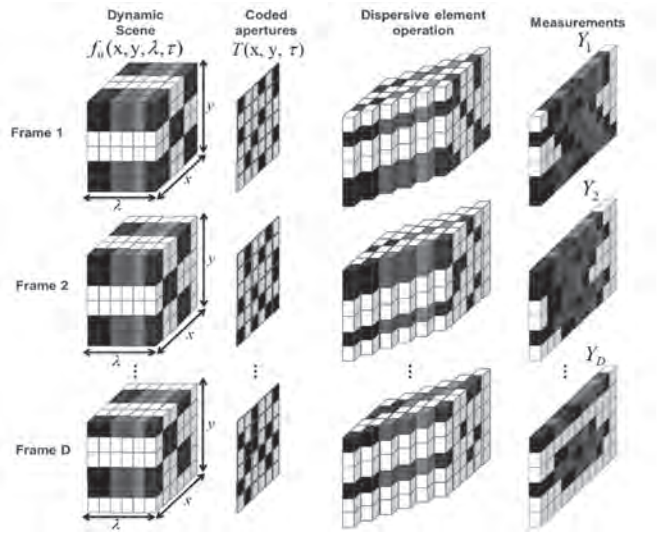


Figure 4. Process of CASSI imaging for a dynamic spectral scene with D frames. Each frame is spatially coded by a different coded aperture and then dispersed by the prism. Each detector pixel contains spectral information from several bands.
Source: Authors.

where $j, \ell = 0, \ldots, N-1$, index the spatial coordinates, $k = 0, \ldots, L-1$, indexes the spectral components, $m = 0, \ldots, D-1$, indexes the frames. This discretization yields a 4-dimensional representation of the dynamic scene, $\mathbf{F} \in \mathbb{R}^{N \times N \times L \times D}$, where $N \times N$ are the spatial dimensions, L is the number of spectral bands and, D is the number of frames. Using these discrete representations, the energy captured on the detector, that comes from the $m-$th frame, can be written as

$$(Y_m)_{j\ell} = \sum_k F_{j(\ell-k)km}(T_m)_{j(\ell-k)} + (\omega_m)_{j\ell} \qquad (7)$$

where the dispersion effect is represented by the shifting in the ℓ-axis and, ω_m is the noise in the system.

The measurement set acquired from a single frame, $\mathbf{Y}_m$, can be represented in vector form as $\mathbf{y}_m$. Similarly, the spatio-spectral source $\mathbf{F}$ can be expressed in vector form as $\in \mathbb{R}^{N^2LD}$, and the relation between the $m-$th source frame and its correspondent measurement set is given by

$$\mathbf{y}_m = \mathbf{H}_m \mathbf{f}_m + \boldsymbol{\omega}_m \qquad (8)$$

where $\mathbf{f}_m$ is the vector representation of the m-th frame and, $\mathbf{H}_m$ is the single-shot CASSI sensing matrix that accounts for the effects of the coded aperture pattern $\mathbf{T}_m$ and the dispersive element. Furthermore, measurements acquired from different frames can also be arranged in a single vector, $\mathbf{y} = [\mathbf{y}_0{}^T \mathbf{y}_1{}^T \ldots \mathbf{y}_{D-1}{}^T]^T$, where $\mathbf{y}_m{}^T$ is the vector representation of the measurement corresponding to the $m-th$ frame. Thus, the system can be modeled in matrix form as follows

$$\mathbf{y} = \mathbf{H}\mathbf{f} + \boldsymbol{\omega}, \qquad (9)$$

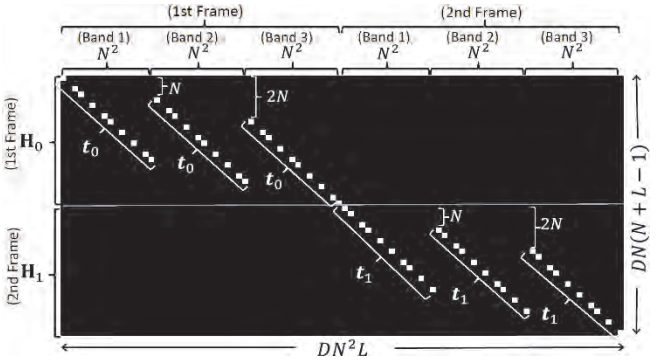


Figure 5. Single shot sensing matrix example for N=4, L=3 and, D=2. Notice that t_0 and t_1 are the vector representations of $T_0$and T_1, respectively.
Source: Authors.

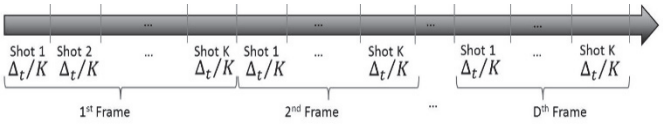


Figure 6. Multiple shot acquisition for a dynamic spectral scene. Each measurement shot has a duration Δ_t/K. K shots are captured for each frame and a detector with integration time Δ_t/K is assumed. Each frame snapshot uses a different coded aperture.
Source: Authors.

where $\mathbf{H} \in \mathbb{R}^{DN(N+L-1) \times DN^2L}$ is the single-shot sensing matrix for the complete dynamic scene. This matrix groups the matrices for all frames as the matrix given by $\mathbf{H} = \text{diag}(\mathbf{H}_0 \mathbf{H}_1 \ldots \mathbf{H}_{D-1})$. Fig. 5 shows an example of the structure of the sensing matrix $\mathbf{H}$, in which the white points correspond to the non-zero elements of the matrices $\mathbf{H}_m$ and, are determined by the coded aperture patterns used for each frame.

3.2. Multiple snapshot mathematical model

In general, a single snapshot in CASSI allows the underlying data cube to be reconstructed. However, multiple snapshots using different coded aperture patterns yield a less ill-posed inverse problem, and better quality reconstructions.

Similarly, several measurement shots can be captured for each single source frame. To this end, the duration of the frame is seen as a set of smaller time intervals, in which the coded aperture pattern is shuffled and, the detector captures a new set of compressive measurements each time. Thus, each measurement shot has duration of Δ_t/K time units, and K measurement shots are captured for each frame. Fig. 6 presents a timeline that illustrates this concept. It can be noticed that a detector with integration time Δ_t/K is assumed.

Consequently, eq. (8) can be rewritten to index the measurement shots. Thus, the i-th shot correspondent to the m-th frame is expressed as

$$\mathbf{y}_m^i = \mathbf{H}_m^i \mathbf{f}_m + \boldsymbol{\omega}_m^i \qquad (10)$$

for $i = 0, \ldots, K-1$. Here, $\mathbf{H}_m^i$ represents the sensing matrix and corresponds to the i-th shot for the m-th frame.

Similarly, all the measurement shots captured for a single frame can be arranged as $\mathbf{y}_m = [(\mathbf{y}_m^0)^T (\mathbf{y}_m^1)^T \dots (\mathbf{y}_m^{K-1})^T]^T$ such that the multi-shot sensing approach can be expressed as in eq. (9) with $\mathbf{y} = [(\mathbf{y}_0)^T \dots (\mathbf{y}_{D-1})^T]^T$. However, $\mathbf{H} \in \mathbb{R}^{KDN(N+L-1)\times DN^2L}$ in this case is the sensing matrix that is associated with the full data using K measurement shots, and is given by the expression

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_0^0 & & & & \\ \mathbf{H}_0^1 & & & & \\ \vdots & & & & \\ \mathbf{H}_0^{K-1} & & & & \\ & \mathbf{H}_1^0 & & & \\ & \vdots & & & \\ & \mathbf{H}_1^{K-1} & & & \\ & \vdots & \ddots & & \\ & & & \mathbf{H}_{D-1}^0 & \\ & & & \vdots & \\ & & & \mathbf{H}_{D-1}^{K-1} \end{bmatrix}. \quad (11)$$

Fig. 7 shows an example of this matrix for $N = 6$, $L = 3$ spectral bands, $D = 2$ frames and, $K = 2$ shots. The upper half of this matrix corresponds to the first frame and the lower half matrix accounts for the second frame. As in Fig. 5, each diagonal stands for a spectral band.

The set of measurements $\mathbf{y}$ is then used to obtain a reconstruction of the underlying 4-dimensional data. This reconstruction is attained by solving the inverse problem $\hat{\mathbf{f}} = \boldsymbol{\Psi}_{4D}(\text{argmin}_{\boldsymbol{\theta}}\|\mathbf{y} - \mathbf{H}\boldsymbol{\Psi}_{4D}\boldsymbol{\theta}\|_2 + \xi\|\boldsymbol{\theta}\|_1)$, where ξ is a regularization constant, $\mathbf{H}$ is the sensing matrix in eq. (11) and, $\boldsymbol{\theta}$ is a sparse representation of $\mathbf{f}$ on the basis $\boldsymbol{\Psi}_{4D}$.

4. Simulations and Results

Simulations were performed in order to first determine the basis that provides the sparsest representation of dynamic spectral images and, second to test the model to sense and recover these types of images using CSI. All the simulations used a test data base composed by 16 frames, each of them with 8 spectral bands and 128×128 pixels of spatial resolution [20]. An RGB false color representation of the frames in this data base is presented in Fig. 8. In addition, the spectral responses of a specific point in the scene over time are depicted in Fig. 9. Random coded aperture patterns were used for all the experiments, in particular the entries of these patterns are realizations of a Bernoulli random variable with parameter $p = 0.5$. All simulations were conducted using an Intel Core i7 3.6 GHz processor and, 64 GB RAM memory.

4.1. Sparse representations

Using eq. (2), different combinations of bases were tested for dynamic spectral scene representation. Previous results show that a Kronecker product between two-dimensional Wavelet Symmlet 8 and DCT bases provides a good sparse representation of spectral images [3,10]. Taking this into account, simulation results are presented for four combinations of Wavelet Symmlet 8 and DCT bases applied to the four dimensions of the test spectral video. More specifically, the kronecker product bases presented in

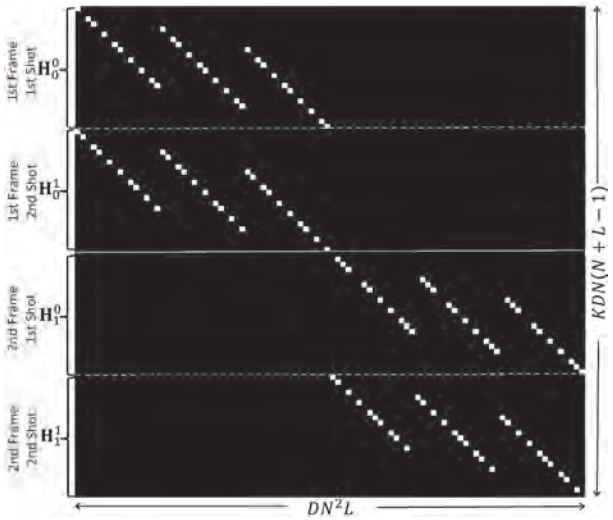


Figure 7. Multi-shot sensing matrix example for N=4, L=3, D=2 and, K=2. Source: Authors.

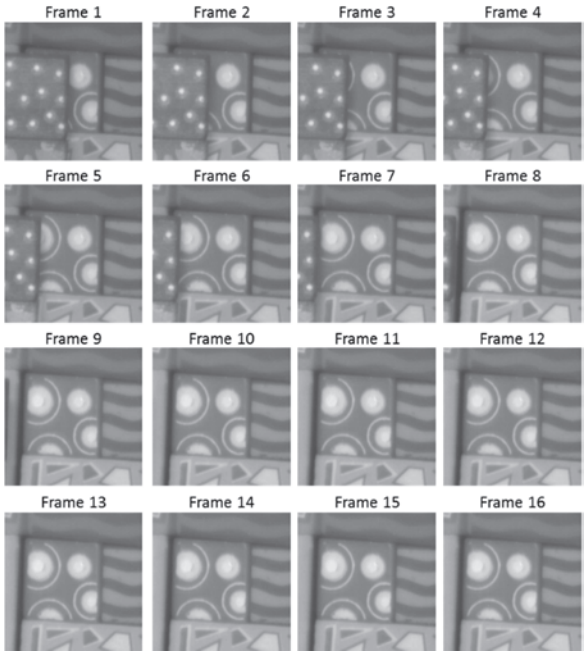


Figure 8. RGB representation of the 16 frames in the test data base. Each frame has 128x128 pixels of spatial resolution and 8 spectral bands. Source: Authors.

Table 1 were tested.

Table 1.
Kronecker product bases used for simulations. Bases' names consist of four letters; the first two represent the bases for the spatial dimensions, the third corresponds to the spectral dimension and the last one accounts for the temporal dimension. W: Wavelet, D: DCT.

Basis Name	**Spatial** $\boldsymbol{\Psi_1 \otimes \Psi_2}$	**Spectral** $\boldsymbol{\Psi_3}$	**Temporal** $\boldsymbol{\Psi_4}$
WWDD	2D-Wavelet	DCT	DCT
WWWW	2D-Wavelet	1D-Wavelet	1D-Wavelet
WWWD	2D-Wavelet	1D-Wavelet	DCT
WWDW	2D-Wavelet	DCT	1D-Wavelet

Source: Authors.

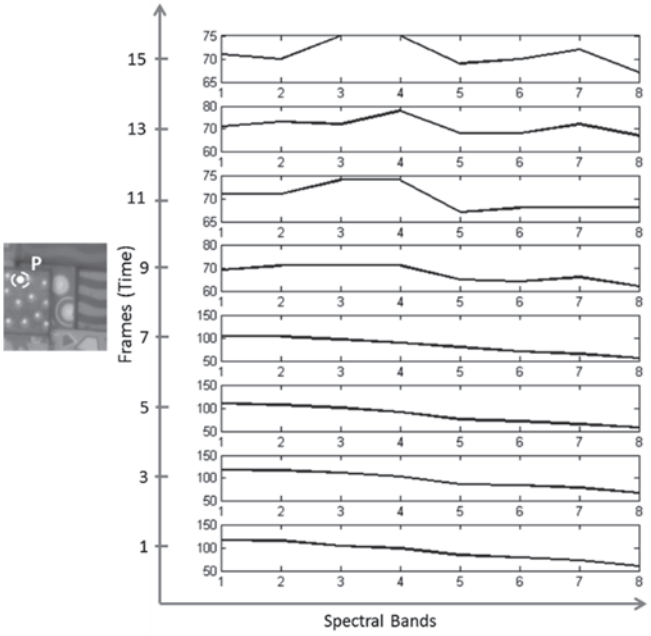


Figure 9. Spectral responses for different frames, of the indicated spatial point P.
Source: Authors.

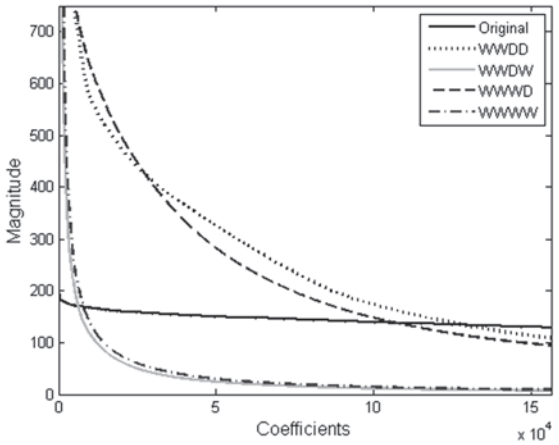


Figure 10. Kronecker sparse bases representation of a test data base for representation in Table 1.
Source: Authors.

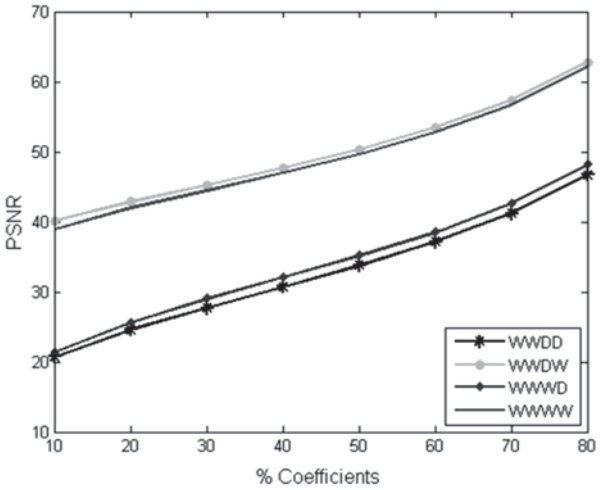


Figure 11. PSNR Representation as a function of the percentage of coefficients used to approximate the data base.
Source: Authors.

Fig. 10 shows the coefficients of the test data base on each basis from Table 1. It can be noticed that the bases WWWW and WWDW provide similar results, as do WWDD and WWWW. However, WWWW and WWDW coefficients experience a more pronounced decay, which indicates that these bases provide the sparsest representations.

The effect of using different bases can be also illustrated by obtaining an approximation of the original data base. This process consists of setting the smallest absolute value coefficients in the basis Ψ to zero, while a percentage of the largest coefficients are preserved and, the reconstruction is obtained applying the inverse transformation.

Fig. 11 shows the Peak Signal-to-Noise Ratio (PSNR) as a function of the percentage of coefficients used to approximate the underlying signal. It can be seen that the best PSNR results are obtained from the sparsest representations; the WWWW and WWDW bases improve the results by up to 30 dB. A comparison of the representations obtained from the different bases, using just the 10% largest coefficients, is shown in Fig. 12(a).

These approximations correspond to a portion of the fourth spectral band from the first frame. As previously mentioned, WWWW and WWDW bases provide accurate quality representations, while objects in the results from the other bases are hardly visible. Similarly, Fig. 12(b) presents the representations obtained from the 50% largest coefficients. It can be seen that a clearer approximation is obtained for all bases. However, the WWWW and WWDW bases still provide better results. In addition, the spectral and temporal approximations for two spatial points of the scene are illustrated in Figs. 13, 14, respectively. These figures demonstrate that the WWDW and WWWW bases provide the most accurate representations of the spectral video signal.

4.2. Reconstruction of dynamic spectral scenes

Several measurement shots were simulated to test the model presented in eq. (9) and eq. (11). In these cases, WWDW and WWWW, the representation bases that provide the sparsest approximations of the scene were used.

The procedure followed in this experiment consists of simulating the measurement set using the multi-shot model described in section 3.1. Then, the measurement set is used as the input of a compressed sensing reconstruction algorithm to obtain an approximation of the original scene. Specifically, the GPSR algorithm was used to solve the inverse problem [31].

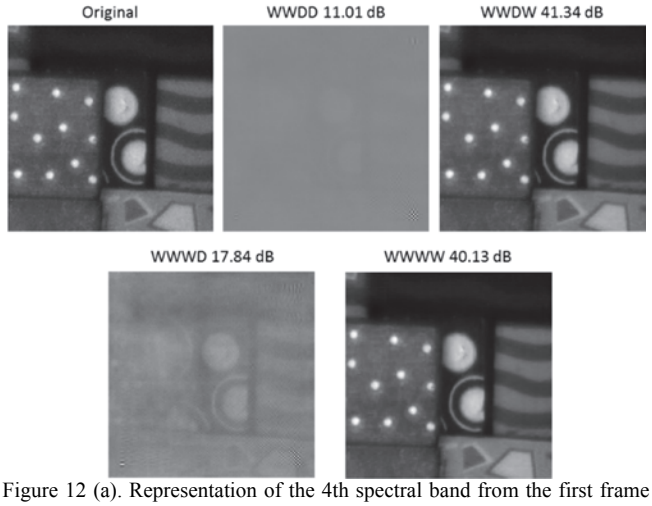


Figure 12 (a). Representation of the 4th spectral band from the first frame using inverse transformation from the 10% largest coefficients on each basis. The PSNR is indicated.
Source: Authors.

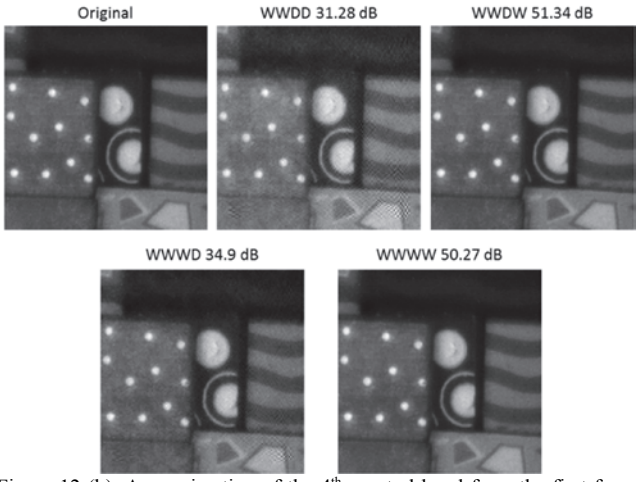


Figure 12 (b). Approximation of the 4th spectral band from the first frame using inverse transformation from the 50% largest coefficients on each basis. The PSNR is indicated.
Source: Authors.

Fig. 15 shows the reconstruction PSNR as a function of the number of measurement shots per frame, K, used to obtain the reconstruction of the scene with 16 frames with 128×128 pixels and 8 spectral bands. The PSNR values are calculated as the average of the PSNR for all the spectral bands and frames. It can be seen that for both representation bases, increasing the number of shots per frame leads to a higher PSNR value. However, the WWDW basis provides a slightly better PSNR value.

Fig. 16 shows the reconstruction of one spectral band obtained from different frames, using both representation bases. In general, this figure shows that both bases provide visually accurate reconstructions.

The performance of the multi-shot model can be demonstrated by comparing the spectral response of a specific point in the original scene with its correspondent reconstruction. Fig. 14 presents this comparison for three

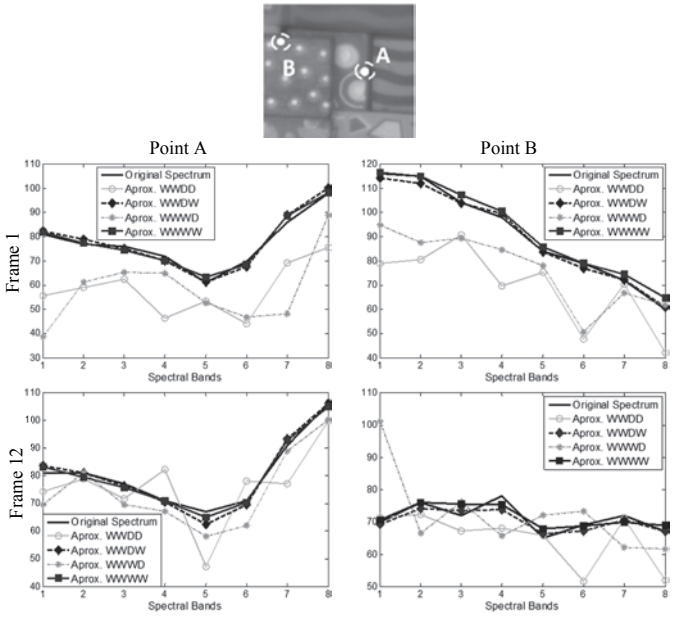


Figure 13 Spectral approximations of two spatial points and two frames using inverse transformation from the 10% largest coefficients on each basis.
Source: Authors.

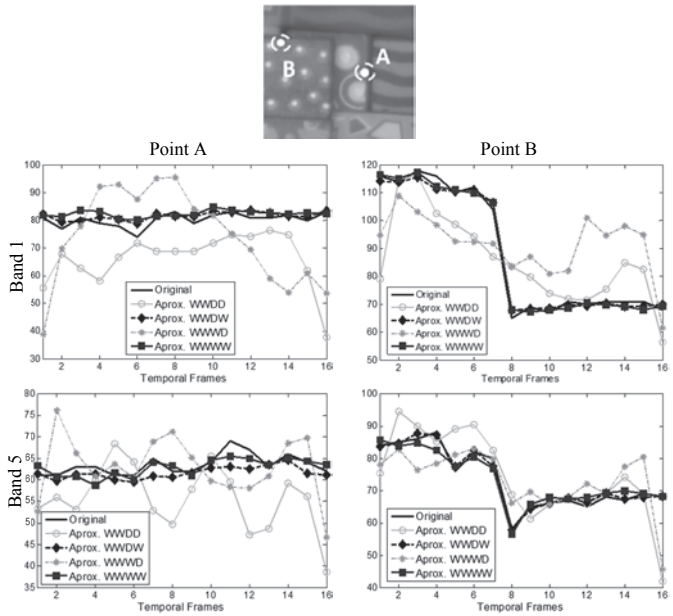


Figure 14. Temporal approximations of two spatial points and two spectral bands using inverse transformation from the 10% largest coefficients on each basis.
Source: Authors.

spatial points as indicated. Specifically, the spectral responses for these points measured in two different frames are shown. These results were obtained using the WWDW representation basis and $K = 3$ measurement shots per frame. Fig. 17 shows that this model provides an accurate spectral reconstruction. The false color representation of frame 1 intends to show the spatial location of the selected points.

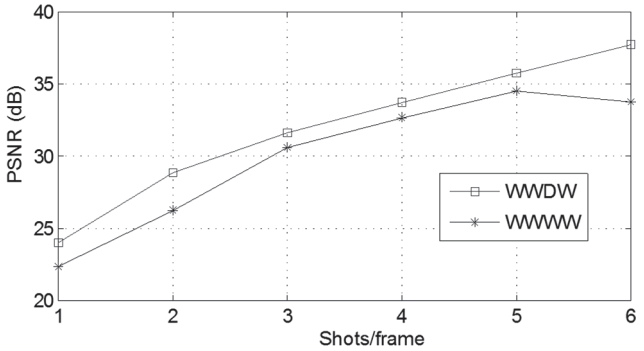


Figure 15. Average reconstruction PSNR as a function of the number of measurement shots used on each frame. The two bases from Section 4.1 that provide the sparsest representation were used. Source: Authors.

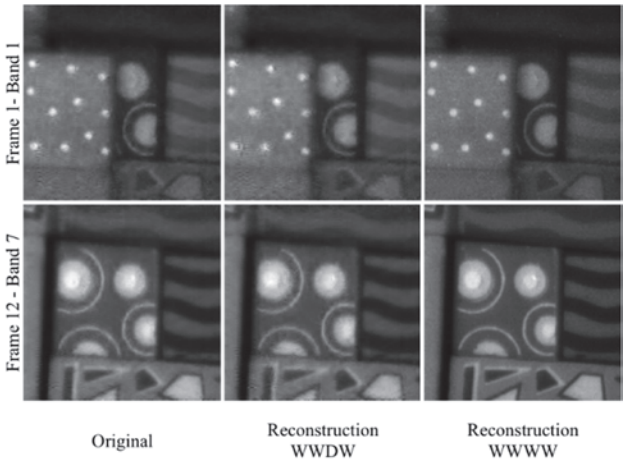


Figure 16. Reconstructions of the test data base using the WWDW and WWWW representation bases and $K = 4$ measurement shots per frame. Two spectral bands from two different frames are shown.
Source: Authors.

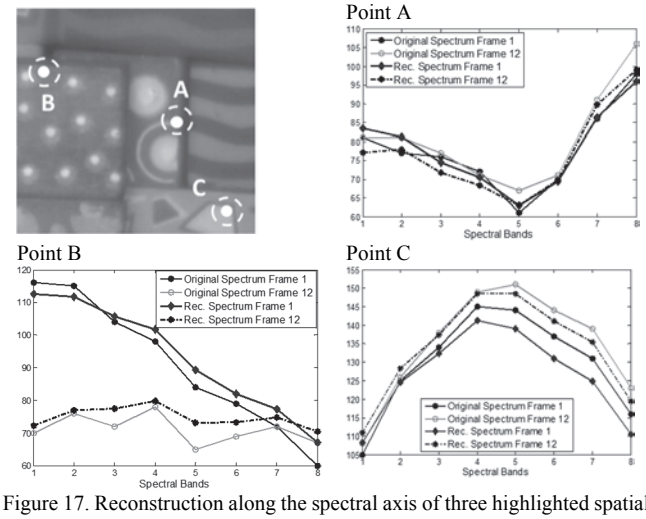


Figure 17. Reconstruction along the spectral axis of three highlighted spatial points from two different frames using the WWDW representation basis. The false color representation of frame 1 intends to show the spatial location of the selected points.
Source: Authors.

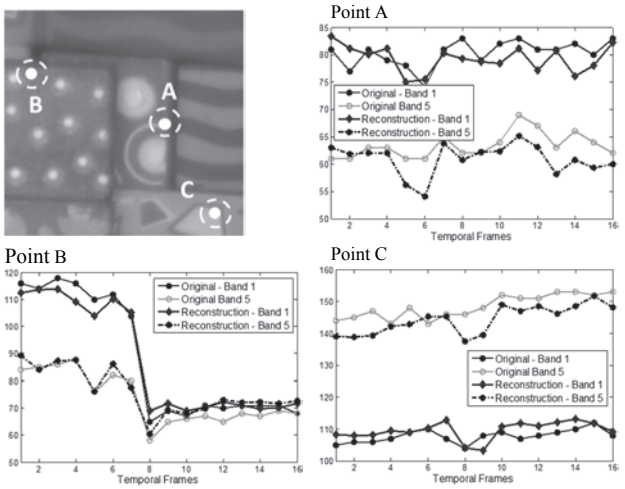


Figure 18. Reconstruction along the temporal axis of three highlighted spatial points from the first and fifth spectral bands using the WWDW representation basis.
Source: Authors.

Similarly, a different strategy to show the accuracy of the model is to compare the behavior of the original scene measured at a specific spatial point and spectral band over time with the correspondent reconstruction. Fig. 18 shows the results for three points in the first and fifth spectral bands, as indicated. These results show that the reconstructions obtained are close representations of the original dynamic spectral scene.

5. Conclusions

A mathematical model for sparse representations of dynamic scenes in compressive spectral video sensing has been presented. Experiments show that the WWDW and WWWW bases provide the sparsest representations of these types of signals. A variation of the CASSI system for compressive spectral video sensing has been also presented. The mathematical models for single-frame and multi-frame capture with the CASSI system have been proposed. Simulation results show the accuracy of the model in spatial, spectral and temporal reconstructions. In general, reconstruction PSNR values of around 30 dB were obtained with the proposed model.

Acknowledgements

The authors gratefully acknowledge the Vicerrectoría de Investigación y Extensión at the Universidad Industrial de Santander and, the University of Delaware for supporting this work registered under the project title "Optimal design of coded apertures for compressive spectral imaging", VIE code 1368.

References

[1] Sarinova, A., Zamyatin, A. and Cabal, P., Lossless compression of hyperspectral images with pre-byte processing and intra-bands correlation. DYNA, 82(190), pp. 166-172, 2015. DOI: 10.15446/dyna.v82n190.43723

[2] Arce, G.R., Brady, D.J., Carin, L., Arguello, H. and Kittle, D., An introduction to compressive coded aperture spectral imaging, IEEE

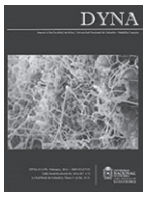
Signal Processing Magazine, 31(1), pp. 105-115, 2014. DOI: 10.1109/MSP.2013.2278763
[3] Arguello, H. and Arce, G.R., Rank minimization code aperture design for spectrally selective compressive imaging, IEEE Trans. Image Processing, 22(3), pp. 941-954, 2013. DOI: 10.1109/TIP.2012.2222899
[4] Candes, E.J. and Wakin, M.B., An introduction to compressive sampling, IEEE Signal Processing Magazine, 25(2), pp. 21-30, 2008. DOI: 10.1109/MSP.2007.914731
[5] Wagadarikar, A.A., John, R., Willet, R. and Brady, D.J., Single disperser design for coded aperture snapshot spectral imaging, Applied Optics, 47(10), pp. B44-B51, 2008. DOI: 10.1364/AO.47.000B44
[6] Arguello, H. and Arce, G.R., Code aperture optimization for spectrally agile compressive imaging, Journal Optical Society of America A, 28(11), pp. 2400-2413, 2011. DOI: 10.1364/JOSAA.28.002400
[7] Wu, Y., Mirza, I.O., Arce, G.R. and Prather, D., Development of a digitalmicromirror- device-based multishot snapshot spectral imaging system, Optics Letters, 36(14), pp. 2692-2694, 2011. DOI: 10.1364/OL.36.002692
[8] Kittle, D., Choi, K., Wagadarikar, A.A. and Brady, D.J., Multiframe image estimation for coded aperture snapshot spectral imagers, Applied Optics, 49(36), pp. 6824-6833, 2010. DOI: 10.1364/AO.49.006824
[9] Rueda, H. and Arguello, H., Spatial super-resolution in coded aperture-based optical compressive hyperspectral imaging systems, Revista Facultad de Ingeniería, 67, pp. 7-18, 2013.
[10] Rueda, H., Arguello, H. and Arce, G.R., On super-resolved coded aperture spectral imaging. SPIE Conference on Defense, Security and Sensing, Baltimore, MD, USA, 2013. DOI: 10.1117/12.2015855
[11] Arguello, H. and Arce, G.R., Colored coded aperture design by concentration of measure in compressive spectral imaging, IEEE Trans. on Image Processing, 23(4), pp. 1896-1908, 2014. DOI: 10.1109/TIP.2014.2310125
[12] Cheng, S.Y., Park, S. and Trivedi, M.M., Multi-spectral and multi-perspective video arrays for driver body tracking and activity analysis, Comput. Vis. Image Underst., 106(2–3), pp. 245-257, 2007. DOI: 10.1016/j.cviu.2006.08.010
[13] Van-Nguyen, H., Banerjee, A. and Chellappa, R., Tracking via object reflectance using a hyperspectral video camera, in 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, CVPRW 2010, pp. 44-51. 22, 2010. DOI: 10.1109/CVPRW.2010.5543780
[14] Banerjee, A., Burlina, P. and Broadwater, J., Hyperspectral video for illumination-invariant tracking, in WHISPERS '09 - 1st Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, 2009. DOI: 10.1109/WHISPERS.2009.5289103
[15] Duran, O. and Petrou, M., Subpixel temporal spectral imaging, Pattern Recognition Letters, 48, pp. 15-23, 2014. DOI: 10.1016/j.patrec.2014.04.005
[16] Leitner, R., De-Biasio, M., Arnold, T., Dinh, C.V., Loog, M. and Duin, R.P.W., Multi-spectral video endoscopy system for the detection of cancerous tissue, Pattern Recognition Letters, 34(1), pp. 85-93, 2013. DOI: 10.1016/j.patrec.2012.07.020
[17] Zuzak, K.J., Naik, S.C., Alexandrakis, G., Hawkins, D., Behbehani, K. and Livingston, E., Intraoperative bile duct visualization using near-infrared hyperspectral video imaging, Am. J. Surg., 195(4), pp. 491-497, 2008. DOI: 10.1016/j.amjsurg.2007.05.044
[18] Arnold, T., De Biasio, M. and Leitner, R., Hyper-spectral video endoscopy system for intra-surgery tissue classification, in Proceedings of the International Conference on Sensing Technology, ICST, pp. 145-150, 2013. DOI: 10.1109/ICSensT.2013.6727632
[19] Yi, D., Kong, L., Wang, F., Liu, F., Sprigle, S. and Adibi, A., Instrument an off-shelf CCD imaging sensor into a handheld multispectral video camera, Photonics Technology Letters, IEEE, 23(10), pp. 606-608, 2011. DOI: 10.1109/LPT.2011.2116153
[20] Mian, A. and Hartley, R., Hyperspectral video restoration using optical flow and sparse coding, Optics Express, 20(10), pp. 10658-10673, 2012. DOI: 10.1364/OE.20.010658
[21] Cao, X., Tong, X., Dai, Q. and Lin, S., High resolution multispectral video capture with a hybrid camera system, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011, pp. 297-304, 2011. DOI: 10.1109/CVPR.2011.5995418
[22] Du, H., Tong, X., Cao, X. and Lin, S., A prism-based system for multispectral video acquisition, 2009 IEEE 12th International Conference on Computer Vision, pp. 175-182, 2009. DOI: 10.1109/ICCV.2009.5459162
[23] Cao, X., Du, H., Tong, X., Dai, Q. and Lin, S., A prism-mask system for multispectral video acquisition, IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(12), pp. 2423-2435, 2011. DOI: 10.1109/TPAMI.2011.80
[24] Llull, P., Liao, X., Yuan, X., Yang, J., Kittle, D., Carin, L., Sapiro, G. and Brady, D.J., Coded aperture compressive temporal imaging, Optics Express, 21(9), pp. 10526-10545, 2013. DOI: 10.1364/OE.21.010526
[25] Xu, L., Sankaranarayanan, A., Studer, C., Li, Y., Baraniuk, R.G. and Kelly, K.F., Multi-scale compressive video acquisition, in Imaging and Applied Optics, OSA Technical Digest, 2013. DOI: 10.1364/COSI.2013.CW2C.4
[26] Llull, P., Liao, X., Yuan, X., Yang, J., Kittle, D., Carin, L., Sapiro, G. and Brady, D.J., Compressive sensing for video using a passive coding element, in Imaging and Applied Optics, OSA Technical Digest, 2013. DOI: 10.1364/COSI.2013.CM1C.3
[27] Koller, R., Schmid, L., Matsuda, N., Niederberger, T., Spinoulas, L., Cossairt, O., Schuster, G. and Katsaggelos, A.K., High spatio-temporal resolution video with compressed sensing, Opt. Express, 23(12), pp. 15992-16007, 2015. DOI: 10.1364/OE.23.015992
[28] Tsai, T., Llull, P., Carin, L. and Brady, D.J., Spectral-temporal compressive imaging, Optics Letters, 40(17), pp. 4054-4057, 2015. DOI: 10.1364/OL.40.004054
[29] Tsai, T., Llull, P., Yuan, X., Carin, L. and Brady, D.J., Coded aperture compressive spectral-temporal imaging, Imaging and Applied Optics 2015, OSA Technical Digest, 2015. DOI: 10.1364/COSI.2015.CTh2E.5
[30] Galvis-Carreño, D., Mejia-Melgarejo, Y. and Arguello-Fuentes, H., Efficient reconstruction of Raman spectroscopy imaging based on compressive sensing. DYNA, 81(188), pp. 116-124, 2014. DOI: 10.15446/dyna.v81n188.41162
[31] Figueiredo, M., Nowak, R. and Wright, S., Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems, IEEE Journal in Selected Topics in Signal Processing, 1(4), pp. 586-597, 2007. DOI: 10.1109/JSTSP.2007.910281

C.V. Correa-Pugliese, received her BSc. Eng. in Computer Science in 2009, her MSc. in Systems Engineering in 2013, both from the Universidad Industrial de Santander (UIS), Colombia. She received her MSc. degree in Electrical Engineering from the University of Delaware in 2013. She is currently a PhD candidate in the Electrical and Computer Engineering Department at the University of Delaware, USA. Her research interests include compressive spectral imaging, computational imaging, and compressed sensing.
ORCID: 0000-0002-1812-287X.

D.F. Galvis-Carreño, received her BSc. Eng. in Chemical Engineering in 2011 from the Universidad Industrial de Santander (UIS), Colombia. She is currently pursuing her MSc. in Chemical Engineering at UIS. Her main research areas include compressive raman spectroscopy, compressed sensing and, image processing.
ORCID: 0000-0002-0392-1281.

H. Arguello-Fuentes, received his BSc. Eng. Electrical Engineering in 2000, his MSc. in Electrical Power in 2003, both from the Universidad Industrial de Santander (UIS), Colombia. He received his PhD in Electrical Engineering from the University of Delaware, USA in 2013. He is an associate professor in the Department of Systems Engineering at the Universidad Industrial de Santander, Colombia. His research interests include high-dimensional signal processing, optical imaging, compressed sensing, hyperspectral imaging, and computational imaging.
ORCID: 0000-0002-2202-253X.

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS



DYNA
http://dyna.medellin.unal.edu.co/

Influence of demand, control and social support on job stress. Analysis by employment status from the V European working conditions survey

Susana García-Herrero [a], Miguel Ángel Mariscal-Saldaña [b], Eva María López-Perea [c] & Martha Felicitas Quiroz-Flores [d]

[a] *Escuela Politécnica Superior, Universidad de Burgos, Burgos, España. susanagh@ubu.es*
[b] *Escuela Politécnica Superior, Universidad de Burgos, Burgos, España. mariscal@ubu.es*
[c] *Facultad de Educación, Universidad de Burgos, Burgos, España. emlperea@ubu.es*
d Instituto Tecnológico de Toluca, México, México. mquiroz@beca.ubu.es

Received: December 16th, de 2014. Received in revised form: August 5th, 2015. Accepted: September 4th, 2015

Abstract
Work stress increasingly affects many workers from different countries. Conditions such as high demand, low social support and low job control are considered predictors of increased stress. With data obtained from the V European Working Conditions Survey (EWCS) a Bayesian network model was made. It provides information on the levels of stress in relation to model demand-control-social support (DCS), differentiating into work situations as they are, self-employed, private and public. To deepen understanding of the interrelationships between these variables sensitivity analysis of individual and overall were performed to check the DCS model assumptions. This model applied in the V EWCS identified the variations and similarities between different work situations, proving that having low levels of demand, together with control and high social support, the likelihood of stress decreases.

Keywords: stress; demand; control; social support; bayesian networks.

Influencia de la demanda, control y apoyo social sobre el estrés laboral. Análisis por situación laboral de la V Encuesta Europea sobre condiciones de trabajo

Resumen
El estrés laboral afecta cada vez en mayor cantidad a trabajadores de diversos países. Condiciones como la alta demanda, bajo apoyo social y bajo control sobre el trabajo se consideran predictores del aumento de estrés. Con datos obtenidos de la V Encuesta Europea sobre Condiciones de trabajo (EWCS), se planteó un modelo de red bayesiana que proporciona información sobre los niveles de estrés en relación al modelo demanda-control-apoyo social (DCS), diferenciado en situaciones laborales como son, autónomo, privado y público. Para profundizar en las interrelaciones existentes entre dichas variables se realizaron análisis de sensibilidad individuales y en conjunto para comprobar las hipótesis del modelo DCS. Este modelo aplicado en la V EWCS permitió identificar las variaciones y similitudes entre las diferentes situaciones laborales, comprobando que al tener niveles bajos de demanda, en conjunto con control y apoyo social alto, la probabilidad de sufrir estrés disminuye.

Palabras clave: estrés; demanda; control; apoyo social; redes bayesianas.

1. Introducción

Esta investigación analiza los niveles de estrés que se tienen al variar el nivel de control, demanda y apoyo social.

En muchos países se considera que el estrés va en incremento. Melchior et al [1] menciona que en los EE.UU. y la UE, más de la tercera parte de los trabajadores son expuestos al estrés laboral y que esta cantidad ha aumentado

desde la década de los noventa. Además de que el estrés laboral se considera causa de enfermedades profesionales, generando consecuencias para el trabajador y el lugar de trabajo.

Para autores como Law et al [2], el estrés laboral es resultante de la exposición prolongada a riesgos psicosociales (aspectos del ambiente laboral, diseño del trabajo, y organización de la administración) que son causas potenciales de daño psicológico y social, mientras que para Noblet and Lamontagne [3], el estrés laboral se ocasiona cuando las demandas y las condiciones externas no coinciden con las necesidades, expectativas o ideales de las personas, al exceder su capacidad física, habilidades o conocimientos para manipular adecuadamente cierta situación.

Actualmente existen diversas investigaciones que muestran los efectos adversos del estrés en las personas, como ejemplo de ello tenemos que el estrés puede afectar la salud de los trabajadores mediante dos mecanismos separados. Primero, actuando directamente en el organismo, el estrés puede afectar física y mentalmente (sistema autónomo, neuroendocrina, e inmunidad) esto implica vulnerabilidad ante patologías. Segundo, de manera indirecta en la toma de decisiones o hábitos generados por el estrés ocupacional, hábitos no saludables como fumar, consumo de alcohol, inactividad física y hábitos no saludables de alimentación [3,4]. Para Chen et al [5] el estrés puede dar lugar a diversos síntomas de mala salud si las situaciones estresantes continúan. Ejemplos de estos síntomas son los que se emplean en la definición de estrés en los trabajos [6,7]: dificultad para dormir, cansancio, dolor de cabeza, alteraciones del apetito, falta de concentración, irritabilidad, bajo estado de ánimo, etc. Además de que existe una creciente preocupación de que el estrés laboral contribuye en el aumento de trastornos psicológicos (ej. depresión y ansiedad) en personas previamente sanas, también se incluye la incapacidad para participar en actividades de ocio y mantener fuertes relaciones sociales, sin dejar de lado la pérdida de productividad laboral [1,3].

Diversos autores afirman que el estrés en el trabajo es un importante obstáculo para el funcionamiento eficaz de la organización al contribuir con absentismo, rotación laboral y desempeño laboral. En estimación, los días perdidos debido al estrés se han duplicado, y en la actualidad con mayor reducción de personal, intensificación del trabajo y racionalización de los recursos, se considera que esta tendencia va a continuar [3]. Como resultado del estrés laboral se sufren pérdidas económicas y humanas, debido a esto se ha prestado mayor atención a la prevención del estrés [5]. Con tal de ser más competitivos, las organizaciones son cada vez más exigentes, por lo que en algún momento, tendrá impacto en la vida de todos los empleados, instintivamente de su posición en la organización o la industria en la que trabajan [3].

Algunos estudios demuestran que las condiciones laborales estresantes como, la alta demanda laboral, bajo apoyo social, inseguridad laboral, trabajo repetitivo, y exposiciones ergonómicas fueron predictores significativos del empeoramiento de la percepción de la salud al limitar la capacidad del individuo para hacer cambios positivos para hábitos de vida o contribuyendo directamente a comportamientos negativos de salud [3,5,8].

Diversas investigaciones indican que el estrés muestra mayores niveles cuando se asocia con la alta carga de trabajo y apoyo social bajo [5,9]. En un estudio se encontró que en trabajadores expuestos a altas demandas laborales, el riesgo de padecer depresión o ansiedad es dos veces mayor que en aquellos con bajas demandas, esto se debe a que al combinar múltiples estresores en el trabajo, el riesgo de sufrir estrés es mayor [1,9].

En las últimas décadas se han realizado diversos estudios aplicando el modelo demanda-control-apoyo (DCS) Theorell [10] a diversas áreas. El modelo DCS, se centra en tres dimensiones del ambiente de trabajo: demanda de trabajo, control y apoyo social. Demanda de trabajo se refiere a "la carga de trabajo", mientras que control del trabajo, también llamado libertad de decisión, se refiere a "la capacidad de la persona para controlar sus actividades en el trabajo", y el apoyo social se refiere a "la presencia o ausencia de recursos de apoyo psicológico provenientes de otras personas" [10,11].

En este modelo, los niveles de estrés laboral altos se producen cuando las cargas de trabajo no se compensan con los niveles adecuados de control y apoyo social. Diversas investigaciones muestran que las variables del modelo DCS son predictivos de los resultados de salud y rendimiento laboral, ya que se han identificado condiciones de trabajo específicas que pueden tener graves consecuencias para la salud física y mental de las personas [3].

A pesar de sus límites, y la aparición de otros modelos, el modelo de Karasek sigue siendo un referente en los estudios de los factores psicosociales en el trabajo [10].

2. Datos y variables utilizados

En primer lugar, ha sido indispensable la obtención de los datos publicados por la V Encuesta Europea sobre Condiciones de trabajo (EWCS), los cuales únicamente están disponibles para Universidades y centros de investigación. Para el tratamiento de datos, se utilizan herramientas como Excel y SPSS, ambas herramientas informáticas son muy útiles para la comprensión de datos. Se procedió al análisis de la EWCS para comenzar a comprender la dimensión y relevancia del estudio.

Para poder realizar un estudio estadístico de una serie de datos, ha de elegirse un número concreto de variables para analizar. En algunos casos, las variables estaban recogidas en una única pregunta, y en otros casos era necesario la agrupación de preguntas independientes referentes a la misma temática para generar la variable deseada.

A continuación, con las variables elegidas se estudió su comportamiento utilizando las Redes Bayesianas para profundizar en las interrelaciones existentes entre dichas variables. En todo momento, cuando se obtenían resultados que no eran coherentes, se reformulaba el proceso y se volvía a trabajar con los datos.

2.1. V Encuesta Europea sobre Condiciones de trabajo

El trabajo desempeña un papel destacado en la vida de las personas, en el funcionamiento de las empresas y en la

sociedad en general. "Las políticas europeas reconocen la necesidad de entender mejor las condiciones de trabajo de los ciudadanos de los países de la Unión Europea (UE) para mejorar la calidad del trabajo, aumentar la productividad y crear más empleos" [12].

La encuesta europea de las Condiciones de Trabajo tiene su origen en el año 1990. Desde su primera convocatoria, la Encuesta ha venido proporcionando una panorámica general sobre las condiciones de trabajo en Europa con el fin de evaluar y cuantificar de manera armonizada las condiciones de trabajo de los empleados, analizar las relaciones entre diferentes aspectos de las condiciones de trabajo y determinar los grupos de riesgo, además de contribuir a la formulación de las políticas europeas. Además, en esta última encuesta se puede analizar el impacto de la crisis mundial.

"El objetivo de la encuesta es ofrecer una imagen exhaustiva de la percepción y valoración que hacen los trabajadores europeos de su vida profesional y sus condiciones de trabajo, lo que proporcionará una referencia a los responsables políticos y servirá de base para otros trabajos de investigación" [12].

La Encuesta ha sido realizada por la Fundación Europea para el Desarrollo de las Condiciones de Trabajo, una agencia independiente de la UE con sede en Dublín, en cooperación con organizaciones europeas e internacionales, así como con instituciones nacionales que llevan a cabo este tipo de encuestas a nivel nacional.

La V EWCS fue realizada entre Enero y Junio de 2010 y se encuestaron a casi 44.000 trabajadores europeos de 34 nacionalidades, contestando a más de cien preguntas de diferentes ámbitos relacionadas con su situación laboral.

Los principales temas tratados son: contexto y tiempo de trabajo, la intensidad, factores físicos y psicosociales, la salud y el bienestar, formación y perspectivas profesionales, organización del trabajo, relaciones sociales, satisfacción, conciliación de la vida laboral y personal.

En un esfuerzo por proporcionar datos comparables y fiables sobre las condiciones de trabajo en toda Europa, Eurofound ha desarrollado un enfoque metodológico unificado y un sistema de aseguramiento de calidad. Con los años, la metodología ha sido mejorada.

La población objetivo son todos los residentes de los países mencionados, desde los 15 años (16 años en España, Reino Unido y Noruega) en adelante, y que tenían un empleo en el momento de la encuesta. Se consideraba tener un empleo, según la definición de la Organización Internacional del Trabajo (OIT), si había trabajado durante al menos una hora remunerada en la semana anterior a la entrevista.

La muestra básica es una muestra al azar de múltiples etapas. El tamaño de la muestra objetivo en la mayoría de los países fue de 1000. Las excepciones fueron Alemania y Turquía (tamaño de la muestra 2000), e Italia, Polonia y el Reino Unido (1.500 entrevistados). Además, tres países decidieron financiar muestras nacionales más grandes. En concreto una muestra de 4.000 en Bélgica, 3.000 en Francia y 1.400 en Eslovenia. El número total de entrevistas en 2010 fue de 43.816.

Tabla 1.
Preguntas situación laboral.

Q6: Principalmente ¿Es Vd.?	Q10: ¿Trabaja Vd. en…?
1. Autónomo sin empleados	1. Sector Privado
2. Autónomo con empleados	2. Sector Público
3. Empleado	3. Empresa Pública-privada
4. Otros	4. ONG
	5. Otros

Fuente: V EWCS.

En el momento en que se elaboró la encuesta, la Encuesta de Población Activa anual (EPA) de 2010 encontró que alrededor de 216 millones de personas trabajaban en la UE-27, y 242 millones en todos los países de la encuesta. Debido a la crisis financiera mundial, el número de desempleados no para de crecer en la mayoría de países dejando cifras de paro dramáticas como es el caso de España o Grecia.

La cantidad de contratos temporales en cada país es muy diferente. La media europea que trabajan a jornada partida es de uno de cada cinco trabajadores, mientras que en Países Bajos es uno de cada dos, y en Bulgaria por el contrario, dos de cada cien. Las diferencias en estos indicadores del mercado laboral, muestran un gran impacto de las condiciones laborales de cada país, y han de ser tenidas en cuenta para interpretar los resultados del estudio.

La mayor parte (82%) de la mano de obra en Europa en 2010 se compone de empleados (es decir, los trabajadores con un contrato laboral, y el 15% de la fuerza laboral trabaja por cuenta propia (es decir, son autónomos).

Como el ritmo de trabajo es cíclico a lo largo del tiempo, hay un crecimiento económico para los negocios, que aumenta considerablemente al principio, hasta que existen periodos en los que desaparece este ritmo de trabajo y se entra en recesión. La reciente crisis económica y la crisis financiera, podrían explicar alguna de estos tipos de contrataciones.

2.2. Variables

Con los datos de la V EWCS, el tratamiento se realizó utilizando herramientas como SPSS (análisis de fiabilidad) y Excel (agrupación de variables, análisis de tablas dinámicas).

Se procedió al estudio estadístico de los datos, eligiendo un número concreto de variables para analizar. Para poder cuantificar una característica a partir de datos fue necesario construir grupos para poder hacer comparativas entre variables.

En algunos casos, las variables estaban recogidas en una única pregunta, y en otros casos era necesario la agrupación de preguntas independientes referentes a la misma temática para generar la variable deseada.

2.2.1. Situación laboral

La situación laboral es la variable más importe de la investigación. Con ella se observó la influencia del estrés para cada situación laboral. Para obtener esta variable se agrupó el resultado de dos preguntas (Tabla 2).

Se diferenciaron cuatro grupos asociando los datos de las dos preguntas de la siguiente manera:

Tabla 2.
Agrupación de la variable Situación Laboral

1. Autónomo	
Q6: Principalmente ¿Es Vd.?	Q10: ¿Trabaja Vd. en…?
1. Autónomo sin empleados 2. Autónomo con empleados	1. Sector Privado
2. Privado	
Q6: Principalmente ¿Es Vd.?	Q10: ¿Trabaja Vd. en…?
3. Empleado 4. Otros	1. Sector Privado
3. Público	
Q6: Principalmente ¿Es Vd.?	Q10: ¿Trabaja Vd. en…?
3. Empleado 4. Otros	2. Sector Público
4. Otros	

Fuente: Elaboración propia.

Tabla 3.
Preguntas apoyo social.

Q51 Elegir la respuesta que mejor describa su situación en el trabajo
Q51-A Sus compañeros de trabajo le ayudan y apoyan
Q51-B Su jefe le ayuda y apoya

Fuente: V EWCS.

Tabla 4.
Respuestas apoyo social

Respuestas Q51	
1	Siempre
2	Casi siempre
3	A veces
4	Raramente
5	Nunca
7	No aplicable

Fuente: V EWCS.

2.2.2. Apoyo social

El apoyo social es la variable que representa el grado de ayuda que recibe el trabajador por parte de sus compañeros y jefes. El resultado se obtuvo de la agrupación de las Tablas 3 y 4.

Al tener cada pregunta un grupo de clasificación de cinco valoraciones, se realizó un promedio de las dos y seguidamente una distribución por percentiles.

Para comprobar la fiabilidad de la unión de las preguntas se realizó el análisis de Alfa de Cronbach, la consistencia de la agrupación resultó de 0,709, que al ser mayor de 0,7 se consideró suficiente para validarla.

Obteniendo la fiabilidad de la unión, se formó la clasificación de la Tabla 5.

Tabla 5.
Agrupación de la variable Apoyo Social

Apoyo social	
Grupo 1	Muy bajo
Grupo 2	Bajo
Grupo 3	Moderado
Grupo 4	Alto
Grupo 5	Muy alto

Fuente: Elaboración propia

2.2.3. Demanda

La demanda describe el grado de exigencia laboral al que está sometido el trabajador. La variable se obtuvo a partir de las preguntas y respuestas de las Tablas 6 y 7.

Al agrupar las dos preguntas, se estableció un número menor de grupos que se distribuyeron mediante percentiles. Posteriormente en el análisis de Fiabilidad se obtuvo un coeficiente de Cronbach de 0,781, con lo cual la asociación resulta suficientemente fiable. La agrupación de datos resultante es la de la Tabla 8.

2.2.4. Control

El control a veces es llamado libertad de decisión y es la variable que se refiere a la capacidad de la persona para controlar sus actividades de trabajo. Fue elaborada a partir de la pregunta de la Tabla 9.

Las respuestas a estas cuestiones son:

1-Si 2-No

Los resultados positivos de cada cuestión se agruparon en la variable Control. Para verificar la fiabilidad de esta variable agrupada se realizó un análisis de fiabilidad para el

Tabla 6.
Preguntas demanda laboral.

Q45 Y ¿podría decirme si su trabajo conlleva…?
Q45-A ¿Podría decir si su trabajo conlleva trabajar a gran velocidad?
Q45-B ¿Podría decir si su trabajo conlleva trabajar con plazos muy ajustados?

Fuente: V EWCS.

Tabla 7.
Respuestas demanda laboral

Demanda laboral	
1	Siempre
2	Casi siempre
3	Tres cuartas partes del tiempo
4	Mitad del tiempo
5	Una cuarta parte del tiempo
6	Casi nunca
7	Nunca

Fuente: V EWCS.

Tabla 8.
Agrupación de la variable Demanda Laboral

Demanda laboral	
Grupo 1	Muy baja
Grupo 2	Baja
Grupo 3	Moderada
Grupo 4	Alta
Grupo 5	Muy alta

Fuente: Elaboración propia

Tabla 9.
Preguntas Control.

Q50 ¿Tiene Vd. La posibilidad de elegir o cambiar…?
Q50-A El orden de sus tareas
Q50-B Sus métodos de trabajo
Q50-C Su velocidad o ritmo de trabajo

Fuente: V EWCS.

Tabla 10.
Agrupación de la variable Control

Control	
Grupo 1	Muy bajo: Ninguna respuesta afirmativa
Grupo 2	Bajo: Una respuesta afirmativa
Grupo 3	Alto: Dos respuestas afirmativas
Grupo 4	Muy alto: Tres respuestas afirmativas

Fuente: Elaboración propia.

Tabla 11.
Agrupación de la variable Estrés

Estrés	
Grupo 1	Nunca
Grupo 2	Raramente
Grupo 3	A veces
Grupo 4	Casi siempre
Grupo 5	Siempre

Fuente: Elaboración propia

que se obtuvo un coeficiente de Cronbach de 0,793.
Se estableció el número de respuestas afirmativas como criterio de clasificación de la variable, obteniendo los grupos mostrados en al Tabla 10.

2.2.5. Estrés

"Las personas experimentan estrés cuando sienten que existe un desequilibrio entre lo que se les exige y los recursos con que cuentan para satisfacer dichas exigencias" [13].

Para construir la variable Estrés, se tomaron los datos de la pregunta:

Q51-N: Experimenta estrés en su trabajo

Las respuestas se clasificaron en cinco grupos, ver Tabla 11.

3. Metodología

En este apartado se describe la herramienta matemática utilizada, el objetivo del estudio y el modelo generado.

3.1. Redes Bayesianas

Existe una enorme cantidad de métodos de estudio en el tratamiento de datos, y la utilización de unos u otros repercute en la calidad de los resultados obtenidos.

Esto se debe, a que cada método tiene una forma de construir el modelo de comparación entre los datos y algunos están destinados a ser empleados para temas específicos donde se ha demostrado su validez. Por ello, se ha decidido utilizar un método que relacione unas variables con otras, pudiéndose entender de una manera muy intuitiva.

En esta investigación se ha realizado el estudio de las variables utilizando redes bayesianas, con el objetivo de obtener conocimiento a partir de los datos de la encuesta europea de condiciones de trabajo. El uso de técnicas bayesianas para detectar y mejorar la prevención de accidentes y las condiciones de trabajo es habitual, sirva como ejemplo su uso para prevención de accidentes en minería [14] o en tráfico [15]:

"Una Red Bayesiana (RB) es un modelo probabilístico de una función de probabilidad conjunta (FPC) definido por un

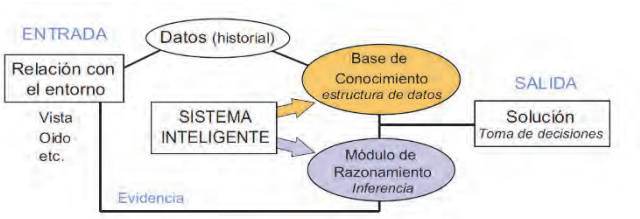


Figura 1. Esquema de un sistema inteligente, con sus componentes principales.
Fuente Gutierrez 2004

grafo dirigido acíclico (DAG) y un conjunto de funciones de probabilidad condicionada, de forma que la estructura de dependencia /independencia mostrada por el DAG puede ser expresada en términos de la función de probabilidad conjunta mediante el producto de varias distribuciones condicionadas, como sigue" [16]:

$$p(x_1, x_2, .. x_n) = \prod_{i=1} p(x_i|\pi_i) \quad (1)$$

Las redes bayesianas son sistemas inteligentes, así que, están compuestos, por un lado de la base de conocimiento y, por otro lado, del módulo de razonamiento (ver Fig. 1). La base de conocimiento almacena la información del problema, y el módulo de razonamiento extrae las conclusiones apropiadas; es decir, a partir de una evidencia (por ejemplo, si los trabajadores están sometidos a unas demandas de trabajo altas) se obtiene información sobre el nivel de probabilidad de estrés esperado [16].

En las redes bayesianas la base de conocimiento es una función de probabilidad conjunta de todas las variables del modelo, mientras que el módulo de razonamiento implementa técnicas eficientes de cálculo de probabilidades condicionadas [16].

Cuando se cuenta con muchas variables definir la probabilidad conjunta no es sencillo (por los grados de libertad que existen), y las redes Bayesianas limitan los grados de libertad para resolver este problema basándose en las dependencias e independencias de las variables obtenidas de la base de conocimiento y reflejadas en el grafo resultante.

La parte más compleja de las redes Bayesianas es el aprendizaje automático del modelo a partir de los datos. Para ello, se han desarrollado distintos algoritmos que permiten obtener, con la mínima supervisión humana, una red bayesiana a partir de los datos disponibles. Actualmente existen varias herramientas que realizan este aprendizaje de forma cómoda y eficiente: Genie, Hugin, Netica, etc.

En este trabajo se utilizó Hugin que es un programa que, partiendo de las variables que deseamos estudiar, crea un modelo de red estudiando si existe cualquier tipo de relación entre ellas, permitiendo discriminar aquellos grupos que no sean interesantes y seleccionar aquellos en los que se quiere profundizar en su estudio.

Después de definir el modelo probabilístico (ver apartado 2.5), la red bayesiana permite calcular las probabilidades "a priori" $p\{Xi = xi\}$, para cada estado xi de cada variable de X, que se corresponden a los estados iniciales de las variables

del modelo, es decir son las frecuencias iniciales de los diferentes estados (ver Fig. 2).

Además, la red bayesiana permite realizar inferencia o razonamiento probabilístico, es decir calcular las probabilidades condicionadas $p(\frac{x_i}{e})$ de cada variable Xi ∉ E , dada una cierta evidencia e (ver apartado 3.1). Así, las nuevas probabilidades indican el efecto de la evidencia en las otras variables del modelo, es decir, se propaga la evidencia.

Analizando las diferencias entre las probabilidades a priori y las condicionadas se determina qué variables influyen más en el nivel de estrés que se quiera estudiar (alto o bajo), de esta manera se analizan y cuantifican los resultados (ver apartado 3).

3.2. Objetivo

Para conocer los niveles de estrés en cada situación laboral, se consideraron las variables relacionadas con el nivel de control, de demanda y de apoyo social, según el modelo demanda-control-soporte (Van Der Doef & Maes, 1990).

Se planteó un modelo de red bayesiana (BN) que proporciona información sobre la relación entre todas las variables estudiadas. El modelo ha sido creado de los resultados de la V EWCS.

3.3. Modelo

En esta red se ha podido diferenciar la influencia de la demanda laboral, control en el trabajo y apoyo social en cada situación laboral. De esta manera, se consiguió analizar si tales condiciones son fuentes claves de estrés para los trabajadores.

La red bayesiana representa gráficamente las variables y sus dependencias, en la cual los nodos representan las variables y los enlaces representan las relaciones de dependencia entre las variables.

Para describir las relaciones probabilísticas de la demanda, el control y el apoyo con la existencia de estrés en las diferentes situaciones laborales se tomó en cuenta el modelo DCS de Karasek, y se desarrolló la red bayesiana como se muestra en la Fig. 2.

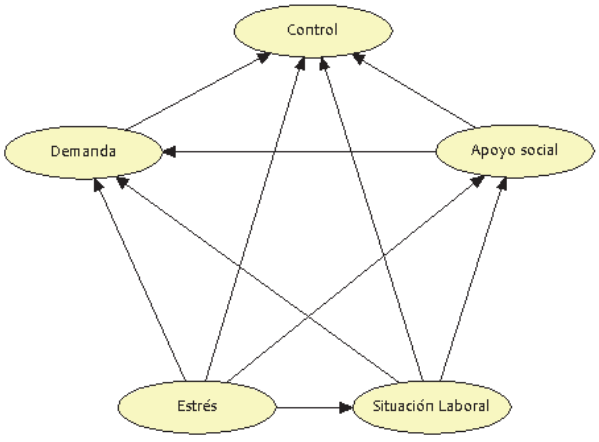


Figura 2. Modelo Demanda-Control-Soporte
Fuente: Elaboración propia

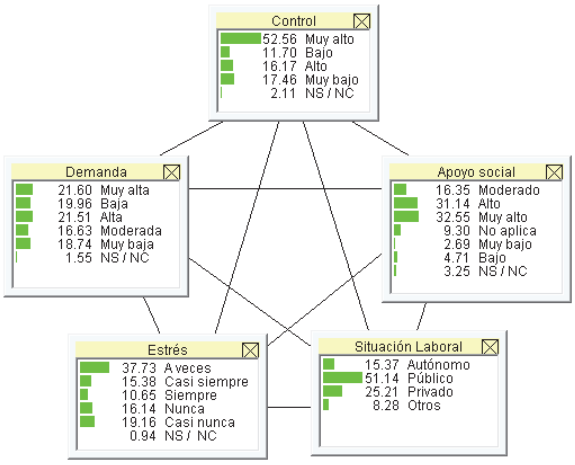


Figura 3. Probabilidades a priori
Fuente: Elaboración propia

Tabla 12
Probabilidades a priori por situación laboral

	Estrés				
Situación laboral	*Nunca*	*Casi nunca*	*A veces*	*Casi siempre*	*Siempre*
Autónomo	14.14	19.92	33.52	15.71	**14.45**
Privado	16.95	19.37	38.11	15.40	**9.56**
Público	15.26	18.43	39.38	15.37	10.90

Fuente: Elaboración propia

La red bayesiana fue generada en Hugin, este programa toma las variables que se desean y crea un modelo de red estudiando si existe cualquier tipo de relación entre ellas, permitiendo discriminar aquellos grupos que no sean interesantes y seleccionar aquellos en los que se quiere profundizar en su estudio.

Se observó que el estrés tiene relación probabilística con la demanda, el control y el apoyo, por lo que se procedió a realizar distintos análisis de sensibilidad.

4. Análisis y Resultados

En primer lugar se muestran en la Fig. 3 las probabilidades a priori de todas las variables que intervienen en el modelo, incluyendo la situación laboral.

Las probabilidades iniciales de sufrir estrés, desglosado por situaciones laborales se muestran en la Tabla 12.

Inicialmente observamos que los autónomos tienen sufren más estrés, mientras que en la situación laboral privado se tiene el nivel más bajo de sufrir estrés de las tres situaciones analizadas.

A continuación se muestran los resultados de los análisis de sensibilidad realizados sobre la demanda, el control y el apoyo social, con ello se consiguen las probabilidades a posteriori de sufrir estrés por parte de los trabajadores.

Análisis de sensibilidad actuando sobre una variable

En este apartado se muestra la variación de la probabilidad de ocurrencia de estrés actuando de forma individual sobre las tres variables de estudio.

Influencia del apoyo social sobre el estrés.

Se observa en la Tabla 13 que los niveles altos de apoyo social consiguen reducir la probabilidad de estrés de los trabajadores en un 46%, partiendo un 14.76% de probabilidad cuando el apoyo social es bajo, y llegando a un 8.1% cuando el apoyo social es alto. Esto ocurre cuando se tiene buen clima laboral resultado del apoyo y ayuda de compañeros y superiores sobre el trabajo.

Posteriormente se analizó la influencia del apoyo social sobre el estrés para cada situación laboral. En el caso de los autónomos no procede que el apoyo social reduce el estrés, ya que estos empleados no tienen un entorno laboral estable.

Para las situaciones laborales privado y público, se muestran las probabilidades de sufrir estrés en las Tablas 14 y 15 respectivamente.

Para las dos situaciones se observa que la probabilidad de sufrir estrés disminuye si el apoyo social es alto. Se aprecia que con apoyo social muy alto no se consiguen mejores resultados en el estrés. Los empleados del sector privado muestran una mayor diferencia en esta reducción, mientras que en los del sector público la diferencia es ligeramente menor.

Influencia del control sobre el estrés.

El control afecta la probabilidad de sufrir estrés muy ligeramente. A mayor control, menor porcentaje de personas que siempre sufren estrés. Resultó interesante observar que la probabilidad más baja de sufrir estrés (10,09%) se tiene cuando el control es alto, cuando podría pensarse que la mejor condición está en el nivel muy alto (ver Tabla 16).

Posteriormente se realizaron análisis individuales de la influencia del control sobre el estrés para cada situación laboral.

Al obtener las probabilidades de las Tablas 17,18 y 19, se constata que en los análisis realizados para cada situación laboral tampoco hay diferencias significativas en la disminución de la probabilidad de sufrir estrés si el control

Tabla 13.
Influencia del apoyo social sobre el estrés.

	Estrés				
Apoyo social	*Nunca*	*Raramente*	*A veces*	*Casi siempre*	*Siempre*
Muy bajo	27.99	17.64	27.91	10.35	14.76
Bajo	13.91	18.02	31.59	19.23	16.33
Moderado	12.13	18.47	39.88	17.7	11.14
Alto	11.59	19.29	42.61	17.94	8.1
Muy alto	20.09	18.96	36.13	12.34	11.81

Fuente: Elaboración propia

Tabla 14.
Influencia del apoyo social sobre el estrés en la situación laboral privado

	Estrés				
Apoyo social	*Nunca*	*Casi Nunca*	*A veces*	*Casi Siempre*	*Siempre*
Muy bajo	32.5	15.21	22.29	13.54	**16.04**
Bajo	13.82	16.84	32.22	19.71	16.6
Moderado	12.3	18.08	40.11	17.69	11.25
Alto	12.2	20.36	42.41	17.58	**7.1**
Muy alto	22.35	19.46	35.78	11.97	9.99

Fuente: Elaboración propia

Tabla 15.
Influencia del apoyo social sobre el estrés en la situación laboral público

	Estrés				
Apoyo social	*Nunca*	*Casi Nunca*	*A veces*	*Casi Siempre*	*Siempre*
Muy bajo	25.33	16.67	30.67	14	**12.67**
Bajo	16.56	17.82	28.09	18.03	18.66
Moderado	11.81	19.15	40.77	16.86	11.07
Alto	11.28	17.81	44.41	17.61	**8.56**
Muy alto	19.11	18.54	36.8	12.63	12.31

Fuente: Elaboración propia

Tabla 16.
Influencia del control sobre el estrés.

	Estrés				
Control	*Nunca*	*Raramente*	*A veces*	*Casi siempre*	*Siempre*
Muy bajo	19.19	17.62	33.76	17	**11.44**
Bajo	17.07	17.77	37.61	16.7	10.48
Alto	14.31	19.9	39.09	15.99	10.09
Muy alto	15.61	19.75	38.65	14.39	**10.65**

Fuente: Elaboración propia

Tabla 17.
Influencia del control sobre el estrés en la situación laboral autónomo

	Estrés				
Control	*Nunca*	*Raramente*	*A veces*	*Casi siempre*	*Siempre*
Muy bajo	17.96	14.64	29.83	19.89	**15.19**
Bajo	20.22	19.12	30.15	14.34	15.07
Alto	13.27	20.88	36.11	15.04	12.92
Muy alto	13.75	20.21	33.65	15.67	**14.56**

Fuente: Elaboración propia

Tabla 18.
Influencia del control sobre el estrés en la situación laboral privado

	Estrés				
Control	*Nunca*	*Raramente*	*A veces*	*Casi siempre*	*Siempre*
Muy bajo	18.85	17.89	34.64	16.95	**10.89**
Bajo	17	18.27	37.86	16.58	9.9
Alto	15.01	20.19	39.52	16.13	8.77
Muy alto	16.78	20.08	39.37	14.06	**9.21**

Fuente: Elaboración propia

Tabla 19.
Influencia del control sobre el estrés en la situación laboral público

	Estrés				
Control	*Nunca*	*Raramente*	*A veces*	*Casi siempre*	*Siempre*
Muy bajo	21.17	17.99	31.79	15.87	**12.23**
Bajo	16.91	16.29	38.46	16.91	11.2
Alto	13.15	18.73	39.4	16.13	11.87
Muy alto	14.03	18.95	42.07	14.42	**9.99**

Fuente: Elaboración propia

Tabla 20.
Influencia de la demanda sobre el estrés.

	Estrés				
Demanda	*Nunca*	*Raramente*	*A veces*	*Casi siempre*	*Siempre*
Muy baja	0	23.07	29.93	8.33	**6.44**
Baja	18.65	26.27	39.28	10.06	4.73
Moderada	12.67	21.86	44.56	13.31	6.92
Alta	10.59	16	43.49	18.77	10.48
Muy alta	9.16	10.41	32.21	24.77	**22.85**

Fuente: Elaboración propia

aumenta. Tampoco se aprecia que por situación laboral haya diferencias significativas.

Influencia de la demanda sobre el estrés.

La demanda laboral ejerce una influencia más significativa en el estrés. Cuando hay un nivel alto de exigencia se llega a triplicar la tasa de estrés con respecto a los trabajadores cuya demanda es muy baja (Ver Tabla 20).

A continuación se analizó cada situación laboral por separado a fin de conocer en cual se tienen las diferencias más significativas de influencia en la reducción de la probabilidad de sufrir estrés.

Los resultados arrojados por los análisis para las tres situaciones laborales (Tablas 21, 22 y 23) muestran diferencias significativas de la probabilidad de sufrir estrés, los incrementos son excesivamente altos, ya que en casos como el de los empleados privados, la probabilidad de sufrir estrés se cuadriplica.

De acuerdo a estos análisis, se observa que la demanda es una detonante en el aumento de las probabilidades de sufrir estrés.

Modelo demanda baja-alto control-alto apoyo social

Una vez obtenidas las probabilidades analizando solo una variable a la vez, se procedió a analizar una las hipótesis del modelo DCS en la que se actúa de forma conjunta sobre las variables de este modelo.

Tabla 21.
Influencia de la demanda sobre el estrés en la situación laboral autónomo

	Estrés				
Demanda	*Nunca*	*Raramente*	*A veces*	*Casi siempre*	*Siempre*
Muy baja	26.19	26.04	26.5	10.4	**8.63**
Baja	15.64	25.57	36.79	11.71	7.36
Moderada	12.74	23.36	38.32	14.02	9.94
Alta	8.61	15.44	38.86	19.29	15.87
Muy alta	7.75	10.11	26.9	23.25	**30.55**

Fuente: Elaboración propia

Tabla 22.
Influencia de la demanda sobre el estrés en la situación laboral privado

	Estrés				
Demanda	*Nunca*	*Raramente*	*A veces*	*Casi siempre*	*Siempre*
Muy baja	34.69	23.59	29.04	7.19	**4.71**
Baja	21.04	27.3	38.48	8.86	3.63
Moderada	13.98	22.39	45.23	12.78	5.03
Alta	11.3	16.85	44.61	18.1	8.69
Muy alta	10	10.85	32.53	24.91	**21.27**

Fuente: Elaboración propia

Tabla 23.
Influencia de la demanda sobre el estrés en la situación laboral público

	Estrés				
Demanda	*Nunca*	*Raramente*	*A veces*	*Casi siempre*	*Siempre*
Muy baja	26.93	21.57	33.28	9.46	**8**
Baja	16.08	25.4	41.61	11.26	5.07
Moderada	11.01	20.48	45.43	13.82	8.98
Alta	10.31	13.64	44.22	19.8	11.61
Muy alta	8.48	9.26	34.27	24.91	**22.67**

Fuente: Elaboración propia

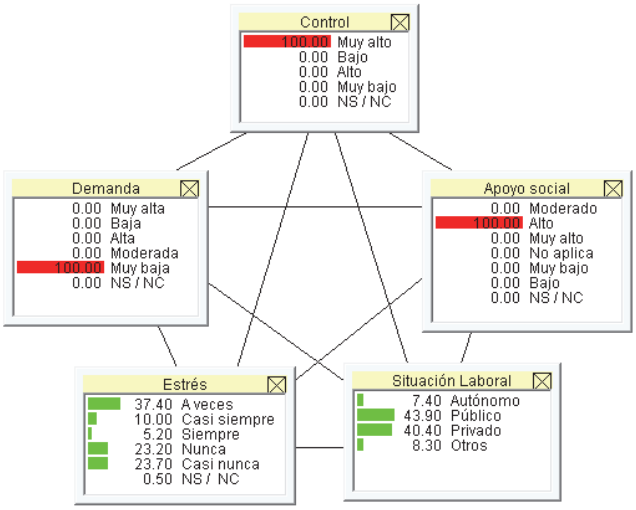


Figura 4. Modelo demanda baja-alto control-alto apoyo social
Fuente: Elaboración propia

Tabla 24.
Influencia de la situación laboral sobre el estrés, en condiciones de baja demanda, alto control y alto apoyo social.

	Situación laboral		
Estrés	*Autónomos*	*Privado*	*Público*
Nunca	19.7	**38.26**	30.39
Raramente	27.09	22.22	20.2
A veces	33.01	27.00	31.92
Casi siempre	8.87	6.19	9.34
Siempre	8.37	**6.05**	7.13

Fuente: Elaboración propia

La hipótesis a probar es que la presencia de muy alto apoyo social en el trabajo reduce el impacto de la demanda muy alta en el bienestar psicológico del empleado y que en el caso donde la demanda es muy baja y el apoyo social muy alto, se tienen los niveles más bajos de estrés (ver Fig.4).

En la Tabla 24 se observa que para las tres situaciones laborales se tienen reducciones considerables de la probabilidad de sufrir estrés, siendo la situación laboral privada la que tiene mayor diferencia al pasar del 38.26% al 6.05%.

También se pudo observar que los empleados autónomos se ven beneficiados por el cambio en dichas variables.

Al comparar los resultados del modelo con las probabilidades a priori, se pudo observar que efectivamente el estrés disminuye bajo estas condiciones del modelo DCS.

Para la situación laboral autónomos, la probabilidad de sufrir estrés inicialmente se encontró en 14.45%, y al realizar el análisis esta probabilidad disminuyó al 8.37%.

Con una disminución ligeramente menos significativa, la situación laboral privada, pasó de 9.56% a 6.05% de probabilidad de sufrir estrés.

Por último, en el sector público, se observó una disminución de aproximadamente 30%, ya que la probabilidad de sufrir estrés inicialmente se encontraba en 10.90%, y con la ejecución del modelo se sitúo en 7.13%.

5. Discusión y resumen

El estrés es una patología emergente que sufren cada vez más empleados, debido a las exigencias del actual entorno laboral. Ambientes laborales con altas exigencias, bajo nivel de control sobre las propias tareas y un bajo apoyo de los compañeros jefes o supervisores, generan las más altas tasas de estrés. El apoyo social, sin duda es un catalizador de los efectos nocivos del estrés. Los empleados de la situación laboral autónomo son los que tienen mayor probabilidad de sufrir estrés y de manera muy frecuente, mientras que los resultados de la situación laboral público y privado son bastante homogéneos.

Se observó que un nivel *muy alto* de apoyo social, genera más estrés que un nivel *alto*, esto puede deberse a la sensación de responsabilidad de cumplir más y mejor con las tareas, a modo de agradecimiento por el apoyo social tan alto que se recibe.

De manera similar, el control sobre las tareas en niveles muy altos genera muy ligeramente que el estrés aumente, analizando los datos, se sugiere que esta situación se da entre los empleados que ocupan puestos jerárquicos con mayor responsabilidad y cargo.

6. Conclusiones

El modelo DCS como predictivo de los resultados de salud y rendimiento laboral, aplicado en la V EWCS permitió identificar las variaciones y similitudes entre las diferentes situaciones laborales que se tienen en la UE. Cumpliendo parte de los objetivos de la V EWCS, se identificaron los niveles demanda, control y apoyo social específicos que generan mayores niveles de estrés. Las EWCS se realizan cada 5 años aproximadamente, por lo que con los datos que surjan de la VI EWCS desarrollada durante el año 2015 se podrá comprobar la evolución de las variables estudiadas.

Referencias

[1] Melchior, M., Caspi, A., Milne, B.J., Danese, A., Poulton, R. and Moffitt, T.E., Work stress precipitates depression and anxiety in young, working women and men. Psychological Medicine, 37, pp. 1119-1129, 2007. DOI: 10.1017/S0033291707000414

[2] Law, R., Dollard, M.F., Tuckey M.R. and Dormann, C., Psychosocial safety climate as a lead indicator of workplace bullying and harassment, job resources, psychological health and employee engagement. Accident Analysis and Prevention, 43, pp. 1782-1793, 2011. DOI: 10.1016/j.aap.2011.04.010

[3] Noblet, A. and Lamontagne A.D., The role of workplace health promotion in addressing job stress. Health Promotion International, 21(4), 2006. DOI: 10.1093/heapro/dal029

[4] Chen, W.Q., Wong, T.W. and Yu. T.S., Association of occupational stress and social support with health-related behaviors among Chinese offshore oil workers. Journal of Occupational Health, 50, pp. 262-269, 2008. DOI: 10.1539/joh.L7149

[5] Elo, A-L, Leppänen A. and Jahkola, A., Validity of a single-item measure of stress symptoms. Scand J. Work Environ Health, 29(6), pp. 444-451, 2003.

[6] Garcia-Herrero, S., Mariscal, M.A., Garcia-Rodríguez, J. and Ritzel, D.O., Influence of task demands on occupational stress: Gender differences. Journal of Safety Research, 43(5-6), pp. 365-374, 2012. DOI: 10.1016/j.jsr.2012.10.005

[7] García-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(9), pp. 1760-1774, 2012. DOI: 10.1016/j.ssci.2012.04.005

[8] Cohidon, C., Niedhammer, I., Wild, P., Guéguen, A., Bonenfant, S. and Chouanière, D., Exposure to job stress factors in a national survey in France. Scand J. Work Environ Health, 30(5), pp 379-389, 2004.

[9] García-Herrero, S., Mariscal, M.A., Gutiérrez, J.M. and Ritzel, D.O., Using Bayesian networks to analyze occupational stress caused by work demands: Preventing stress through social support. Accident Analysis and Prevention, 57(0), pp. 114-123, 2013. DOI: 10.1016/j.aap.2013.04.009

[10] Van der Doef, M. and Maes, S., The job demand-control (-Support) Model and psychological well-being: A review of 20 years of empirical research. Work & Stress: An International Journal of Work, Health & Organisations, 13(2), pp. 87-114, 1999. DOI: 10.1080/026783799296084

[11] Pérez, J. and Martín, F., El apoyo social. Instituto Nacional de Seguridad e Higiene en el Trabajo, 1991.

[12] EUROFOUND. Cuarta Encuesta Europea sobre las condiciones de trabajo. [Online]. Available at: www.eurofound.europa.eu/pubdocs/2006/78/es/1/ef0678es.pdf

[13] EU-OSHA. [Online]. Available at: https://osha.europa.eu/es/topics/stress/ definitions_and_causes

[14] de Felipe-Blanch, J.J., Freijo-Álvarez, M., Alfonso, P., Sanmiquel-Pera, L. and Vintró-Sánchez, C., Occupational injuries in the mining sector (2000-2010). Comparison with the construction sector, DYNA, 81(186), pp. 153-158, 2014.

[15] Guerrero-Barbosa, T.E. and Amarís-Castro, G.E., Application of bayesian techniques for the identification of accident-prone road sections, DYNA, 81 (187), pp. 209-214, 2014. DOI: 10.15446/dyna.v81n187.41333

[16] Gutierrez, J.M, Cano, R., Cofiño, A y Sordo, C., Redes probabilísticas y neuronales en las ciencias atmosféricas. Ministerio de Medio Ambiente. Madrid. ISBN: 84-8320-281-6. 2004

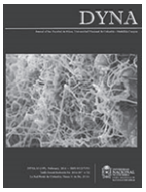
S. García-Herrero, received the MSc. Eng in Industrial Engineering in 1996 from the Universidad de Cantabria, Spain and the PhD degree in Industrial Engineering in 2002, from the Universidad de Burgos, Spain. Currently, he is a full professor in the Civil Engineering Department, Escuela Politécnica Superior, Universidad de Burgos, Spain and director of the reach group "Industrial organizational Engineering". Her research interests include: simulation, safety and prevention, culture organization, and industrial engineering using time-series analysis and computational intelligence techniques.
ORCID: 0000-0002-8061-9537

M.A. Mariscal-Saldaña, received the MSc. Eng in Industrial Engineering in 1997 from the Universidad de Cantabria, Spain and the PhD degree in Industrial Engineering in 2002, from the Universidad de Burgos, Spain. Currently, he is a full professor in the Civil Engineering Department, Escuela Politécnica Superior, Universidad de Burgos, Spain. His research interests include: simulation, safety and prevention, culture organization, and industrial engineering using time-series analysis and computational intelligence techniques.
ORCID: 0000-0002-7897-9891

E.M. López-Perea, received the MSc. degree in inclusive education and society in 2014 from the Universidad de Burgos, Spain. Currently, he is a full professor in the Education Department, Faculty of Education, Universidad de Burgos, Spain. His research interests include: simulation, safety and prevention and culture organization using time-series analysis and computational intelligence techniques.
ORCID: 0000-0002-1352-0430

M.F. Quiroz-Flores, studied the degree in Industrial Engineering from the Instituto Tecnológico de Toluca, México. Her research interests include: lean manufacturing, safety and prevention, culture organization, stress management, industrial engineering and Bayesian networks.
ORCID: 0000-0002-4904-8609

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS



DYNA
http://dyna.medellin.unal.edu.co/

Saraz method adjustment for the quantification of ammonia emissions generated in opened or hybrid animal production facilities

Fernanda Campos-de Sousa [a], Ilda de Fátima Ferreira-Tinôco [b], Jairo Alexander Osório-Saraz [c], Keller Sullivan Oliveira-Rocha [d] & Maximiliano Arredondo-Ramirez [e]

[a] *Department of Agricultural Engineering, Federal University of Viçosa, Viçosa, Brazil. fernanda.sousa@ufv.br*
[b] *Department of Agricultural Engineering, Federal University of Viçosa, Viçosa, Brazil. iftinoco@ufv.br*
[c] *Department of Agricultural Engineering, Universidad Nacional de Colombia, Medellín, Colombia. aosorio@unal.edu.co*
[d] *Department of Agricultural Engineering, Federal University of Viçosa, Viçosa, Brazil. kellersullivan@yahoo.com.br*
[e] *Department of Agricultural Engineering, Universidad Nacional de Colombia, Medellín, Colombia. marredondor@unal.edu.co*

Received: January 23th, 2015. Received in revised form: October 5th, 2015. Accepted: October 14th, 2015.

Abstract
The objective of this study was to evaluate the efficiency of the Saraz method in order to quantify ammonia emissions generated in opened or hybrid animal production facilities, and to determine an equation for the adjustment method. To do this, we developed beacon equipment, with input and output gas sectors, hoods and absorbent porous material. After the collection, the amount of ammonia captured in the environment was determined in the laboratory. Different ammonia concentrations were evaluated in addition to the different speeds of the exhaust air. Considering the results, it can be concluded that for the situations analyzed the Saraz method is efficient, but as with other methods, with an increase of air velocity and concentration, its efficiency decreases. An equation for the adjustment of the Saraz method was generated to determine the concentration and the rate of ammonia emissions inside animal facilities.

Keywords: air quality; gas emission; natural ventilation.

Método Saraz ajustado para cuantificar emisiones de amoníaco generadas en instalaciones de producción animal abiertas o híbridas

Resumen
El objetivo de este estudio fue evaluar la eficacia del método Saraz utilizado para cuantificar las emisiones de amoníaco generadas en instalaciones de producción animal abiertas o híbridas y determinar una ecuación ajustada para el método. Para esto, fue desarrollado un equipo con entrada y salida de gases, campanas y materiales porosos absorbentes. Después de la recolección, la cantidad de amoníaco capturado en el medio ambiente fue determinado en el laboratorio. Diferentes concentraciones de amoníaco fueron evaluadas, además de diferentes velocidades del aire. Con base a los resultados, se puede concluir que el método Saraz es eficiente dentro de las situaciones analizadas, pero al igual que otros métodos, con el aumento de la velocidad del aire y la concentración, su eficiencia disminuye. Una ecuación ajustada para el método Saraz fue encontrada, para determinar la concentración y la tasa de emisión de amoniaco en el interior de instalaciones animales.

Palabras clave: calidad del aire; emisión de gases; ventilación natural.

1. Introduction

Ammonia is the most common pollutant found in high concentrations on the premises of animal production facilities [1]. The presence of this gas directly influences the growth of young animals, especially birds. Exposure to an excess of ammonia causes problems and diseases, such as reduced appetite and respiratory rate, respiratory tract lesions, conjunctivitis, and increased susceptibility to viral infections [2]. Besides the economic issues related to ammonia gas in animal husbandry environments, due to it damaging the health of animals and workers, it is also worthwhile to note that ammonia is a greenhouse gas that

leads to undesirable environmental consequences when emitted in high concentrations [3]. It may also contaminate water, air and soil, affecting the planet's future of and animal production sustainability [4]. Thus, it is necessary to implement technological strategies to minimize the environmental impact caused by the emission of greenhouse gases [5].

For about two decades, emissions rates and/or concentrations of ammonia in animal production environments have been the focus of studies by researchers in various European countries and in North America, where inventories have been carried out on greenhouse gas emissions, establishing protocols. For these countries, determining ammonia emissions is relatively simple since facilities are closed, and, therefore, there is control over the volume of air in the premises [6].

For regions with tropical and subtropical climates, such as Brazil, determining the concentration of ammonia emissions is much more complex since virtually all livestock shelters are kept open, constituting open or partially open (hybrid) systems, which have interference currents and uncontrollable external wind [7] thermodynamic systems.

Among the methods available to naturally ventilate in predominantly open facilities is that of passive flow. The Saraz method proposed by [8] is noted for its simplicity, efficiency, and applicability to being able to determine ammonia emissions in livestock production in open and general facilities.

The Saraz method was evaluated in field conditions [8] and was considered to be efficient in measuring ammonia emissions from broiler litter coming in very low concentrations, such as 0.5 ppm and 1.0 ppm in ventilation conditions for natural air velocity above 0.1 m s^{-1}. However, according to [8], the method needs to be improved and more research should be conducted into different concentrations and environmental conditions, with the aim of gaining a better understanding of the efficiency and applicability in inventories of greenhouse gases such as ammonia gas.

Despite the great diversity of methods to quantify the concentration of ammonia in the atmosphere , the majority of these are expensive and involve undertaking a series of steps in the laboratory that may contaminate the sample. Therefore, comparison studies are still required, as well as the adaptation and application of methods to quantify ammonia [3].

Thus, there is a need to assess the efficiency of the proposed Saraz method in different environmental conditions, with variations in ambient air velocities, and with different ammonia concentrations studied by [8] in the conditions in the field shelters for livestock. In order to do this, it is important to find a reliable method of calibration, which allows for the actual and possible values of ammonia levels to be quantified. The pickup device employed in the Saraz method is capable of measuring in environments with climate variability and, therefore, it captures the rate of ammonia emissions in a controlled situation by generating an equation for the Saraz method environmental setting.

2. Material and methods

This work was conducted in the Department of Agricultural Engineering at the Federal University of Viçosa in climatic chambers, at Ambiagro – the Center for Research in Agro-industrial ambience, and also at the Engineering Systems and laboratories in the areas of Rural Buildings and Ambience and Energy in Agriculture.

2.1. The beacon equipment development

To enable the improvement of the Saraz method, fully sealed beacon equipment was developed, consisting of a box of translucent glass and polycarbonate with the following dimensions: 30.0 x 38.0 x 28.0 cm (length, width, height). This was inserted in the climatic chamber in order to obtain controlled environmental conditions, enabling the analysis in an environment in which the temperature and humidity are controlled and known.

The box, called a normalizing box, was divided into two compartments, an entry sector, and a gas outlet with the same dimensions 30.0 x 18.0 x 28.0 cm (length, width, height). The sector gas inlet was composed of a housing provided with openings for gases to enter(ammonia and fresh air) through silicone tubing connected to the external gas cylinders for the purpose of this study. This compartment was provided with two circular openings of 4.0 cm radius, sealed by absorbent porous material (described in 2.2) positioned above the hoods, which when in use, forced the removal of the outside air to the box. The sector gas outlet, was, in turn, subdivided into two equal sized compartments to allow two samples of air through the porous absorbent material via the exhaust and, hence, two replicates for each variable were analyzed (Figs. 1 and 2).

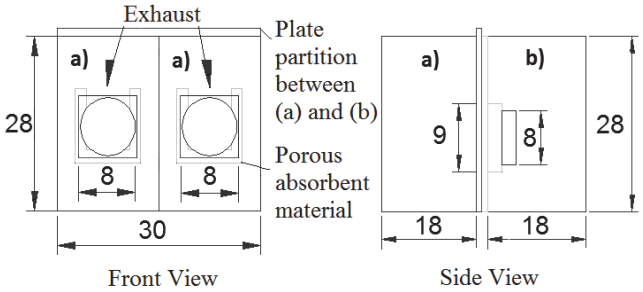


Figure 1. Schematic of box. Front and side view of the box showing the extraction fans and the compartments: a) gas inlet, b) gas outlet (measured in cm).
Source: The authors.

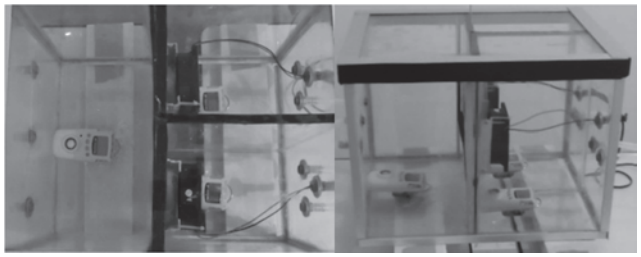


Figure 2. Box top view (left) and (right) lateral view.
Source: The authors.

2.2. Absorbent porous material

The absorbent porous material used was composed of polyurethane sponges with a density D18 (0.0162 g cm^{-3}) and dimensions of 8.0 cm in length, 9.0 cm in height, and 2.0 cm in thickness.

For each repetition, the sponges were impregnated in 25 ml of sulfuric acid (H_2SO_4), 1.0 mol L^{-1}, and glycerin ($C_3H_5(OH)_3$) 3% v/v before sealing the holes of each of the box's input compartments, according to the method proposed by [9].

2.3. Exhaust fans

The exit velocity of the gas at a known concentration through the porous absorbent material was different for each treatment and had a different speed to the exhaust air (0.1, 1.2, and 2.4 m s^{-1}). This was achieved using hoods installed inside the box that forced the gas through the passage foams, drawing air from inlet compartment to the outlet compartment. The exhaust fans were made by fans or cooler "sleeve bearings" with dimensions of 80 x 80 mm, had a DC voltage of 12 V, a current of 0.15 A, and a speed of 1800 rpm. The speed of the exhaust was controlled by a universal energy source model FT-1462P with a power of 18 W, a maximum load of 1000 mA, a 110/220 Vac input, and seven outputs.

Initially, an exhaust calibration curve was taken to determine the rate of entry of gas into the foam; one hot wire term anemometer (TESTO model 425) was used, measuring with a range from 0 to 20 m s^{-1} and with an accuracy of ± 0.03 m s^{-1} and a resolution of 0.01 m s^{-1}.

2.4. Electrochemical Sensors

Electrochemical detectors were used as the instruments to make reference measurements of the various adopted ammonia concentrations (5, 10, 15, 20, and 25 ppm). BW "Gas Alert Extreme Ammonia Detectors", which are compact and affordable, present a variation in measuring accuracy from 0 to 100 ppm 2%, and operate at a relative humidity from 15 to 90% and a temperature of -20 to + 50° C.

The sensors were initially unloaded, and, despite having an automatic setting to grade the security every time you are connected, a simple calibration of all sensors used in the experiment was additionally performed. A total of six sensors were used, three for each sample, with one positioned on each of the three compartments of the normalizing box, the initial compartment for the air containing a known concentration of ammonia and the two sequential compartments with the same dimensions containing the gas after it had passed through the absorbent porous material.

At the end of each collection, a change of the position of sensors occurred, which allowed for a reading of the ammonia level in each sample. Different sensors were positioned in different compartments, thus consecutive readings were avoided. Simultaneously, the sensors were calibrated, and any necessary correction factors were set.

2.5. Climatic chamber

The experiment was conducted in a climatic chamber in order to control the thermal environment for each gas concentration studied. The climatic chamber, located in the Annex to AMBIAGRO, DEA, UFV, was equipped with hot/cold air conditioning of 12,000 BTU h^{-1}, an electric resistance heater with 2000 W of power, a humidifier with a capacity of 4.5 liters and a 300 mL/hr output medium mist. The heater and the humidifier were controlled by an electronic temperature and humidity controller with a 531Ri MT-plus serial communication. Each device was powered on or off automatically in order to ensure temperature control and humidity patterns in the study set.

In this experiment, samples were collected at a temperature of 25° C, and relative humidity was maintained at around 60% in order to simulate creation conditions that were within the range of appropriate comfort for animals such as birds from their second week of life [10]. At this point the generation of ammonia in the production environment is intensified.

2.6. Calibration of the NH3 sensors

The calibration process of the NH_3sensors was achieved by using a gas with a known concentration of ammonia and another gas that did not contain ammonia. A gas cylinder with a known ammonia concentration of 25 ppm and an air cylinder with ultra pure synthetic zero with 0 ppm ammonia were used. The cylinders were connected to silicone tubes, and these were connected to measuring instruments following the methodology described by [11].

2.7. Functioning of the normalizing box

At the beginning of each sample collection, the inlet chamber of the housing of the normalizing box received a volume of air with different ammonia concentrations, namely 5, 10, 15, 20, and 25 ppm. These values are situated in the range between the maximum and ideal values for ammonia in an environment for continuous, intermittent and sporadic exposure to animals. These different concentrations of ammonia were obtained from mixtures of ammonia gas up to 25 ppm and with zero ultra pure synthetic air and were monitored by means of electrochemical detectors placed inside the box. As the gases were added to the inlet chamber, the readings were observed by sensors until the desired concentration was obtained. After a time for the sensors to be stabilized and about five equal consecutive readings of the same concentration, we initiated the process of capturing the ammonia absorbed by the porous material that was positioned next to the exhaust fans in contact with the air inlet.

Different speeds of the exhaust air were assessed: 0.1, 1.2, and 2.4 m s^{-1}; these were the upper and lower limits of air typical in some animal production facilities. Birds have a slower speed compared with other animals such as pigs [12-14].

With the exhaust in operation, after the concentration of ammonia was stabilized, a hatch (separating the

compartments of input and output gas) was opened that allowed the passage of gas through the porous absorbent material and it to be simultaneously captured. Two other sensors were positioned in the gas exit compartments, one in each compartment, which were monitored to assess possible leaks and to verify the absorption capacity of the absorbent porous material. The final collection time given by the sensors indicated a value of 0 ppm for five consecutive readings, thus indicating that all ammonia present within the gas inlet chamber had been captured by the absorbent porous material.

2.8. Concentration and emission rate of ammonia

The ammonia concentration was determined from the amount of ammonia that the sponges captured, impregnated first with 25 ml of sulfuric acid (H_2SO_4), 1.0 mol L^{-1}, and glycerin (C_3H_5 $(OH)_3$) 3% v/v solutions; they were responsible for fixing the diffused ammonia.

After the capture period, the exhaust fans were turned off and the absorbent porous material was removed, stored and refrigerated in plastic wrap to be taken to the laboratory for the ammonia to be extracted according to the Kjeldahl method, in accordance with the methodology adopted by [9].

The ammonia concentration captured by the sponges, that is, the concentration obtained by the Saraz method, was compared to the concentration of ammonia present in the actual normalizing box in terms of the amount of ammonia contained in the foam and calculated in the box, considering the volume of the foam and the normalizing box. Thus, for the amount of ammonia present in the box, the volume of gas input was considered to be equal to 15,120 cm^3, taking into account the dimensions of 30.0 cm length, 18.0 cm width, and 28.0 cm in chamber height. For the sponge, the amount was 144 cm^3, considering the dimensions of 2.0 cm thickness, 8.0 cm width, and 9.0 cm height. An equation to adjust the concentration obtained from the absorbing porous material due to the expected concentration to be collected by the sponge was generated.

According to the Saraz method, the emissions rate of ammonia through the porous material can be calculated by applying eq. (1):

$$SARAZ_{NH3} = \frac{NH_3}{t} \quad (1)$$

Where:

Saraz NH_3: ammonia emission observed in the Saraz method rate (mg s^{-1});

NH_3: Ammonia captured by the absorbent material (mg NH_3);

t: Time of exposure to ammonia gas absorbent material (s).

By calculating the quantity values of the NH_3 and NH_3 concentrations and the rate of ammonia emissions observed with the expected values in the environment, the efficiency was analyzed by the Saraz method according to the time of exposure to the different concentrations of the ammonia situations and different speeds of the exhaust air.

2.9. Experimental design

The experimental design was completely randomized with four replications. The treatments consisted of the combination of five different concentrations of ammonia (5, 10, 15, 20, and 25 ppm) and three different speeds of the exhaust air (0.1, 1.2, and 2.4 m s^{-1}) to analyze the amount of ammonia recovered by the foam and the evaluation of the ammonia emissions rate under different conditions. There were a total of 60 samples (n = 60).

The data amount of ammonia obtained by foam and the ammonia emissions rate were subjected to analysis of variance (ANOVA). When the F test of ANOVA showed difference, the means were compared by Tukey's test. For all statistical analysis, the value of 5% for the probability of type I error was adopted. For an analysis of the model under study, the following hypotheses were tested:

The null hypothesis (H_0): concentration data calculated in a real ammonia environment ($ERNH_3$) are equal to the data indirectly through the absorbent material ($SARAZNH_3$).

The alternative hypothesis (H_1): concentration data calculated in a real ammonia environment ($ERNH_3$) differ from the data indirectly through the absorbent material ($SARAZNH_3$).

If the alternative hypothesis (H_1) was true, and validated using the ANOVA test, we would then carry out a linear regression analysis to determine the coefficients of the model expressed by eq. (2) using the Sigma Plot version 12.0 program.

$$ER_{NH3} = a \cdot SARAZ_{NH3} \pm b \quad (2)$$

3. Results and discussion

Table 1 presents data for mean recovery of ammonia using the Saraz method for the different concentrations tested. It is observed that, for lower concentrations, the Saraz method displayed the highest ammonia recovery efficiencies of around 81% for the 5 ppm concentration and 51% for the 10 ppm concentration. These results are in agreement with work done by [8], which was undertaken in field conditions with a recovery of volatilized ammonia poultry manure in the 68–82% range.

The Saraz method proves to be more efficient, compared to studies by [15], [16], [17] and [18], which use a collection chamber to determine volatilized ammonia from poultry manure and soil that would normally have maximum values of 70% recovery of ammonia.

Table 1.
Average recovery of NH_3 using the Saraz method in different concentrations analyzed and their standard deviations

NH_3 Concentration (ppm)	Average recovery (%) ± Standard deviation
5	81.10±8.22[a]
10	51.54±5.80[b]
15	41.00±3.10[c]
20	36.90±6.90[d]
25	34.31±3.20[e]

Means followed by different letters in the emissions of ammonia rates column differ (P <0.05) from Tukey's test.
Source: The authors.

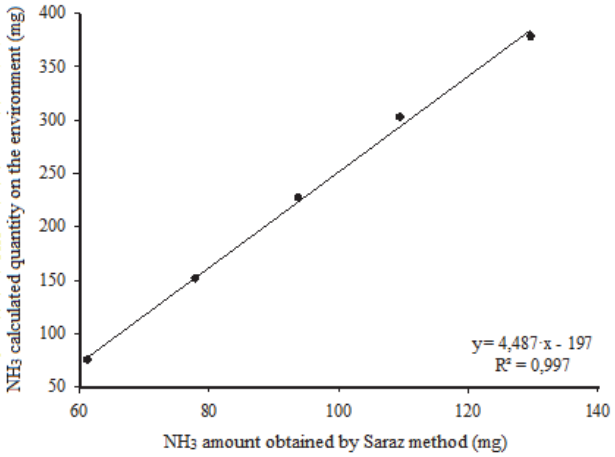


Figure 3. The linear regression curve for the different concentrations for the amount of ammonia obtained by the absorbent porous material (Saraz) and the amount of ammonia present in the environment.
Source: The authors.

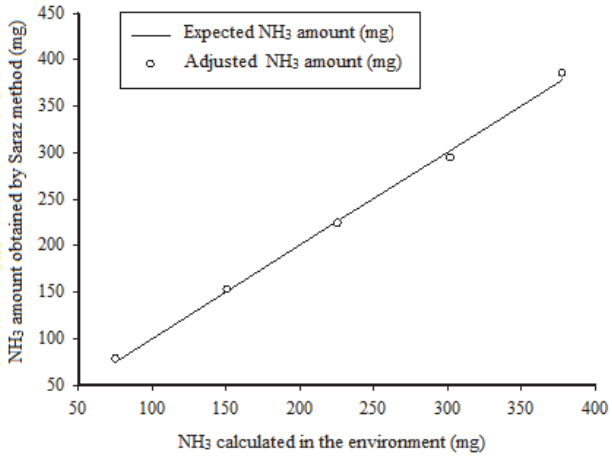


Figure 4. Representation of setting the values for the quantity of NH_3 absorbent porous material obtained by using the Saraz method.
Source: The authors.

This higher recovery efficiency of ammonia through the collector box sponge, compared to what happens in practice, can be explained by the fact that, in this experiment, ammonia present in the ambient air was volatilized. This is unlike what happens in other studies in which the ammonia was still in the process of volatilization in the middle bed or the ground and was then captured by the collector. It was thus dependent on the environment's conditions of volatilization, which can interfere with the recovery efficiency. Similar results were found in [19,17]'s studies, which showed that the efficiency of the collection chamber for open semi-static NH_3 varies with the amount of volatilized ammonia, both under greenhouse and in field conditions.

It can be inferred, therefore, that the method used in this study has a good recovery efficiency of ammonia, without an adjustment curve for environments, with concentrations of up to 10 ppm.

Fig. 3 shows the linear regression curve for the amount of ammonia obtained by the sponge and the amount of ammonia present in different concentrations. It can be observed in Fig. 3 that, independent of the velocity of the exhaust air, one can adjust an equation for any concentration within the values that are being analyzed (up to 25 ppm), since there is a high correlation between the values obtained by foam ammonia and the ammonia values calculated in the environment (R^2 = 0.997). This correlation is linear, showing that the method is valid and can be used for the conditions that are analyzed.

Using the regression equation shown in Fig. 3, the calibration equation for the Saraz method, eq. (3), was obtained to determine the actual amount of ammonia captured in situations in which the environment contained less than 25 ppm of ammonia.

$$NH_3(mg) = 4.487 \cdot SARAZ_{NH_3}(mg) - 197 \qquad (3)$$

Fig. 4 shows an overlapping set of values with expected values in all observations, which means that an equation fit

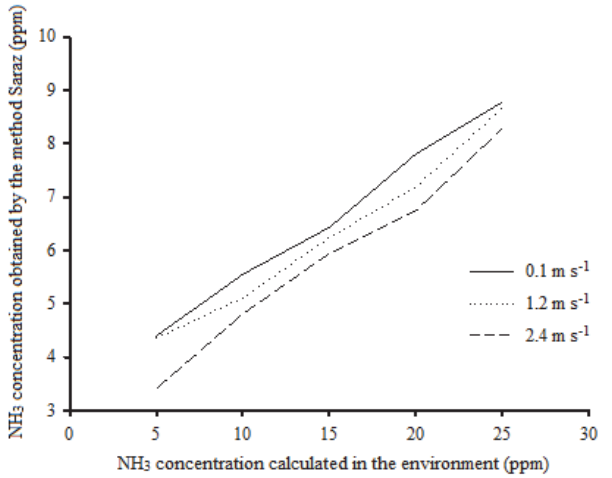


Figure 5. Representative curves of the speeds of the exhaust air with the ammonia concentration obtained by the absorbent porous material in different ambient concentrations.
Source: The authors.

of the data obtained can be applied. It is therefore indicated that the use of the setting values in the equation for ammonia concentrations in absorbent porous material obtained by the Saraz method proves efficient at any concentration for those tested (up to 25 ppm).

Fig. 5 shows the relationship between the concentration of ammonia present in the environment and the ammonia concentration obtained by the absorbent porous material using the Saraz method, at different speeds of exhaust air. No statistically significant difference (P> 0.05) was found using analysis of variance among the three adopted speeds (0.1, 1.2 and 2.4 m s-1). This indicates that the use of the Saraz method enables the efficient recovery of ammonia without being affected by the air velocity, and, as was expected, the amount of ammonia recovered decreases with increasing air velocity.

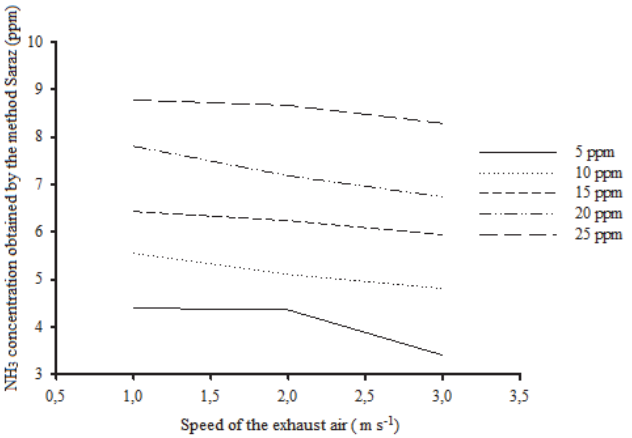


Figure 6. Representative ammonia curves according to the amount absorbed by the porous material and different speeds of the analyzed exhaust air environment.
Source: The authors.

Similarly to Fig. 5, Fig. 6 shows that, independent of the velocity of the exhaust air, the behavior of the curves is the same for all the velocities studied. The amount of ammonia captured by the collector was directly proportional to the concentration of ammonia in the environment, i.e., the higher the amount of water present in the environment, the greater the amount collected by the Saraz ammonia method. In contrast, there is an inversely proportional relation when analyzing the exhaust air speed and the amount of ammonia captured by the collector; the higher the speed of the exhaust air, the smaller the amount of ammonia captured by the Saraz method.

In Fig. 7 shows the regression curve for the rate of ammonia emissions observed from the data obtained by ammonia absorbent porous material ($Saraz_{NH3}$) and the emission rate of ammonia calculated by the expected environmental concentration (ER).

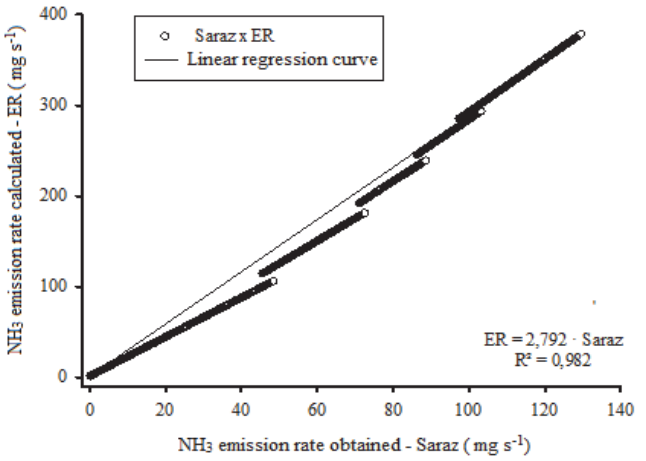


Figure 7. Relationship between the rate of ammonia emission observed (Saraz) and the emission rate of ammonia calculated for different concentrations (ER).
Source: The authors.

There is high correlation ($R^2 = 0.982$) between the data and the values of the ammonia emission rate obtained by the Saraz method, and the rate values calculated for ammonia emissions in the environment that exhibit a linear relationship, showing that the method is valid under the conditions analyzed.

By using the linear regression curve shown in Fig. 7, it appears that we can estimate the rate of actual emission of ammonia by using a correction factor of 2.79 on the emission rate values obtained by the Saraz method. That is, the real value of the ammonia emission rate (ER) for air velocities between 0 and 2.4 m s^{-1} and concentrations up to 25 ppm, can be calculated using eq. (4), which is shown below.

$$ER = 2.79 \cdot SARAZ_{NH_3} \quad (4)$$

In Fig. 8 (where rates represent ammonia emissions obtained at the various speeds of the exhaust air), the same behavior found in the analysis of the amount of ammonia obtained can be observed. There were no statistically significant differences found (P> 0.05) between the emissions of this gas for each of the different speeds of the exhaust air.

Applying the analysis of variance (ANOVA) to the data rate of ammonia emission, it is possible to observe the significance of this gas' emissions in various concentrations of ammonia in the environment (Table 2). It can be observed that the values of ammonia emission rates increased with increasing concentration in the environment, as was expected.

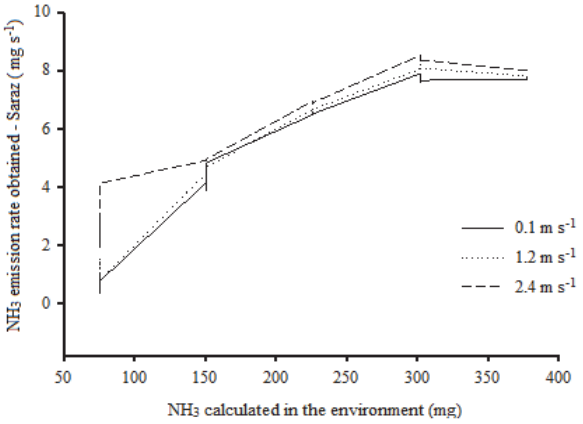


Figure 8. Representation of the behavior of the exhaustion velocity curves to calculate the ammonia emission rate at different concentrations.
Source: The authors.

Table 2.
Average values of ammonia emission rate at the different concentrations analyzed, with their respective standard deviations

Concentration NH3 (ppm)	Average emission rate of ammonia ($mg.s^{-1}$)
5	1.59±0.45 a
10	4.61±0.16ab
15	6.70±0.03 bc
20	8.10±0.09 d
25	7.82±0.02 cd

Means followed by different letters in the ammonia emission rate column differ (P <0.05) from Tukey's test.
Source: The authors.

4. Conclusions

The Saraz method is an efficient method. To determine concentrations of up to 10 ppm it does not need adjustment and presents a good recovery efficiency. For concentrations higher than 10 ppm, the method can be used with confidence through the use of generated equations to adjust both the concentration (eq. 3) and for the rate of ammonia emission obtained by either method (eq. 4), regardless of air velocity.

With air speeds of up to 2.4 ms^{-1} and ammonia concentrations of up to 25 ppm, the method can be used reliably, but as with other methods, when an increase in air velocity and concentration occurs, efficiency decreases.

Acknowledgments

The authors are thankful to the CNPq, Capes and FAPEMIG.

References

[1] Owada, A.N., Nãas, I.A., Moura, I.J. y Baracho, M.S., Estimativa de bem-estar de frango de corte em função da concentração de amônia e grau de luminosidade no galpão de produção. Eng. Agríc., 27(3), pp. 611-618, 2007. DOI: 10.1590/S0100-69162007000400003

[2] Osorio, J.A., Tinoco, I.F.F. and Ciro, H.J., Ammonia: A review of concentration and emission models in livestock structures. DYNA, 76(158), pp. 89-99, 2009.

[3] Felix, E.P. and Cardoso, A.A., A method for determination of ammonia in air using oxalic acid-impregnated cellulose filters and fluorimetric detection. J. Braz. Chem. Soc., 23(1), pp. 142-147, 2012. DOI: 10.1590/S0103-50532012000100020

[4] Carolla, C., Sánches, R. y Montiel, E., Modelo de superfície de respuesta que permite inferior concentración de nitrógeno em "compost" producido a partir de desechos orgánicos. Ing. Investig., 29(3), pp. 128-133, 2009.

[5] Vargas-Nieto, C., Carriazo, J.G. y Castillo, E., Estudio de materiales adsorbents de bajo costo para remover Cr(VI) de efluentes acuosos. Ing. Investig., 31(1), pp. 154-162, 2011.

[6] Osorio-Saraz, J.A., Tinoco, I.F.F., Gates, R.S., Paula, M.O. and Mendes, L.B., Evaluation of different methods for determining ammonia emissions in poultry buildings and their applicability to open facilities. DYNA, 80(178), pp. 51-60, 2013.

[7] Tinoco, I.F.F. y Osorio, J.A., Control ambiental y la agroindustria de produccion animal en el Brasil y America Latina. In: Congreso Nacional de Ingenieria Agricola, Medellín, Colombia, 2008.

[8] Saraz, J.A.O., Tinoco, I.F.F., Gates, R.S., Rocha, K.S.O., Caballero, E.M.C. and Sousa, F.C., Adaptation and validation of a methdology for determing ammonia flux generated by litter in naturally ventilated poultry houses. DYNA, 81(187), pp. 137-143, 2014. DOI: 10.15446/dyna.v81n187.40806

[9] Osorio, J.A., Determinação experimental e modelagem em CFD das taxas de emissões de amônia de camas de aviários e distribuições de concentrações, temperatura e velocidade do ar no interior de galpões avícolas. Thesis, Department of Agricultural Engineering, Federal University of Viçosa, Viçosa, Minas Gerais, Brasil, 2010.

[10] Cassuce, D.C., Tinôco, I.F.F., Baêta, F.C., Zolnier, S., Cecon, P.R. and Vieira, M.F.A., Thermal comfort temperature update for broiler chickens up to 21 days of age. Eng. Agríc., 33(1), pp. 28-36, 2013. DOI: 10.1590/S0100-69162013000100004

[11] Amaral, M.F.P., Gates, R.S., Wilkerson, E.G., Overhults, D.G. and Tinôco, I.F.F., Comparison between two systems for ammonia emission monitoring in broiler houses. In: Proceedings, International symposium on air quality and waste management for agriculture, Broomfield, Colorado, 2007.

[12] Medeiros, R., Santos, B.J.M., Freitas, M., Silva, O.A., Alves, F.F.y Ferreira, E., A adição de diferentes produtos químicos e o efeito da umidade na volatilização de amônia em cama de frango. Ciência Rural, 38(8), pp. 2321-2326, 2008. DOI: 10.1590/S0103-84782008000800035

[13] Vigoderis, R.B., Cordeiro, M.B., Tinôco, I.F.F., Menegali, I., Souza Júnior, J.P. y Holanda, M.C.R., Avaliação do uso de ventilação mínima em galpões avícolas e de sua influência no desempenho de aves de corte no período de inverno. Rev. Bras. Zootec., 39(6), pp. 1381-1386, 2010. DOI: 10.1590/S1516-35982010000600030

[14] Menegali, I., Tinoco, I.F.F., Carvalho, C.C.S., Souza, C.F. y Martins, J.H., Comportamento de variáveis climáticas em sistemas de ventilação mínima para produção de pintos de corte. Rev. bras. eng. agríc. ambient., 17(1), pp. 106-113, 2013. DOI: 10.1590/S1415-43662013000100015

[15] Hernandes, R. y Cazetta, J.O., Método simples e acessível para determinar amônia liberada pela cama aviária. Rev. Bras. Zootec., 30(3), pp. 824-829, 2001. DOI: 10.1590/S1516-35982001000300030

[16] Da Ros, C.O., Aita, C. y Giacomini, S.J., Volatilização de amônia com aplicação de uréia na superfície do solo, no sistema plantio direto. Ciência Rural, 35(4), pp. 799-805, 2005. DOI: 10.1590/S0103-84782005000400008

[17] Araujo, E.S., Marzola, T., Miyazawa, M., Soares, L.H.B., Urquiaga, S., Boddey, R.M. and Alves, B.J.R., Calibration of a semi-opened static chamber for the quantification of volatilized ammonia from soil. Pesq. agropec. bras., 44(7), pp. 769-776, 2009. DOI: 10.1590/S0100-204X2009000700018

[18] Alves, A.C., Oliveira, P.P.A., Herling, V.R., Trivelin, P.C.O., Luz, P.H.C., Alves, T.C., Rochetti, R.C. and Barioni-Júnior, W., New methods to quantify NH3 volatilization from fertilized surface soil with urea. Rev. Bras. Ciênc. Solo, 35(1), pp. 133-140, 2011. DOI: 10.1590/S0100-06832011000100012

[19] Lara-Cabezas, W.A.R. and Trivelin, P.C.O., N-NH_3 losses from nitrogen sources applied over unburned sugarcane straw. Rev. Bras. Ciênc. Solo, 14(1), pp. 345-352, 1990. DOI: 10.1590/S0100-06832003000400007

F. Campos-de Sousa, received her BSc. Eng. in Agricultural and Environmental Engineering in 2012, her MSc. in Agricultural Engineering in the area of Rural Ambience Constructions in 2014, both from the Federal University of Viçosa, Viçosa, Brazil. She is currently a doctoral student in the field of agricultural engineering and rural constructions, a member of AMBIAGRO (Center for Research in Agro-Industrial Ambience and Engineering Systems) at the Department of Agricultural Engineering, Federal University of Viçosa, Brazil.
ORCID: 0000-0002-5584-728X

I.deF. Ferreira-Tinôco, received her BSc. Eng. in Agricultural Engineering in 1980, her MSc. in Animal Structures in 1988, her DSc. in Animal Sciencesm in1996, all from the Federal University of Lavras, Minas Gerais state, Brazil. She is currently an associate professor at the Department of Agricultural Engineering of the Federal University of Viçosa, Brazil, and coordinates the UFV branch of the following scientific exchange programs: (1) CAPES/FIPSE University of Illinois and University of Purdue, U.S.A, (2) Umbrella Agreement between UFV and Iowa State University, and the University of Kentucky, (3) Scientific and Technical Agreement between UFV and University of Évora (Portugal), Universidad Nacional de Colombia (Colombia), Iowa State University and University of Kentucky (U.S.A.). She also coordinates the Center for Research in Animal Microenvironment and Agri-Industrial Systems Engineering (AMBIAGRO), and is the President of the DEA-UFV International Relations Committee.
ORCID: 0000-0002-8152-0301

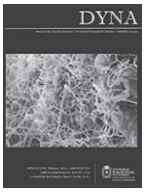
J.A. Osório-Saraz, received his BSc. Eng. in Agricultural Engineering in 1998, his Sp in Environmental Legislation in 2001 and his MSc. in Materials Engineering in 2006, all from the Universidad Nacional de Colombia, Medellín, Colombia. In 2011 he received a Dr. in Rural Constructions from the Federal University of Viçosa, Minas Gerais, Brazil. Since 2003 he has been a professor at the Universidad Nacional of Colombia, Medellín, Colombia, teaching and researching 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, Brazil and of the research group in agricultural engineering at the Universidad Nacional de Colombia.
ORCID: 0000-0003-4867-9158

K.S. Oliveira-Rocha, received he BSc. in Information Systems with an option in Computer Science, in 2005 from the University Center UNA, his MSc. in Agricultural Engineering in 2008, and PhD with an emphasis in Energy in Agriculture in 2012, all from the Department of Agricultural Engineering in the Federal University of Viçosa, Brazil. He has experience in the areas of instrumentation, process control, data acquisition using electronic devices addressable systems, numerical methods applied to engineering, simulation, grain drying and aeration, ventilation systems, machine vision, image processing. He is a member of AMBIAGRO (Center for Research in Agro-Industrial Ambience and Engineering Systems), at the Department of Agricultural Engineering, Federal University of Viçosa, Brazil.
ORCID: 0000-0003-4305-4837

M. Arredondo-Ramirez, received his BSc. Eng. in Agricultural Engineering in 2014, at the Universidad Nacional of Colombia, Medellín, Colombia and was then an intern at the UFV. He is currently a MSc. Engineering student in water resources. He is a member of OCEÁNICOS (Research Group about Simulation Software Sea and Engineering Systems from ocean), at the Facultad de Minas, Universidad Nacional de Colombia, Medellin, Colombia.
ORCID: 0000-0003-4096-1212

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS



DYNA
http://dyna.medellin.unal.edu.co/

Implementation of a non-conventional method to characterize voltage sags and swells

Jorge Enrique Celis-Montero [a], Diego Fernando-Navas [b] & Ferley Castro-Aranda [b]

[a] Facultad de Ingeniería, EIEE, GRALTA, Universidad del Valle, Cali, Colombia. jorge.celis@correounivalle.edu.co
[b] Facultad de Ingeniería, EIEE, GRALTA, Universidad del Valle, Cali, Colombia. diego.navas@correounivalle.edu.co
[c] Facultad de Ingeniería, EIEE, GRALTA, Universidad del Valle, Cali, Colombia. ferley.castro@correounivalle.edu.co

Received: July 21th, 2014. Received in revised form: March 14th, 2015. Accepted: December 15th, 2015.

Abstract
Voltage sags and swells are among the power quality disturbances that represent the biggest economic losses for affected users. It is therefore necessary to undertake a suitable characterization of those events to conduct studies that permit the causes and possible mitigation techniques to be identified. This work describes the development of monitoring modules to characterize voltage sags and swells that could be used in Electromagnetic Transients Programs - EMTP (for example the ATP – Alternative Transients Program). The implemented module uses a novel method to characterize these disturbances. The results of the implementation show that the voltage sags and swells are appropriately characterized; furthermore, less sampled data is required from a voltage signal with respect to the conventional RMS voltage method. This could optimize the capture and analysis process of information in power quality monitoring.

Keywords: Electromagnetic transient analysis, parameter estimation, power distribution faults, power quality, power system modeling, simulation, simulation software, time-domain analysis.

Implementación de un método no convencional para caracterizar hundimientos y elevaciones de tensión

Resumen
Dentro de los problemas de calidad de potencia que acarrean mayores pérdidas económicas sobre los usuarios afectados se encuentran los hundimientos y elevaciones de tensión. Debido a ello, se hace necesario caracterizarlos adecuadamente para emprender estudios que permitan identificar sus causas y posible formas de mitigación. En el presente trabajo se describe el desarrollo de un módulo de monitoreo para caracterizar hundimientos y elevaciones de tensión, el cual se puede usar en el Programa de Transitorios Electromagnéticos - EMTP (como ATP – Programa Alternativo de Transitorios). El módulo implementado usa un novedoso método para la caracterización de estos disturbios. Los resultados de la implementación muestran que los hundimientos y elevaciones son caracterizados adecuadamente, además, que el método requiere menos muestras de la señal de tensión respecto al método convencional de tensión RMS, lo cual podría optimizar el proceso de captura y análisis de la información durante el monitoreo de calidad de potencia.

Palabras clave: Análisis de transitorios electromagnéticos, análisis en el dominio del tiempo, calidad de potencia, estimación de parámetros, fallas en sistemas de distribución, modelado de sistema de potencia, simulaciones, software de simulación.

1. Introduction

For the development of smart grids, efficiency in detecting events of power quality is an important task in monitoring the power system.

The power quality disturbances that currently cause the greatest damage to industrial and commercial users are voltage sags [1-3]. A voltage sag is the sudden reduction of voltage at a particular point in an electricity supply system below a specified threshold, followed by its recovery after a brief interval between half cycle and 1 min [3].

When these disturbances are manifested, many industrial processes are halted because of inadequate operation of electronic equipment [4] like programmable logic controllers

(PLC), computers and variable speed drives [6], as well electric induction motors [7]. Furthermore, controlled bridge rectifiers, generally used in DC motor speed controls, as well as in HVDC links, are also affected by voltage sags [8].

Sags are usually accompanied by voltage swells, which are sudden voltage increases above a threshold, between 110% and 180% of nominal voltage [3]. Both are caused by faults in electrical supply lines and are produced either by short circuits or by insulation faults due to lightning strokes [9], the starting up of large induction motors [10], operation of arc furnaces, or the load transfer between power supplies [11].

Current standards of power quality are quite precise in defining the necessary parameters to characterize voltage sags and swells [12]. The IEC 61000-4-30 standard [13] defines magnitude and duration as the sole parameters to characterize these disturbances.

Nevertheless, magnitude and duration are not the only parameters of voltage sags and/or swells that influence equipment behavior. The imbalance of sag and/or swell based on the type of fault (single-phase, phase-to-phase, or three-phase), the missing voltage, phase angle or point-on-wave, and phase shift are other parameters that also differentiate one event from another and that have a distinct impact on susceptible equipment [13,14].

Therefore, there is a needed to employ additional methods that characterize non-standard parameters of voltage sags and swells in order to evaluate other effects. There are other non-conventional methods that have been used to characterize some voltage sags and swells parameters.

The fundamental voltage component uses the Sliding Discrete Fourier Transform (SDFT) method to analyze non-stationary signals including voltage sags and swells [15,16].

The DQ transformation, also known as Park's transformation, is commonly used when modeling synchronous and induction motors. This transform has been used to study several power quality disturbances [17,18]..It is possible given that the three-phase AC voltages can be represented as DC voltages when applying the DQ transformation through which any disturbance in AC voltage is reflected as a disturbance in DQ values.

The Waveform Envelope method [19], is an approximate alternative that seeks to determine the initial and final times of a voltage sag, as of the waveform of the instantaneous voltage. However, other characteristics such as magnitude or phase shift cannot be obtained by this method.

The Kalman Filter is a method based on models that has been used in power systems to trace harmonics in real time [20], to estimate voltage and current parameters in protection systems and to estimate transient parameters [21]. There are at least two models to analyze voltage sags and swells with this method [22]: the linear model, which considers the magnitudes and angles of each harmonic component as state variables, and the extended model, which considers the frequency and the state variables of the linear model as its state variables.

The Wavelet Transform (WT) is a useful mathematical tool to analyze non-stationary signals [23]. It has been used to analyze electrical signals in power systems [24]. To study the voltage sags and swells, the multi-resolution analysis (MRA) is quite useful. The detail components provide information about the beginning and end of the said disturbances while the magnitude can be determined by two methods: the Discrete Wavelet Transform (DWT) coefficients method and the phasor method [22].

The Numerical Matrix method is generally used as a voltage sag detection technique for a Dynamic Voltage Restorer (DVR) [25]; however, the method is so adaptable and easy to apply that it could be used to characterize voltage sags and swells parameters.

This paper presents the implementation of Numerical Matrix method to characterize voltage sags and swells in EMTP.

This document is organized as follows. In Section II, the Numerical Matrix method is explained. In Section III, standard voltage sag signals are used to compare the performance of some characterization methods, including the Numerical Matrix method. Section IV describes the method implementation in EMTP and its performance in analyzing some events that were simulated in the IEEE 13-node test feeder. Finally, Section V presents this paper's conclusions.

2. Numerical Matrix

This method has been used as part of the DVR's control system [25]. It has small response times and yields results that can be directly interpreted. As such, the method is applied to each phase independently and it, monitors the start and end of the voltage sag, the remaining voltage and the phase shift.

The method involves sampling the voltage signal and storing it in matrix format. Whenever the dominant frequency components from the source are known, the following equations may be used:

$$V_{p1} = |V_1|\cos(\omega_1 t + \phi_1) + |V_5|\cos(\omega_5 t + \phi_5) \quad (1)$$

$$V_{p2} = |V_1|\cos(\omega_1 t - \omega_1 T_s + \phi_1) + |V_5|\cos(\omega_5 t - \omega_5 T_s + \phi_5) \quad (2)$$

$$V_{p3} = |V_1|\cos(\omega_1 t - 2\omega_1 T_s + \phi_1) + |V_5|\cos(\omega_5 t - 2\omega_5 T_s + \phi_5) \quad (3)$$

$$V_{p4} = |V_1|\cos(\omega_1 t - 3\omega_1 T_s + \phi_1) + |V_5|\cos(\omega_5 t - 3\omega_5 T_s + \phi_5) \quad (4)$$

where V_{p1} is the current sample of the voltage; V_{p2}, V_{p3}, and V_{p4} are the samples of the voltage in prior sampling periods, $|V_1|$ and $|V_5|$ are the magnitudes of the fundamental and 5th harmonic, $\Phi 1$ and $\Phi 5$ are the corresponding phase angles, $\omega 1$ and $\omega 5$ are the angular frequencies of the fundamental and 5th harmonic, and Ts is the signal's sampling period.

Depending on the number of harmonics detected in the voltage signal, the number of equations to be solved can decrease or increase so that the number of simultaneous equations required to find a solution is always double the total number of frequencies considered.

If the dominant harmonics are not included in the equations, or if they are unknown, the solution can contain errors. This

denotes the sensitivity of the matrix, which is related to the size of the error [25]. This is how the sensitivity of the matrix increases with the order of the unknown harmonics. It has also been found that the sensitivity increases if the sampling period (T_s) is very small.

If we reorganize (1), (2), (3) and (4) in matrix format, we obtain:

$$\underbrace{\begin{bmatrix} V_{p1} \\ V_{p2} \\ V_{p3} \\ V_{p4} \end{bmatrix}}_{b} = \underbrace{\begin{bmatrix} 1 & 0 & 1 & 0 \\ \cos\omega_1 T_s & \sin\omega_1 T_s & \cos\omega_5 T_s & \sin\omega_5 T_s \\ \cos 2\omega_1 T_s & \sin 2\omega_1 T_s & \cos 2\omega_5 T_s & \sin 2\omega_5 T_s \\ \cos 3\omega_1 T_s & \sin 3\omega_1 T_s & \cos 3\omega_5 T_s & \sin 3\omega_5 T_s \end{bmatrix}}_{A} \cdot \underbrace{\begin{bmatrix} |V_1|\cos\phi_1 \\ |V_1|\sin\phi_1 \\ |V_5|\cos\phi_5 \\ |V_5|\sin\phi_5 \end{bmatrix}}_{x} \quad (5)$$

Knowing the voltage data (V_{p1-4}) and the relation matrix "A", it is possible to determine the magnitudes and phases of the frequency components considered in the model.

The response time to detect changes in the magnitude and phase of the fundamental component depend on the sampling period and on the number of dominant harmonic components that are present in the voltage signal.

Finally, this method is useful to characterize voltage sags and swells, given that it delivers magnitude and phase shift directly. Also, it requires less samples of the voltage in respect to the standard RMS voltage method, which could be an advantage for the time sequence of events.

3. Comparison of characterization methods

The fundamental voltage component (employing SDFT), Kalman filter, Wavelet Transform, and Numerical Matrix are methods that could determine magnitude, duration and phase shift of voltage sags and swells. The standard RMS voltage method [13] only characterizes magnitude and duration of those events; however, it is the most commonly used [26]. The aforementioned methods can be compared to determine which one that delivers these parameters (magnitude and duration) with the least error. We must note that due to the lack of a standard procedure to measure the phase shift, only magnitude and duration were compared for the methods analyzed.

To conduct the comparison, a basic implementation of the selected methods was carried out: we used a sliding window of a half cycle for the RMS voltage method; for the Kalman filter method, we used the extended model [22]; for the Wavelet transform method, we used Daubechies 6 as "mother wavelet" for first level MRA analysis [27], and we used the phasor method to determine the magnitude of voltage; for the Numerical matrix method, we assumed fundamental component and 5th harmonic.

With the basic implementation of the methods described above, we were able to analyze standard signals of known magnitudes and durations. The standard signal was a 60-Hz pure sinusoidal wave, 1 per unit of nominal voltage, and a sampling period of 65.1 μs (256 samples per cycle). Waves of different magnitudes were superimposed on this signal (from 10 to 90% in 10 steps), and different points on the wave (from 0° to 360º in steps of 10°) for each magnitude were superimposed. The phase angle was always conserved, which is equivalent to simulating different voltage sags without phase shift.

The relative error of magnitude was calculated from the magnitude of the known standard signal and the magnitude estimated through each method by considering the detection methodology described in the IEC 61000-4-30 standard [13].

According to the results, the relative errors of magnitude for the RMS voltage and the fundamental component (SDFT) methods were negligible. The error in estimating the magnitude with the Kalman filter method can be observed in Fig. 1, and the error of magnitude obtained with the wavelet transform method is shown in Fig. 2.

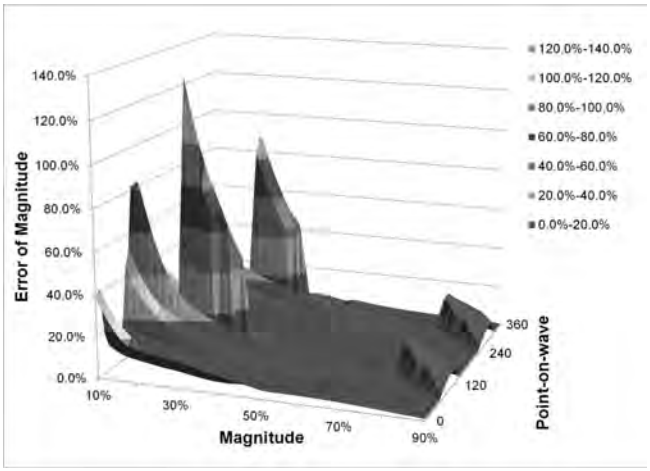


Figure 1. The Kalman Filter method's error of magnitude when analyzing standard signals.
Source: The authors.

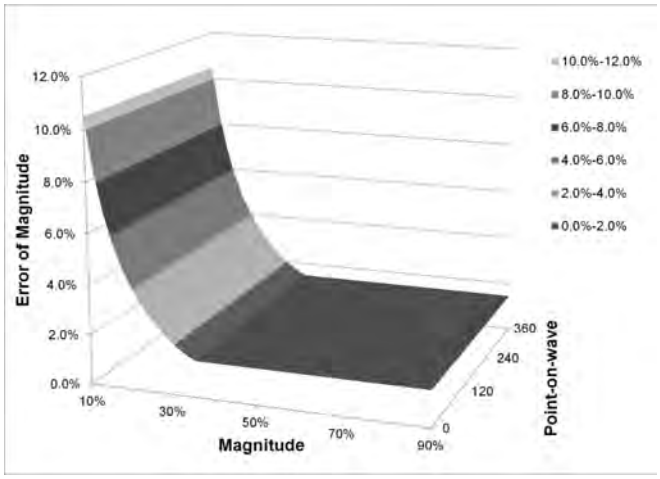


Figure 2. The wavelet transform method's error of magnitude when analyzing standard signals.
Source: The authors.

Fig. 3 shows the error of magnitude estimated with the numerical matrix method. According to the figure, the error obtained with this method can be considered negligible.

Fig. 4 shows the relative error of duration obtained with the RMS voltage method. It was calculated using the standard signal's known data of duration , as in the case of error of magnitude. Only those methods that yielded the lowest error in estimating the duration were compared.

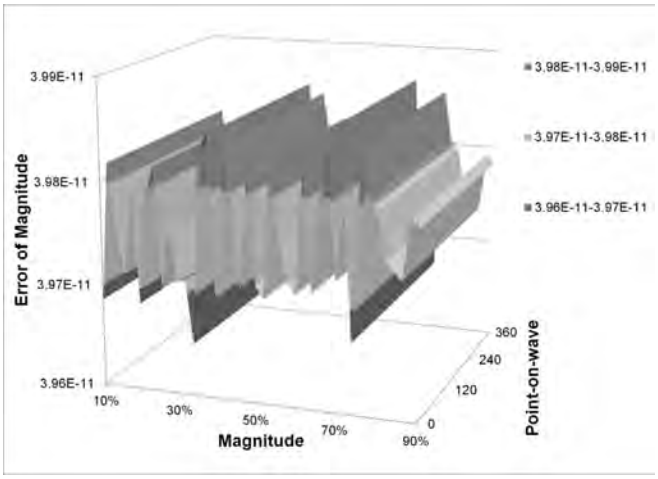


Figure 3. The Numerical Matrix method's error of magnitude when analyzing standard signals.
Source: The authors.

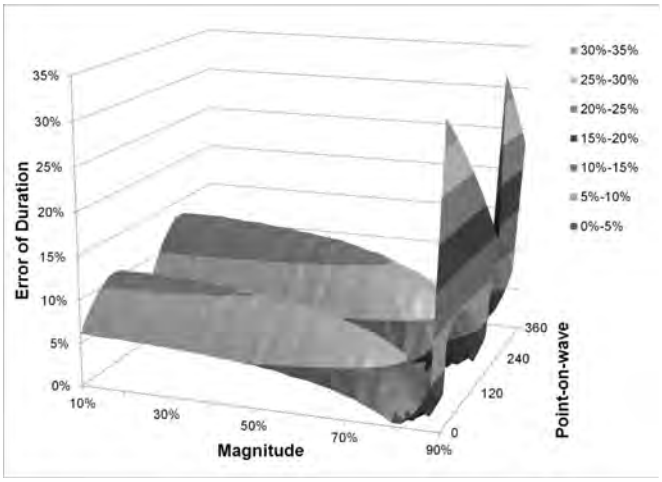


Figure 4. The RMS Voltage method's error of duration of when analyzing standard signals.
Source: The authors.

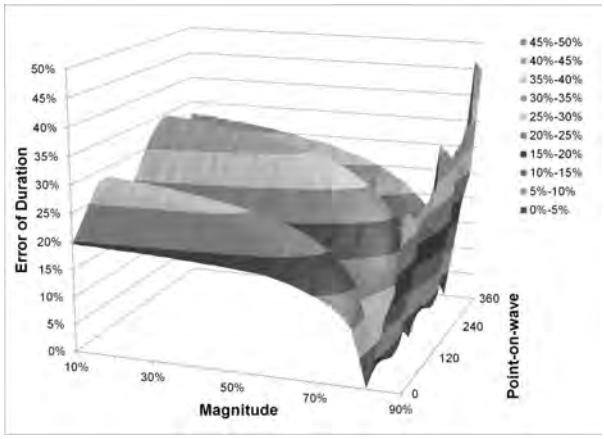


Figure 5. The Voltage Fundamental Component method's (SDFT) error of duration of when analyzing standard signals.
Source: The authors.

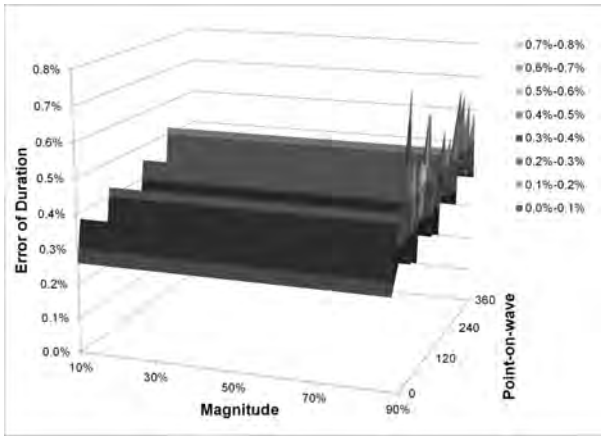


Figure 6. The Numerical Matrix method's error of duration when analyzing standard signals.
Source: The authors.

The relative error of duration obtained with the fundamental voltage component method (SDFT) is shown in Fig. 5. It can be seen that the error is higher than with the RMS voltage method; this is because the SDFT needs to analyze a whole period of the signal to deliver a result.

Finally, Fig. 6 shows the lowest error in estimating duration, which is obtained with the Numerical Matrix method.

When comparing results the Numerical Matrix method is suitable to characterize voltage sags and swells because it has the lowest error in estimating magnitude and duration of these disturbances. As such, this method has been selected to be used in the electromagnetic transient analysis software EMTP.

4 Implementation of numerical matrix method in EMTP

By comparing the results yielded by existing voltage sag and swell characterization methods, it was possible to determine that the Numerical Matrix method delivers the least errors in estimating the magnitude and duration of these disturbances. Therefore, this method will be adapted to create a monitoring module to characterize voltage sags and swells.

The monitoring module must be able to analyze a sinusoidal-type signal, of any magnitude, frequency, and phase angle; a signal that must be delivered by the simulation tool employed. The results obtained by this module must be the sum of events occurring during the time of signal analysis indicating the type of disturbance, the magnitude and duration, the graphic of the magnitude of the fundamental component and additionally, the graphic of the phase shift during the period analyzed.

The algorithm developed for the monitoring module must include the execution of the Numerical Matrix method over simulation time. As was previously explained, this method delivers the magnitude of the fundamental voltage of the signal analyzed. Thereafter, the algorithm must carry out, in each simulation step, the comparison of the voltage magnitude with the detection threshold; it must register the instants during which that threshold is surpassed and, thus,

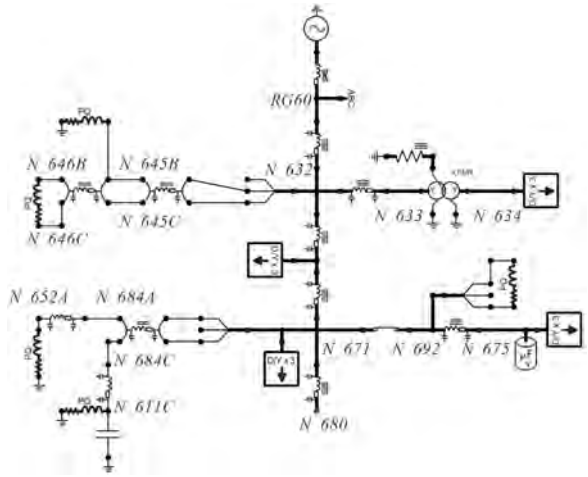


Figure 7. IEEE 13 node reference system implemented using EMTP (ATP).
Source: [28]

determine the duration of the disturbance. Finally, the algorithm must deliver the record of the events found along with the characteristics of magnitude and duration.

The monitoring module was developed using the transient analysis tool EMTP (ATP), and the programming of the algorithm was carried out using FORTRAN.

For the corresponding test of the monitoring module that was implemented, we modeled the IEEE 13-node test feeder [28]. However, the voltage regulator was omitted and all loads were modeled as constant impedances (Fig. 7).

The N_675 node was chosen as the measuring node on the reference system while the failures were generated on N_632 node.

With a sampling time of 1.302E-4 s (128 samples per cycle at 60-Hz), the voltage signals in the measuring node were recorded when a single-phase-to-ground fault of 0.6Ω-impedance and 74 milliseconds in duration was generated in N_632 node. Fig. 8 shows the resulting signals and it is evident that there is a voltage sag on phase C along with a slight increase in the maximum voltage in other phases.

The monitoring module was localized in the measuring node for characterizing the voltage sags and swells.

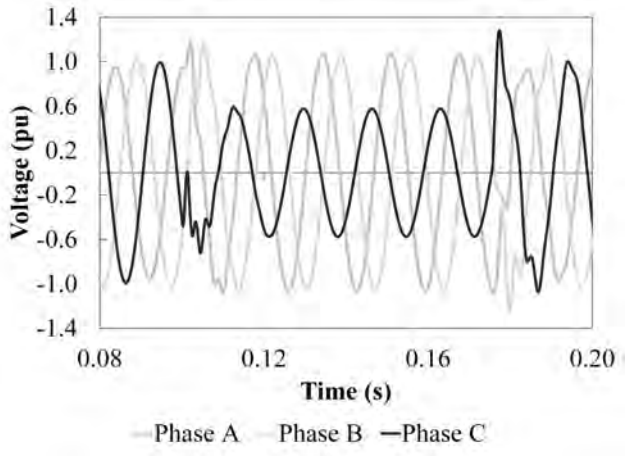


Figure 8. Instantaneous voltages in the measuring point.
Source: The authors.

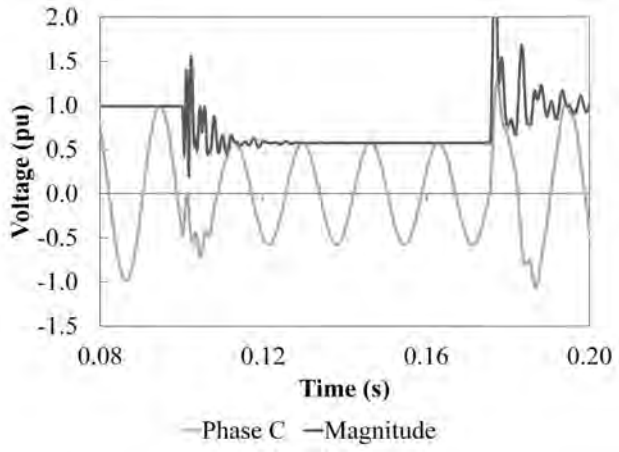


Figure 9. Magnitude of the fundamental voltage of phase C from node N_675, obtained with the monitoring module.
Source: The authors.

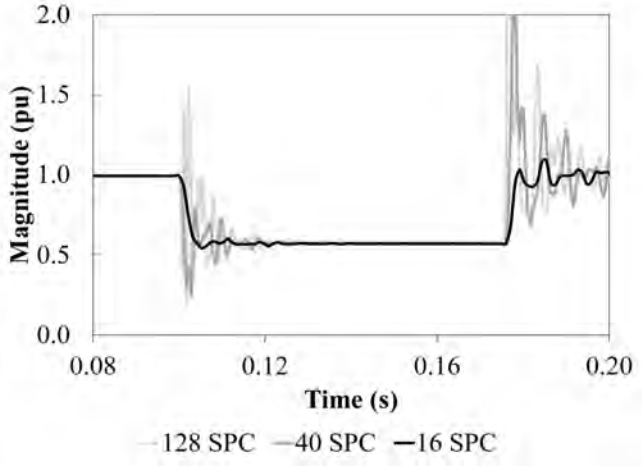


Figure 10. Magnitudes of the fundamental voltage of phase C considering different sampling times (SPC: Samples per cycle).
Source: The authors.

Fig. 9 shows the comparison of the voltage input signal of phase C with the magnitude of the fundamental voltage given by the monitoring module. As seen in the figure, there are oscillations of fundamental voltage magnitude at the beginning and end of the voltage sag. These are related to the low-frequency oscillatory transients that the voltage signal presents in those instants.

As the sensitivity of the matrix in the numerical matrix method depends on the amount of harmonics in the model and the sampling period, the monitoring module was adjusted to include four harmonic components in the model (fundamental, 5th, 7th and 11th harmonics), and the sampling times were varied to reduce the error in the results.

Fig. 10 shows the magnitudes of fundamental voltages obtained for the stated case while considering different sampling times. As can be verified, as the sampling times increase, the oscillations in the magnitude are reduced.

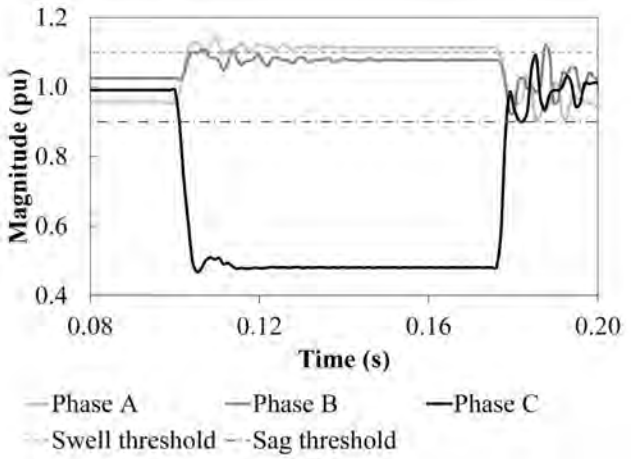


Figure 11. Magnitudes of N_675 node's voltages for a single-phase fault on N_671 node.
Source: The authors.

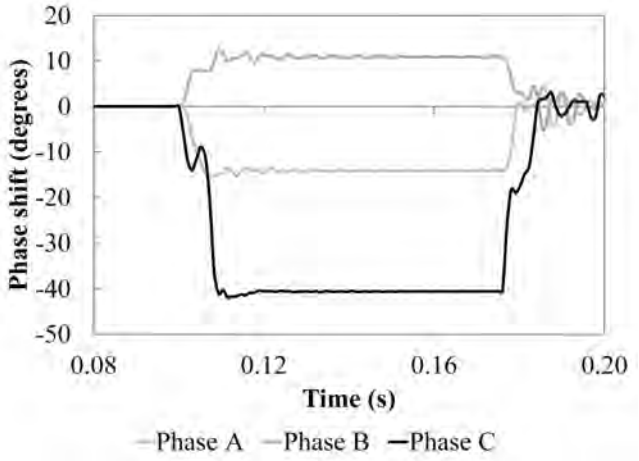


Figure 12. Phase shift for the voltages of the N_675 node, obtained with the monitoring module.
Source: The authors.

Accordingly, we chose a 1.04167E-3 s sampling time (16 samples per cycle) to analyze the various types of faults in the reference system. Fig. 11 shows the magnitudes of the three-phase voltages in the measuring node when a single-phase fault exists on phase C of the N_671 node with a fault impedance of 0.7 Ω. In this case, a voltage sag in phase C and a voltage swell in phase A were observed.

Fig. 12 shows the phase shift of each phase given by the monitoring module. This parameter was calculated independently with respect to the angle of each signal.

To determine if the results obtained by the monitoring module were accurate, the relative error percentage was found, which was calculated from "real values". To establish the real duration of each event, the time difference between the beginning and end of the fault current was determined, while the result given by the RMS voltage method according to IEC 61000-4-30 [13] was used for the magnitude.

Table 1 shows the results given by the monitoring module for various simulated cases.

Table 1.
Comparison of results of the monitoring module implemented in EMTP

Disturbance	Event detected	EMTP Result	Real Result	Relative Error
N_671: Ph-Gr in phase C Zf = 0.7 Ω	Ph. A: Swell	112.63%	113.83%	-1.055%
		68.75 ms	76.04 ms	-9.589%
	Ph. B: No event	NE	NE	NE
		NE	NE	NE
	Ph. C: Sag	47.70%	47.80%	-0.221%
		76.04 ms	76.04 ms	0%
N_671: Ph-Ph in phases AB Zf = 1.0 Ω	Ph. A: Sag	73.77%	73.06%	0.968%
		81.25 ms	80.21 ms	1.299%
	Ph. B: Sag	40.30%	40.61%	-0.765%
		84.38 ms	80.21 ms	5.195%
	Ph. C: Swell	118.20%	118.15%	0.047%
		73.96 ms	80.21 ms	-7.792%
N_633: Ph-Ph-Gr in phases CA Zf = 0.5 Ω	Ph. A: Sag	44.33%	44.39%	-0.137%
		81.25 ms	80.21 ms	1.299%
	Ph. B: Swell	112.68%	113.09%	-0.362%
		70.84 ms	80.21 ms	-11.688%
	Ph. C: Sag	63.36%	64.27%	-1.419%
		76.04 ms	80.21 ms	-5.195%

Ph-Gr: Phase-to-ground fault
Ph-Ph: Phase-to-phase fault
Ph-Ph-Gr: Phase-to-phase-to-ground fault
Zf: Fault impedance
NE: No Event
Source: The authors.

From the results shown in this table, it can be stated that the monitoring module implemented in EMTP furnishes results with very low errors; hence, it is quite effective in characterizing these disturbances.

5. Conclusions

It is necessary to study other characteristics of voltage sags and swells, which will be useful to develop an analysis of the causes and the impacts they have. Phase shift may be considered to be another characteristic that could be studied. However, no standard procedure exists to determine the phase shift presented by voltage when it is affected by sag or swell. In terms of future work, a methodology must be established to measure this characteristic, as well as the identification of the correlation between phase shift and fault cause.

The monitoring module that was developed to characterize voltage sags and swells showed good results for the analysis of simulated signals. However, the accuracy of the results depends on the presence of noise in the signals that were analyzed. Therefore, if we want to implement it in a real system we must consider additional possibilities in real time signal processing, such as applying low pass filters, establishing the predominant harmonics of voltage signals at the start of the measurements, and setting the sampling time according to the noise signal present, etc.

The advantage of using the Numerical Matrix method to characterize voltage sags and swells is that it requires less data from voltage signals in comparison to the RMS voltage

conventional method. This means that the process of capture, analysis and storage of information in power quality monitoring can be accelerated.

In terms of future work, the monitoring module could be complemented by analyzing real signals, and the results compared with those obtained by conventional measurement equipment.

References

[1] Waskito, F. and Banmongkol, C., Simulation of the voltage sag effects on an induction motor, International Conference on Consumer Electronics, Communications and Networks (CECNet), pp. 731-734, 2011. DOI: 10.1109/CECNET.2011.5768900.

[2] Villate-Martinez, J. y Garcia-Tejedor, J., La calidad del suministro eléctrico y su mejora mediante convertidores de potencia. Revista DYNA, 79(3), pp. 30-34, 2004.

[3] Yamamoto, K., Ikeda, K., Tsurusaki, Y. and Ikeda, M., Characteristics of voltage Sag/Swell compensator utilizing single-phase matrix converter, International Conference on Electrical Machines and Systems, pp. 1863-1868, 2013. DOI: 10.1109/ICEMS.2013.6713300.

[4] IEC. Electromagnetic compatibility (EMC). Environment. Voltage dips and short interruptions on public electric power supply systems with statistical measurement results. PD IEC/TR 61000-2-8, 1, Geneva: IEC, 2002.

[5] Gil-Montoya, F., Manzano-Agugliaro, F., Gómez-Lopez, J. y Sánchez-Alguacil, P., Técnicas de investigación en calidad eléctrica: Ventajas e inconvenientes. Revista DYNA, 79(173), pp. 66-74, 2012.

[6] Almeida, C. and Kagan, N., Using genetic algorithms and Fuzzy programming to monitor voltage sags and swells. IEEE Intelligent Systems [online]. 26(2), 2011. [date of reference July 3th 2014]. Available at: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5696715. DOI: 10.1109/MIS.2011.2.

[7] Gonçalves, J., Baptista, J., Neves, L. and Tadeu-Oliveira. F., Simulation of the effect of voltage transients on an induction motor with ATP / EMTP, International Conference on Renewable Energies and Power Quality (ICREPQ), pp. 63-69, 2009.

[8] Dofnas, L., A novel method to mitigate commutation failures in HVDC systems, International Conference on Power System Technology (PowerCon), pp. 51-56, 2002. DOI: 10.1109/ICPST.2002.1053503.

[9] Bollen, M.H.J., Voltage sags characterisation. Understanding power quality problems: Voltage sags and interruptions, 1st ed. New York, USA, Wiley, 2000, pp. 139-168.

[10] Gomez, J.C, Reineri, C., Campetelli, G. and Morcos, M.M., A study of voltage sags generated by induction motor starting, Electric. Power Components and Systems, 32(6), pp. 645-653, 2010. DOI: 10.1080/15325000490228423.

[11] Chen, C., Wang, H. and Chin, Y., A simple rule-based approach for detection and classification of voltage sag, swell, and interruption in power systems, International Conference on Power Electronics and Drive Systems, (IEEE PEDS), pp. 5-8, 2011. DOI: 10.1109/PEDS.2011.6147338.

[12] IEEE. Recommended practice for monitoring electric power quality. IEEE Std 1159, 2009.

[13] IEC. Electromagnetic compatibility (EMC) Part 4-30: Testing and measurement techniques – Power quality measurement methods, IEC 61000-4-30, 2008.

[14] IEEE. Recommended practice for the design of reliable industrial and commercial power systems, IEEE Std 493, 2007.

[15] Bollen, M. and Gu, I., Processing os stationary signals, signal processing of power quality disturbances, 1st ed. New York, USA, Wiley, 2006, pp. 222-226.

[16] Jacobsen, E. and Lyons, R., The sliding DFT, IEEE Signal Processing Magazine. [online]. 20(2), 2003. [date of reference July 3th 2014]. Available at: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=01184347. DOI: 10.1109/MSP.2003.1184347.

[17] Montero-Hernandez, O.C. and Enjeti, P.N., A Detection algorithm suitable for mitigation of numerous power quality disturbances. IEEE Transactions on Industry Applications [online]. 41(6), 2005. [date of reference July 3th of 2014]. Available at: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1542324. DOI: 10.1109/IAS.2001.955994.

[18] Zhao, Y., Yonghai, X., Xiangning, X., Yongqiang, Z. and Chunlin, G., Power quality disturbances identification based on DQ conversion, Wavelet Transform and FFT, Asia-Pacific Power and Energy Engineering Conference (APPEEC), pp. 1-4, 2010. DOI: 10.1109/APPEEC.2010.5448526.

[19] Djokic, S.Z., Milanovic, J.V. and Rowland, S.M., Advanced voltage sag characterisation II: Point on wave. IET Generation, Transmission & Distribution [online]. 1(1), 2007 [date of reference July 3th of 2014]. Available at: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=04082380. DOI: 10.1049/iet-gtd:20050434.

[20] Mostafa, M.A., Kalman filtering algorithm for electric power quality analysis: Harmonics and voltage sags problems, Large Engineering Systems Conference Power Engineering, pp. 159-165, 2007. DOI: 10.1109/LESCPE.2007.4437371.

[21] González, M. and Cardenas, V., The kalman filter in power quality – theory and applications, Kalman Filter. 1st ed. USA, InTecch, 2010, pp. 101-125. DOI: 10.5772/9582.

[22] Perez-Fernandez, E., Nuevo método de detección y análisis en tiempo real de eventos en la tensión de suministro de energía eléctrica empleando un modelo combinado wavelets-filtro de Kalman extendido, Tesis Doctorado en Ingeniería, Departamento de Ingeniería Informática y Electrónica, Universidad de Cantabria, Santander, España, 2006.

[23] Christy, X., Vedamani, J., and Karthikeyan, S., Wavelet based detection of power quality disturbance - A case study, International Conference on Signal Processing, Communication., Computing and Networking Technologies (ICSCCN), pp. 157-162, 2011. DOI: 10.1109/ICSCCN.2011.6024534.

[24] Gallego-Tejada, J and Castro-Aranda, F., Análisis de señales por transformada Wavelet. La Transformada Wavelet aplicada a los sistemas eléctricos de potencia: Estado del Arte. 1st ed. Alemania, Editorial Académica Española, 2011, pp. 40-63.

[25] Fitzer, C., Barnes, M. and Green, P., Voltage sag detection technique for a dynamic voltage restorer. IEEE Transactions on Industry Applications [online]. 40(1), 2004. [date of reference July 3st of 2014]. Available at: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1268198. DOI: 10.1109/TIA.2003.821801.

[26] Flores-Arias, J, Bellido-Outeiriño, J. y Moreno-Muñoz, A., A fast RMS meter for detecting sag events in household environments, International Conference on Consumer Electronics (ICCE), pp. 321-322, 2014. DOI: 10.1109/ICCE.2014.6776024.

[27] Parsons, A.C, Grady, W.M. and Powers, E.J., A wavelet-based procedure for automatically determining the beginning and end of transmission system voltage sags, IEEE Power Engineering Society, pp. 1310-1315, 1999. DOI: 10.1109/PESW.1999.747406.

[28] Distribution test feeders working group, 13-Bus Test Feeder [online]. Available at: http://ewh.ieee.org/soc/pes/dsacom/testfeeders/.

J.E. Celis-Montero, received his BSc. in Electrical Engineering in 2009, from the Universidad del Valle, Santiago de Cali, Colombia. He is currently undertaking his PhD. in Electrical and Electronic Engineering at the Universidad del Valle, Cali, Colombia. He has worked as a researcher at the High Voltage Research Group - GRALTA, Universidad del Valle, Cali, Colombia. His research interests focus on power quality, modeling and the simulation of power system transients, transformers, ferroresonance and smart grids.
ORCID: 0000-0003-2488-8238

D. Fernando-Navas, received his BSc. in Electronic Technology in 2002, from the Universidad del Valle, Palmira, Colombia and his BSc. and MSc. in Electrical Engineering in 2009 and 2012 respectively, from the Universidad del Valle, Santiago de Cali, Colombia. He has worked as a

researcher at the Grupo de Investigación en Alta Tensión - GRALTA and has been an active part of the Laboratorio de Alta Tensión at the Universidad del Valle, Cali, Colombia. He is currently an assistant professor in the Faculty of Engineering of the Electric and Electronic Engineering School at the Universidad del Valle (EIEE). His areas of interest are electricity markets, transformers, electrical equipment testing and vegetable oils. He is a member of the Committee 130 - Transformers ICONTEC Colombia.
ORCID: 0000-0002-7316-7368

F. Castro-Aranda, received his BSc. and MSc. in 1992 and 1995 respectively, in Electrical Engineering from the Universidad del Valle, Santiago de Cali, Colombia and his PhD. in 2005 from the Universitat Politècnica de Catalunya, Spain. He is an associate professor at the Electric and Electronic Engineering School at the Universidad del Valle - EIEE (Santiago de Cali, Colombia). His research interests are focused on the areas of insulation coordination and systems modeling for transient analysis using EMTP. He is President of the Committee 130 - Transformers ICONTEC Colombia, and Director of the High Voltage Lab at the Universidad del Valle.
ORCID: 0000-0002-4858-0222

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS

DYNA
http://dyna.medellin.unal.edu.co/

Influence of strain rate and heat treatments on tensile and creep properties of Zn-0.15Cu-0.07Ti alloys

María José Quintana [a], José Ovidio García [b], Roberto González [c] & José Ignacio Verdeja [d]

[a] *Facultad de Ingeniería, Universidad Panamericana, México, D.F., México. mquintana@up.edu.mx*
[b] *Escuela de Ingeniería de Minas, Energía y Materiales de Oviedo, Universidad de Oviedo, Oviedo, España. jovidio@uniovi.es*
[c] *Facultad de Ingeniería, Universidad Panamericana, México, D.F., México. robglez@up.edu.mx*
[d] *Escuela de Ingeniería de Minas, Energía y Materiales de Oviedo, Universidad de Oviedo, Oviedo, España. metalotecnia@etsimo.uniovi.es*

Received: August 14th, 2014. Received in revised form: December 10th, 2015. Accepted: January 10th, 2016.

Abstract
The use of Cu and Ti in Zn alloys improves mechanical properties as solid solution and dispersoid particles (grain refiners) may harden the material and reduce creep deformation. This is one of the main design problems for parts made with Zn alloys, even at room temperature. In this work the mechanical behavior of a Zn-Cu-Ti low alloy is presented using tensile tests at different strain rates, as well as creep tests at different loads to obtain the value of the strain rate coefficient *m* in samples parallel and perpendicular to the rolling direction of the Zn strip. The microstructure of the alloy in its raw state, as well as heat treated at 250°C, is also analyzed, as the banded structure produced by rolling influences the strengthening mechanisms that can be achieved through the treatment parameters.

Keywords: Zn alloys; creep; strain rate coefficient; solid solution hardening; texture.

Influencia de la velocidad de deformación y tratamientos térmicos en las propiedades de tensión y fluencia del Zn-0.15Cu-0.07Ti

Resumen
El uso de Cu y Ti como aleantes del Zn mejora las propiedades mecánicas a través de solución sólida y partículas dispersas (promoviendo el afino de grano), resultando en materiales duros y resistentes al creep, lo que es, incluso a temperatura ambiente, uno de los principales problemas al diseñar piezas con aleaciones de Zn. El trabajo presenta el comportamiento mecánico de una aleación Zn-Cu-Ti a través de pruebas de tensión a diferentes velocidades de deformación, así como ensayos de fluencia a diferentes cargas para obtener el valor del coeficiente *m* de velocidad de deformación tanto en muestras paralelas a la dirección de laminación del material como en muestras perpendiculares. También se analizó la microestructura de la aleación en su estado de laminación, así como muestras tratadas a 250°C, ya que la estructura bandeada producto del rolado influye en los mecanismos de endurecimiento que se pueden alcanzar a través del tratamiento térmico.

Palabras clave: aleaciones de Zn; fluencia; coeficiente de velocidad de deformación; endurecimiento por solución sólida; textura.

1. Introduction

Zn alloys microalloyed with Cu and Ti are mainly used by the construction industry in roofing, gutters and drains [1]. The material is usually supplied as strips that, after solidification, have undergone hot rolling (between 250 and 100 °C). In some cases, room temperature rolling is also performed. Concepts of hot and cold rolling must be carefully considered because at 300 K (27 °C) Zn alloys have a temperature equivalent to 0.43 T_M (the melting temperature is 649 K), which makes them behave like metals deformed at high temperatures (dynamic restoration, creep, static recrystallization after straining, etc.) [2].

A specific requirement for this material is its creep strength [3] when loaded at room temperature. For example, with stresses higher than 90 MPa, stationary strain rate must not reach or exceed $1 \times 10^{-7}\ s^{-1}$ in both parallel and perpendicular

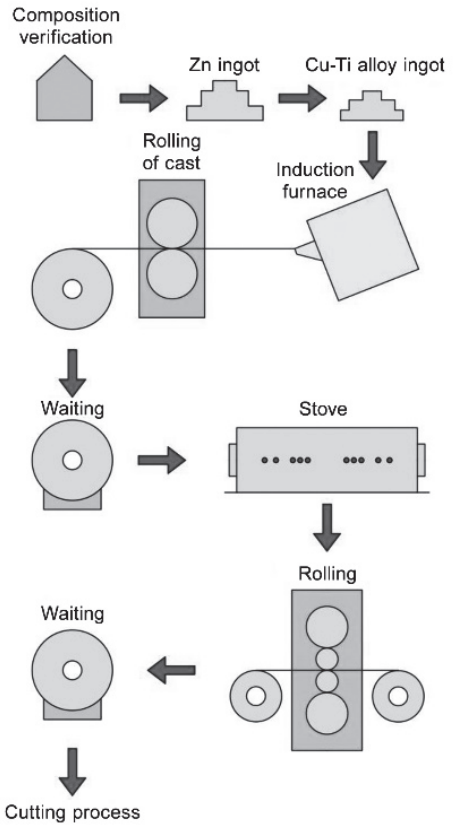


Figure 1. Schematic representation of the rolling process for a 0.7 mm strip at the ASLA facilities
Source: The authors

directions compared to the rolling axis (in strips with thickness of 0.7 mm). Adding the microalloying elements Cu and Ti improves hot deformation strength by solid solution of Cu in Zn or by grain refining of the dispersoid $Zn_{16}Ti$, which is formed during solidification [4-6]. The tensile and creep properties of this alloy are the product of hardening by grain refinement due to $Zn_{16}Ti$ [7-9], solid solution hardening by Cu in Zn [10-12] and also the increase in resistance in the direction perpendicular to the rolling axis caused by texture hardening [13-17].

The material analyzed in this work was manufactured by "Asturiana de Laminados (ASLA)" in facilities located at Asturias, Spain, using continuous casting equipment of the "Twin roll casting" type, which produces 8 mm thick slabs in the form of coils (Fig. 1).

After a 10-hour homogenization treatment at 250°C, hot rolling is undertaken in a reversible "Steckel" train at temperatures between 250 and 100°C in 5 phases. A final thickness of 0.7 mm is obtained (Fig. 2).

As the industrial application of parts manufactured with Zn alloys is conditioned by room temperature creep, this work analyzes the behavior of a Zn – Cu – Ti alloy that has been deformed at room temperature in tensile tests (different strain rates) or applied with a constant load (creep tests) in both the raw state and heat treated conditions. The anisotropy of the rolled strip is also considered.

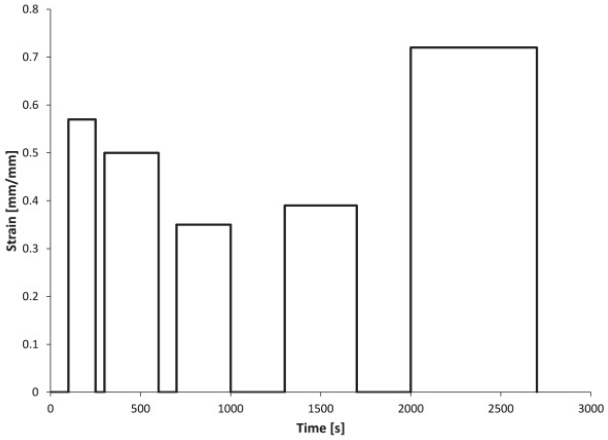


Figure 2. Strain amount during the rolling passes of 0.7 mm strips
Source: The authors

2. Experimental procedure

The strip produced by ASLA that is 0.7 mm thick is a Zn-0.15% Cu- 0.07% Ti used in industrial applications. Samples for mechanical testing were obtained from the Zn strip in both parallel and perpendicular directions relative to the rolling axis and were then machined as tension test samples with a calibrated gage length of l_0 = 100 mm (ASTM E8-04). The samples were tested using an Instron 5583 standard universal machine, while varying crosshead speeds (Table 1) until total fracture of the specimens. Also, creep tests were performed in samples with the same geometries by applying a constant load until steady strain rate behavior was observed for a considerable period of time (Table 2).

Furthermore, tension tests changing the crosshead speed (from 0.1 mm/min to 1 mm/min) every 2 or 3 mm were undertaken in order to calculate the strain rate equation coefficient. To compare the behavior of the raw state rolled material with that of the samples modified by heat treatments, tension test specimens were heat treated at 250°C for periods of 1 and 24 hours.

Microstructural characterization of the samples was undertaken using a metallographic cutter to observe longitudinal sections of the thin slab casting as well as the parallel direction of the thin strip. Traditional grinding and polishing techniques were used to obtain a mirror finish, as well as etching using Palmerton reactive. Metallographic characterization was performed through optical microscopy with Nikon Epiphot equipment and the quantitative analysis of the grain sizes of the α-Zn grains was undertaken using a Buehler Omnimet image analyzer connected to the optic microscope. This allowed for automatic grain identification and ASTM grain size measurement to be used, although in the case of the heat treated samples manual counting techniques were employed.

3. Results

The microstructure of the thin slab casting is shown in Fig. 3. Columnar grains of α-Zn (disperse constituent) are surrounded by an eutectic matrix (α+$Zn_{16}Ti$), just as is expected from the phase diagram (Fig. 4) [18]:

$$L\ (0.07\%) \rightarrow 74\%\ \alpha_{proeut} + 26\%\ (\alpha + \mathrm{Zn_{16}Ti}) \quad (1)$$

The phase diagram indicates a 74% presence of proeutectic phase and 26% of eutectic, which results in approximately 1.1% of $Zn_{16}Ti$. It is, therefore, an abundant dispersoid capable of grain refinement and a barrier to the restoration and recrystallization processes during hot deformation.

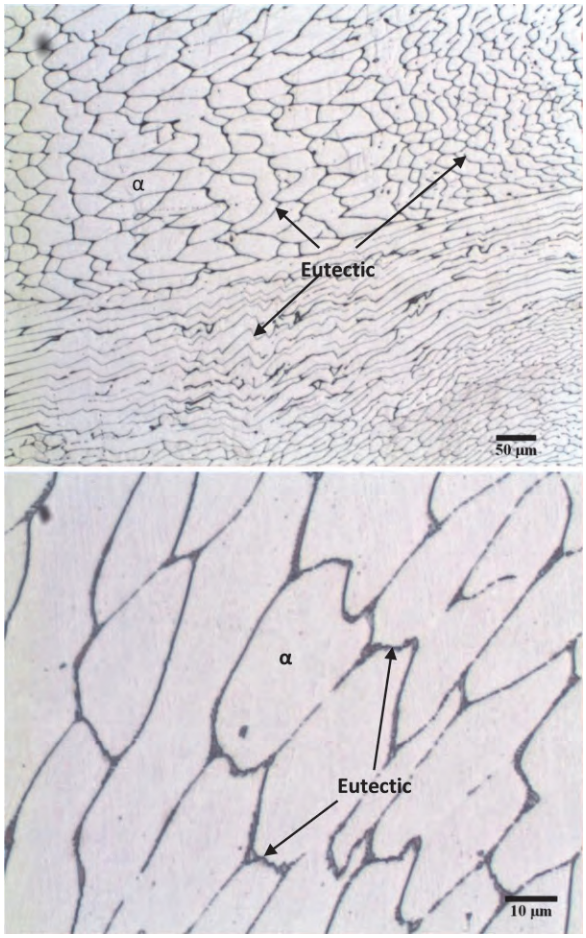


Figure 3. Solidification microstructure: α grains and eutectic phase (α + $Zn_{16}Ti$)
Source: The authors

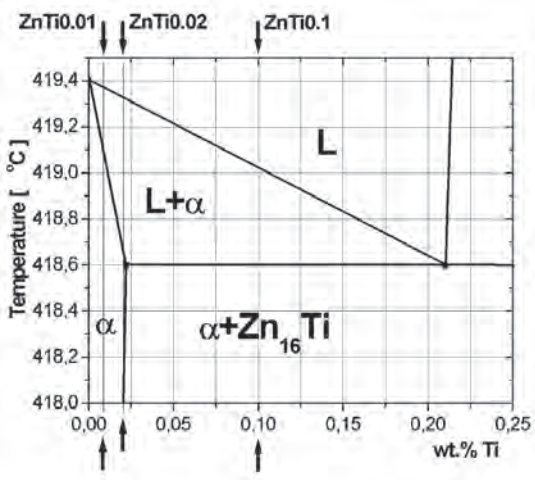


Figure 4. Zn-Ti phase diagram.
Source: [18]

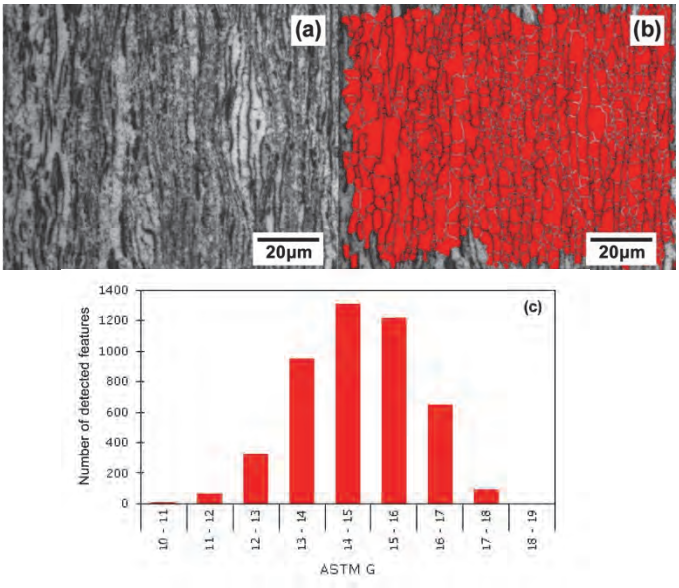


Figure 5. Raw state microstructure (a), detected grain pattern (b) and ASTM G grain size distribution histogram (c)
Source: The authors

Moreover, the microstructure of the raw hot rolled state strips (0.7 mm in thickness) observed in a section parallel to the rolling direction (Fig. 5a) is that of a restored material, partially recrystallized, and with bands in the rolling direction. Just as Kurz [19] observes, lighter zones are α-Zn grains and darker ones are α+$Zn_{16}Ti$ eutectics which were not separated nor spheroidized after the homogenization and hot deformation treatments. The optical microscopy does not show $Zn_{16}Ti$, neither at the boundaries nor at the interior of the α-Zn grains. Also, the size of the α-Zn grains where the eutectic is deformed is smaller than at the bands where the eutectic is not present. The quantitative measurements (Fig.s 5b-c) indicate a very small grain size with a mean length of α-Zn, close to 2 µm (14.5 ASTM).

In order to evaluate the response of this material to strain rate, tensile tests were made at different crosshead speeds in samples both parallel and perpendicular to the rolling direction, as indicated by Table 1. As expected, the higher the strain rate the higher maximum stress the material may withstand, as shown in the engineering stress-strain curves for perpendicular samples (Fig. 6). Though the sample tested at 10 mm/min may have fractured due to defects produced during machining, the rest of the samples have a tendency to deform by larger amounts (values as high as 37%) when low strain rates are used.

On the other hand, samples parallel to the rolling direction present lower maximum stresses (Table 1), although deformations between 50 and 55% are reached in all samples. This confirms the effect of the banded structure and crystallographic texture in the anisotropic mechanical behavior.

Table 1.
Tension stresses at various crosshead speeds for samples perpendicular and parallel to the rolling direction. RS – Raw State, 1h – 250°C + 1 hour heat treated, 24h – 250°C + 24 hours heat treated.

Crosshead speed		Tension stresses (MPa) 300 K			
		Perpendicular	Parallel		
mm/min	$\dot{\varepsilon} \times 10^{-5}$ (s^{-1})	RS	RS	1h	24h
0.1	1.67	143.3	97.2		
1	16.67	179.0	116.9		
2	33.33	188.5	119.0	129.3	123.3
5	83.33	201.8	133.2	135.5	132.0
10	166.67	207.4	140.3		
20	333.33	213.9			
50	833.33	223.0			

Source: The authors

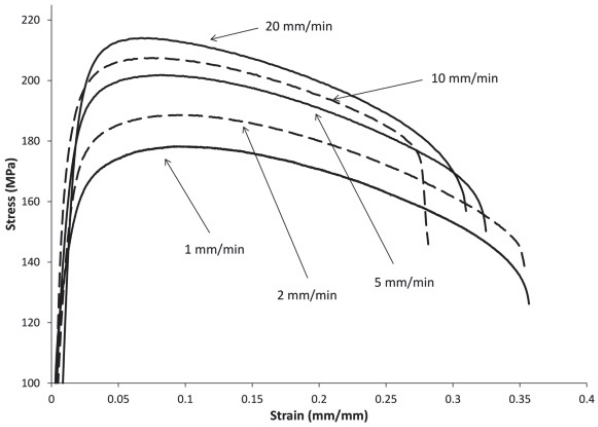


Figure 6. Effect of strain rate on engineering stress-strain curves for samples perpendicular to the rolling direction
Source: The authors

Table 2.
Steady state strain rates at constant stress values, for samples perpendicular and parallel to the rolling direction. RS – Raw State, 1h – 250°C + 1 hour heat treated, 24h – 250°C + 24 hours heat treated

Creep Stress (MPa)	$\dot{\varepsilon} \times 10^{-9}$ (s^{-1}) 300 K			
	Perpendicular	Parallel		
	RS	RS	1h	24h
70	3.2	210.0	30.0	10.0
80	9.8			
90	45.0	7100.0	250.0	1000.0
100	290.0			
110	1090.0			

Source: The authors

As the resistance of the material under constant load is very important in Zn alloys, creep tests were performed in order to determine the strain rate (steady state creep) [16,20,21] for the material (Table 2). Although samples with stresses above 90MPa were tested until they completely deformed and fractured, showing the typical creep curve [22] of strain vs. time (Fig. 7), lower stress levels showed very low strain values. These tests were only carried out until the steady state strain rate was evident.

If the values of stress (maximum stress for tensile tests and constant stress for creep) vs. strain rate are plotted for all the samples tested, it becomes evident that parallel and perpendicular samples have a similar profile: as strain rate is higher, stress reaches a plateau of approximately 190 MPa for the perpendicular samples and 125 MPa for the parallel ones (Fig. 8).

Furthermore, in order to analyze the possibility of improving creep behavior of parallel samples through the modification of microstructural features [23], heat treatments at 250°C were undertaken for periods of 1 and 24 hours. Fig. 9 shows evident microstructural changes due to the activation of diffusional processes [17]: the strip has recrystallized, grain size has grown (13 µm or 9 ASTM for 1 h and 15 µm or 8.5 ASTM for 24 h) and is more homogeneous, and the eutectic phase has fragmented which results in bands diminishing. Nevertheless, $Zn_{16}Ti$ are still not observed at either boundaries or at the interior of α-Zn grains.

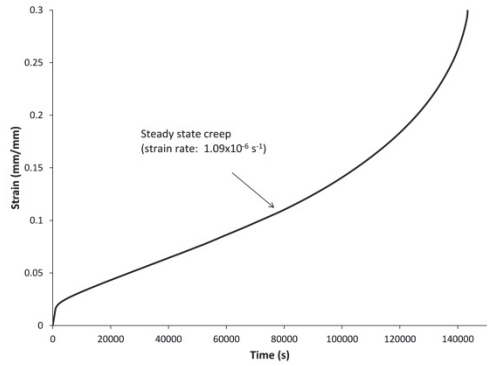


Figure 7. Creep curve for a sample perpendicular to the rolling direction with a constant stress of 110 MPa
Source: The authors

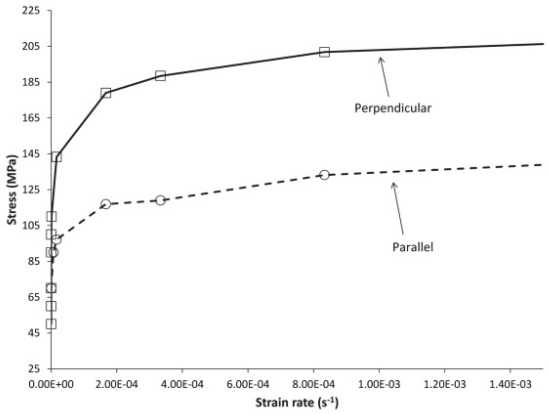


Figure 8. Maximum tension stress vs. strain rate for samples perpendicular and parallel to the rolling direction
Source: The authors

If tensile tests are compared in the case of samples parallel to the rolling direction (Fig. 10), the one hour treatment seems to increase mechanical properties (also indicated in Table 1) caused by the disappearance of the continuous eutectic bands. The 24 hour treatment can only slightly improve mechanical resistance (or even decrease it) depending on the strain rate (Fig. 9 and Table

1): grain growth and $Zn_{16}Ti$ coarsening during recrystallization softens the material [3] and the disappearance of continuous bands results in a continuous α-Zn deforming phase.

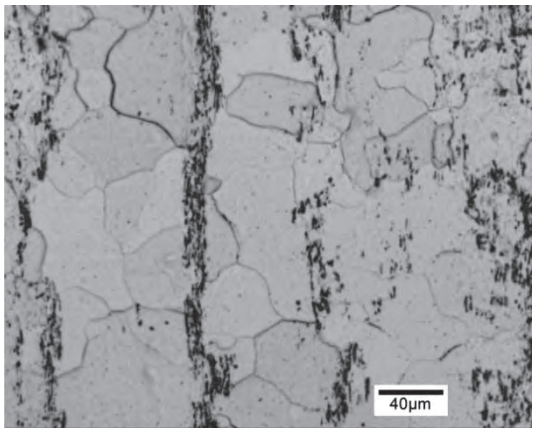


Figure 9. Microstructure of the material heat treated at 250°C
Source: The authors

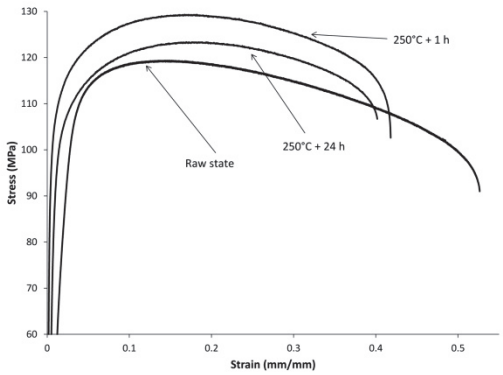


Figure 10. Effect of heat treatments on engineering stress-strain curves for samples parallel to the rolling direction that were tested with a constant crosshead speed of 2 mm/min
Source: The authors

4. Discussion

If the data for maximum stress (in the case of tensile tests) or constant stress (in the case of creep tests) is plotted vs. strain rate (strain rate of the tensile tests or steady state creep rate of the creep tests), using logarithm of both variables, a linear behavior is observed in the parallel and perpendicular directions (Fig. 11), especially above 90 MPa.

In order to confirm that the value of the m coefficient ($\sigma = K\dot{\varepsilon}^m$) is actually the one indicated by the data in Fig. 11 (slopes), a tensile test varying strain rate every 2 or 3 mm from 0.1 to 1 mm/min (Fig. 12) was undertaken. Consider the following equation:

$$m = \frac{\log(\sigma_2/\sigma_1)}{\log(\dot{\varepsilon}_2/\dot{\varepsilon}_1)} \quad (2)$$

where the $\dot{\varepsilon}_2/\dot{\varepsilon}_1$ ratio in this test is 10, σ_2 is the peak value of each step at the 1 mm/min curve sections and σ_1 is the last value of the 0.1 mm/min curve section before the crosshead speed changes. The m coefficient was calculated: for the samples in the parallel direction $m \approx 0.077$, while for the perpendicular direction $m \approx 0.080$. Both values are very similar to the ones observed in Fig. 11, and much lower (0.13~0.33) than the ones reported by Reed-Hill [24].

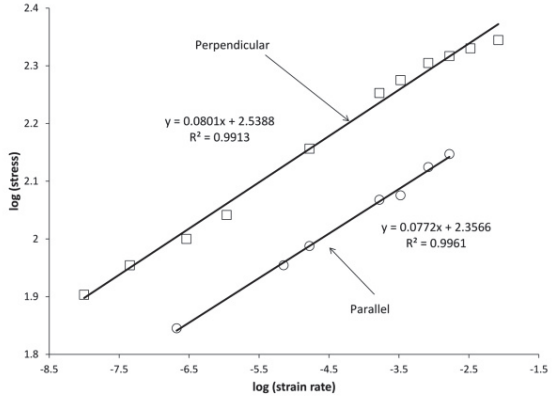


Figure 11. Tension stress vs. strain rate (log of both values) for samples perpendicular and parallel to the rolling direction
Source: The authors

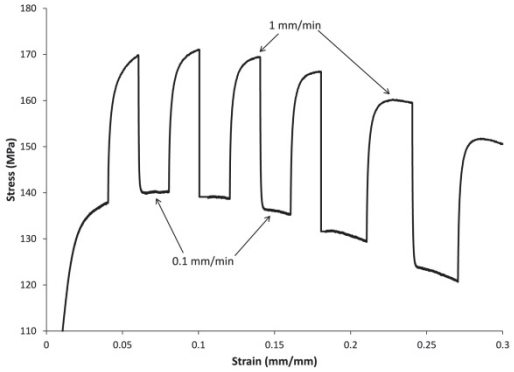


Figure 12. Stress-strain for a sample perpendicular to the rolling direction changing from 0.1 mm/min to 1 mm/min and vice versa every 2 mm (first 4 peaks) and then every 3 mm (last 2 peaks)
Source: The authors

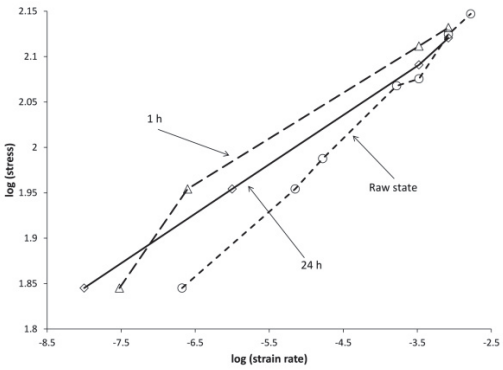


Figure 13. Tension stress vs. strain rate (log of both values) for samples parallel to the rolling direction in their rolled raw state and treated at 250°C for 1 and 24 hours
Source: The authors

In Fig. 13 Stress vs. Strain rate of samples heat treated for 1 and 24 hours at 250°C and the data of the material in its raw state is presented. Although at low strain rates the heat treated samples show higher strength and higher creep resistance than the raw state material, at high strain rates the values converge. This may be an indication that plastic mechanisms activated at low strain rates are different to the ones at high strain rates: at 300 K, creep in these type of alloys is characterized by intergranular sliding, thus the increase in grain size by the 250°C treatments will slow the effect of this mechanism [9].

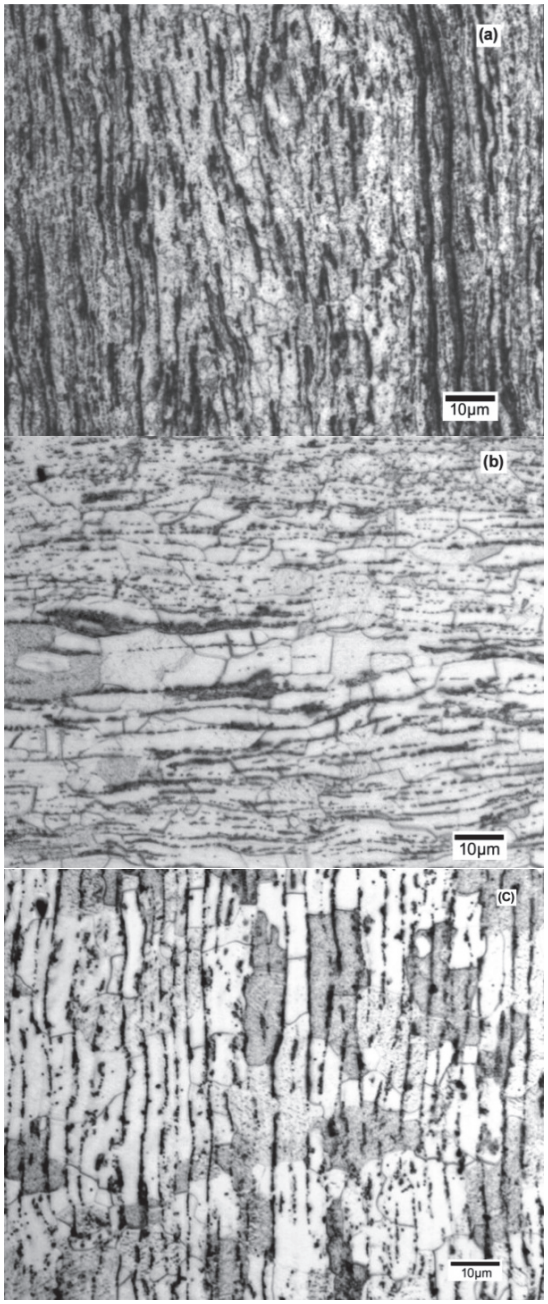


Figure 14. Microstructure of the material a) in its raw state and b) heat treated for 1hour at 250°C and c) heat treated for 24 hours at 250°C
Source: The authors

A comparison of microstructures in a raw state and after both heat treatments (250°C) for 1 and 24 hours is presented in Fig. 14. The increase in resistance for the one hour sample may be explained by the appearance of small, more homogeneously distributed dispersoids (precipitation hardening); however the 24 hour sample no longer presents the fine precipitates. Instead, there are coarser ones (overaging), which explains the decrease in mechanical properties. Furthermore, evidence of twinning is observed in the heat treated samples, which also explains the higher mechanical properties compared to the raw state ones [25-28].

5. Conclusions

Although the mechanical resistance of this alloy in the direction perpendicular to the rolling axis is higher than the parallel one, the strain rate coefficient m in both directions is very similar (0.07~0.08). Moreover, heat treatments at 250°C may increase mechanical properties (tensile and creep resistance), especially at low strain rates. Deformation mechanisms at high strain rates are very similar for raw state and heat treated samples.

The strengthening of heat treated samples deformed at low strain rates may be explained by the combination of the disappearance of continuous bands, twinning of the α-Zn grains and the appearance of homogeneously distributed precipitates.

Acknowledgments

The authors wish to thank the technical staff at the Asturiana de Laminados – ASLA (www.elzinc.es) for providing the material used for this research and detailed information of the rolling process for the Zn-Cu-Ti alloys.

References

[1] Barnhurst, R.J., Zinc and Zinc alloys, in Metals Handbook Vol. 2, 10th ed., USA, ASM, 1990. pp. 527-542.

[2] Jonas, J.J. and Mc-Queen, H.J., Recovery and recrystallization during high temperature deformation, in Mise en Forme des Métaux et Alliages, Paris, Editions de CNRS, 1976, pp. 99-143.

[3] Matsunaga, T., Kameyama, T., Takahashi, K. and Sato, E., Constitutive relation for ambient-temperature creep in hexagonal close-packed metals. Materials Transactions, 50(12), pp. 2858-2864, 2009. DOI: 10.2320/matertrans.M2009223

[4] Valadez, S., Zanatta, A., Robles, S. and Valdez, S., Microstructural analysis of Al-Mg-Si-Zn Alloy. DYNA, 77(163), pp. 9-12, 2010.

[5] Boczkal, G., Second phase morphology in the Zn-Ti0.1-Cu0.1 single crystals obtained at different growth rates. Archives of Metallurgy and Materials, 57(2), pp. 479-484, 2012.

[6] Faur M. and Cosmeleata G., Effects of hot and cold rolling on the microstructure of low alloy Zn-Cu and Zn-Cu-Ti zinc alloy with improved corrosion resistance. U.P.B Scientific Bulletin Series B, 68(1), pp. 67-74, 2006.

[7] Jaoul, B., Etude de la plasticité et application aux metaux. Paris: Dunod, 1965, pp. 362-400, 481-511.

[8] Das, S.K., Kang, D.H. and Jung, I.H., Experimental and diffusion simulation for the homogenization of As-cast Mg-Al, Mg-Zn and Mg-Al-Zn alloys. Metallurgical and Materials Transactions A, 45A(11), pp. 5212-5225, 2014. DOI: 10.1007/s11661-014-2443-6

[9] Xiao, L.R., Zhang, X.M., Wang, Y., Li, W., Sun, Q.S. and Geng, Z.J., High temperature creep behavior of Zn-1.0Cu-0.2Ti Alloy. Advanced

Materials Research, 287-290(7), pp. 769-776, 2011. DOI: 10.4028/www.scientific.net/AMR.287-290.769
[10] Hirth, J.P. and Lothe, J., Theory of dislocations. New York: Wiley, 1968. pp. 637-753.
[11] Dieter, G.E., Mechanical metallurgy. UK: McGraw-Hill, 1988. pp. 185-240.
[12] Pero-Sanz, J.A., Ciencia e Ingeniería de Materiales. Madrid: Cie Dossat, 5th ed., 2006. pp. 393-446.
[13] Cosse P., D´Heeyer, R., Coutsouradis, D. et Habrajen, L., Le comportement à l´emboutissage du Zinc et d´alliages de Zinc. Mémoires Scientifiques Revue de Métallurgie, pp. 517-528, 1978.
[14] Penelle, R., Influence de l´anisotropie sur la limite d´élasticité et la loi de comportement, in Mise en Forme des Métaux et Alliages, Paris, Editions de CNRS, 1976, pp. 75-85.
[15] Barrett, C.S., Structure des Métaux. Paris: Dunod, 1957, pp. 417-478.
[16] Deka, D., Joseph, D., Ghosh, S. and Mills, M., Crystal plasticity modeling of deformation and creep in polycrystalline Ti-6242. Metallurgical and Materials Transactions A, 37A(5), pp. 1371-1388, 2006. DOI: 10.1007/s11661-006-0082-2
[17] Doiphode, R.L., Narayana Murty, S.V.S., Prabhu, N. and Kashyap, B.P., High-temperature tensile flow behavior of caliber-rolled Mg-3Al-1Zn Alloy, Metallurgical and Materials Transactions A, 46A(7), pp. 3028-3042, 2015. DOI: 10.1007/s11661-015-2883-7
[18] Mikulowski, B. and Boczkal, G., Zn-Ti single crystals deformed along the basal slip system. Archives of Metallurgy and Materials, 54(1), pp. 197-203, 2009.
[19] Kurz, W. and Trivedi, R., Banded solidification microstructures. Metallurgical and Materials Transactions A, 27A(3), pp. 625-634, 1996. DOI: 10.1007/BF02648951
[20] Blum, W., Eisenlohr, P. and Breutinger, F., Understanding Creep – A review. Metallurgical and Materials Transactions A, 33A(2), pp. 291-303, 2002. DOI: 10.1007/s11661-002-0090-9
[21] Peschanskaya, N.N., Shpeizman, V.V., Sinani, A.B. and Smirnov, B.I., Micrometer-Scale deformation jumps at different stages of creep in solids. Physics of the Solid State, 45(11), pp. 2058-2062, 2004. DOI: 10.1134/1.1825549
[22] Winter, R., EZAC – High strength, creep resistant, Zinc die casting alloy. Die Casting Engineer, (3), pp. 32-35, 2011.
[23] Mahmoud, S.A., Semary, M.A., Farid, Z.M. and El-Naquib, N., On two types of creep behaviour of Cu-Zn solid solution. Materials Science and Engineering A, 118(10), pp. 107-111, 1989. DOI: 10.1016/0921-5093(89)90062-2
[24] Reed-Hill, R.E., Physical metallurgy principles. Second Edition, University Series in Basic Engineering, 1994, pp. 827-887.
[25] Zhu, Y.T., Liao, X.Z., Wu, X.L., Narayan, J., Grain size effect on deformation twinning and detwinning. Jounal of Materials Science, 48, pp. 4467-4475, 2013. DOI: 10.1007/s10853-013-7140-0
[26] Jenei, P., Gubicza, J., Dirras, G., Lábár, J.L. and Tingaud, D., Indentation creep study on ultrafine-grained Zn processed by powder metallurgy. Materials Science & Engineering A. 596, pp. 170-175, 2014. DOI:10.1016/j.msea.2013.12.050
[27] Boczkal, G., Mikulowski, B., Oertel, C.G. and Skrotzki, W., Work-hardening characteristics of Zn-Ti alloy single crystals. Crystal Research and Technology, 45(2), pp. 111-114, 2010. DOI: 10.1002/crat.200900537
[28] García-Bernal, M.A., Hernandez-Silva, D. and Sauce-Rangel, V., Superplastic behavior of coarse-grained Al-Mg-Zn alloys. Journal of Material Science, 42, pp. 3958-3963, 2007. DOI: 10.1007/s10853-006-0368-1

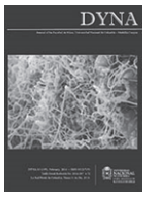
M.J. Quintana, received her BSc. Eng in Mechatronics Engineering in 2006, her Sp. degree in Advanced Manufacturing Systems in 2007, her MSc. degree in Engineering in 2010, all of them from the Universidad Panamericana, Mexico City, Mexico. She was awarded her PhD in Science and Technology of Materials in 2013 from the Universidad de Oviedo. Oviedo, Spain. She has worked since 2007 in the Faculty of Engineering at the Universidad Panamericana, first as an invited professor, and since 2011 as an associate professor and researcher. Her research interests include: finite element simulation, material testing and characterization, selection of materials for engineering applications.
ORCID: 0000-0002-1926-9910

J.O. García, received his Adv. Technician degree in 1975, from the Polytechnic School of Oviedo, Spain. Has worked in the Universidad de Oviedo since 1982 as a lab teacher in the Materials Department, and now also works as an associate professor on the Science and Engineering course. His research interests include: optical and electronical microscopy, mechanical testing, thermal treatments and alumino-thermal welding.
ORCID: 0000-0002-9156-7349

R. González, received his BSc. Eng in Mechanical and Electrical Engineering in 1991 from the Universidad Panamericana, Mexico City, Mexico and his PhD in Mechanical Properties and Structure of Materials in 1996 from CEIT - Universidad de Navarra, San Sebastián, Spain. Since 1996 he has worked in the Faculty of Engineering at the Universidad Panamericana as associate professor and researcher and is currently Dean of the Mechatronics and Mechanic Engineering programs. His research interests include: finite element simulation, material testing and characterization, tribology, design theory.
ORCID: 0000-0003-2319-1631

J.I. Verdeja, received his Mining Eng. degree from the Adv. Tech. School of Mining Eng. in Madrid in 1968, his Nuclear Eng. degree from the Nuclear Energy Board in 1969, his PhD in Mining from the Adv. Tech. School of Mining Eng. of Oviedo in 1974, and a PhD form the University of Paris in 1978. He began to work in the Adv. Tech. School of Mining Eng. of Oviedo in 1974 as an invited and associate professor and then as an associate chairman in 1983. His research interests include: steels, castings and metallic alloys in general.
ORCID: 0000-0002-4446-6457

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS



DYNA
http://dyna.medellin.unal.edu.co/

Non-destructive *in situ* analysis of garnet by combining scanning electron microscopy and X-ray diffraction techniques

John Deiver Bonilla-Jaimes [a], Jose Antonio Henao-Martínez [a], Carolina Mendoza-Luna [b], Oscar Mauricio Castellanos-Alarcón [c] & Carlos Alberto Ríos-Reyes [d]

[a] *Escuela de Química, Universidad Industrial de Santander, Bucaramanga, Colombia, johnrayosx@gmail.com*
[a] *Escuela de Química, Universidad Industrial de Santander, Bucaramanga, Colombia, jahenao@uis.edu.co*
[b] *Escuela de Física, Universidad Industrial de Santander, Bucaramanga, Colombia, caroluna8007@gmail.com*
[c] *Programa de Geología, Universidad de Pamplona, Pamplona, Colombia, oscarmca@yahoo.es*
[d] *Escuela de Geología, Universidad Industrial de Santander, Bucaramanga, Colombia, carios@uis.edu.co*

Received: October 19th, 2014.Received in revised form: August 11th, 2015. Accepted: January 21th, 2016.

Abstract
By using the X-ray powder diffraction (XRPD) micro X-rays diffraction (μXRD) and scanning electron microscopy, the structural characterization of minerals is far more reliable and accurate. The identification and elemental and compositional quantification of minerals by these non-destructive techniques improve the quality of the results and allow a full analysis of the material. The data obtained by these techniques revealed the presence of garnet-type spessartine, in addition to trace elements and compounds that form the overall material. The structural refinement of spessartine was performed using the Rietveld method from data obtained by conventional diffraction and by using the MDI RIQAS analysis software. With the data acquired by μXRD using an area detector, a shorter exposure time (compared to that required by the 0L and 1D detectors) was achieved, and there was no need for particle size reduction of the mineral. It was also possible to identify the spessartine and other compounds in smaller concentrations (in situ measurements). By combining scanning electron microscopy and X-ray diffraction techniques, both worked from a characterization point of view. The examination by micro X-ray diffraction did not require physical separation of the sample. Using this information and the above advanced analytical techniques, the identification of garnet can be undertaken much more reliably.

Keywords: scanning electron microscopy; micro X-ray diffraction; analytical techniques; mineral; garnet.

Análisis no destructivos *in situ* de granate, combinando las técnicas de microscopía electrónica de barrido y difracción de rayos-X

Resumen
Mediante el uso de la difracción de rayos-X de polvo (DRXP), microdifracción de rayos-X (μDXR) y microscopía electrónica de barrido, la caracterización estructural de minerales resulta ser mucho más fiable y precisa. La identificación y cuantificación elemental y composicional de los minerales mediante estas técnicas no destructivas, mejoran la calidad de los resultados y permiten realizar un análisis completo del material. Los datos obtenidos mediante estas técnicas revelaron la presencia de granate tipo espesartina, además de los elementos y compuestos trazas que conforman el material en general. El refinamiento estructural de la espesartina fue realizado mediante el método Rietveld a partir de los datos obtenidos por difracción convencional y con ayuda del software de análisis MDI RIQAS. Con los datos adquiridos por μDXR usando un detector de área, un menor tiempo de exposición (comparado con el requerido en detectores 0D y 1D) y sin la necesidad de la disminución del tamaño de partícula del mineral, fue posible la identificación de la espesartina y otros compuestos en menor concentración (mediciones "in situ"). Mediante la combinación de las técnicas de microscopía electrónica de barrido y microdifracción de rayos X, tanto de trabajo desde un punto de vista de la caracterización. El examen por difracción de micro-rayos X no requiere la separación física de la muestra. Usando esta información y las técnicas analíticas avanzadas anteriores, la identificación de granate puede ser mucho más fiable.

Palabras claves: microscopía electrónica de barrido; micro difracción de rayos-X; técnicas analíticas; mineral; granate.

1. Introduction

The destructive character of many analytical techniques has restricted studies in mineral characterization, since the conservation of mineral properties has been prioritized above any information that may be obtained. However, the continuous improvements in scientific instruments has made it possible to study minerals in a non-destructive way, without any degradation or alteration in their appearance. The scanning electron microscopy (SEM) and X-ray diffraction (XRD and μXRD) techniques enable the mineral characterization in a non-destructive way. The μXRD is a structural analysis technique that allows very small sample areas to be examined to obtain information about the structure of crystalline materials. Recent developments in specialized laboratories have extended the application of μXRD to the examination of geomaterials, which have a broad spectrum of applications [1] in the fields of mineralogy [2], petrology [3], material sciences, environmental sciences [4], and Earth and planetary sciences [5-8]. The SEM is a powerful tool for the microstructural analysis of geomaterials. The μXRD is a versatile technique that uses a two-dimensional (2D) general area detector diffraction system (GADDS), which acquires textural and crystallinity information and easily discriminates between multiple phases, providing mineral identification using crystal structural parameters [5]. Moreover, in situ μXRD of minerals preserves spatial relationships, enabling the study of orientational phenomena, such as strain-related mosaicity (giving "streaked" diffraction lines). The objective of this study is to carry out the analysis of garnet by combining SEM and μXRD techniques.

2. Analytical techniques

2.1. Materials

The sample that was analyzed is from a garnet amphibolite from the Arquía Complex, cropping out on the southern Pijao (Quindío). Several studies [9-12] report the occurrence of this type of rock in the Arquía Complex, which represents a tectonic mixture of blocks of different origins (e.g., oceanic crust, subduction zone, marine sedimentary rocks, etc.) that were created by the shearing zone between the Caribbean-Colombian Cretaceous Igneous Province and the Colombian western margin during the Cretaceous period [13]. According to Kerr et al. [14], these processes result in the formation of ophiolitic complexes, metamorphic belts and the amalgamation and strong interaction of several tectonic terranes.

2.2. Analysis

2.2.1. Petrographic analysis

The garnet amphibolite samples were first analyzed by transmitted light microscopy, by using a Nikon trinocular microscope (model Labophot2-POL) in order to capture photomicrographs with the 5x objective of the optical properties of garnet, as well as its textural relationships with other mineral phases. The Kretz [14] system of mineral abbreviations is used.

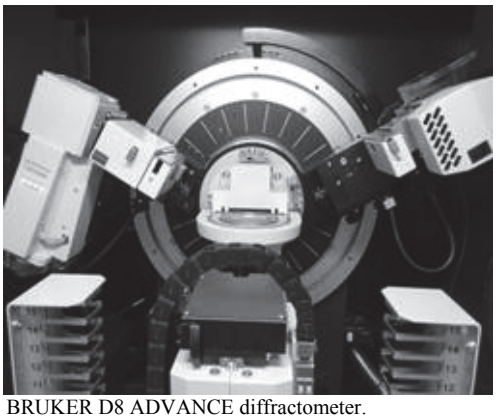


Figure 1. BRUKER D8 ADVANCE diffractometer.
Source: The authors.

2.2.2. Scanning electron microscopy (SEM)

Backscatter electron (BSE) imaging and EDS analysis of garnet were carried out using a FEI QUANTA FEG 650 environmental scanning electron microscope (ESEM), under the following analytical conditions: magnification = 100-20000x, WD = 9.0-11.0 mm, HV = 20 kV, signal = BSE in Z CONT mode, detector = BSED, EDS Detector EDAX APOLO X with resolution of 126.1 eV (in. Mn Kα).

2.2.3. X-ray powder diffraction (XRPD)

The sample was milled in an agate mortar and then mounted on a sample holder of polymethylmethacrylate (PMMA) using the filling front technique . The XRPD pattern of garnet was recorded by X-ray diffraction using a BRUKER D8 ADVANCE diffractometer (Fig. 1) operating in Da Vinci geometry and equipped with an X-ray tube (Cu-Kα1 radiation: λ = 1.5406 Å, 40 kV and 30 mA), a 1-dimensional LynxEye detector (with aperture angle of 2.93º), a divergent slit of 0.6 mm, two soller axials (primary and secondary) of 2.5º and a nickel filter. Data collection was carried out in the 2θ range of 12-80°, with a step size of 0.01526° (2θ) and counting time of 1 s/step. Phase identification was performed using the crystallographic database Powder Diffraction File (PDF-2) from the International Centre for Diffraction Data (ICDD) and the Crystallographica Search-Match program. The unit-cell constants, atomic positions, factors of peak broadening and phase concentrations were refined and calculated by using the MDI RIQAS program based on Rietveld method.

2.2.4. Micro X-Ray Diffraction (μXRD)

For a powder sample measurement, an agate mortar was first milled and then mounted on a polymethylmethacrylate (PMMA)sample holder using the filling front technique. Finally, the measurement was adjusted in the Eulerian cradle.

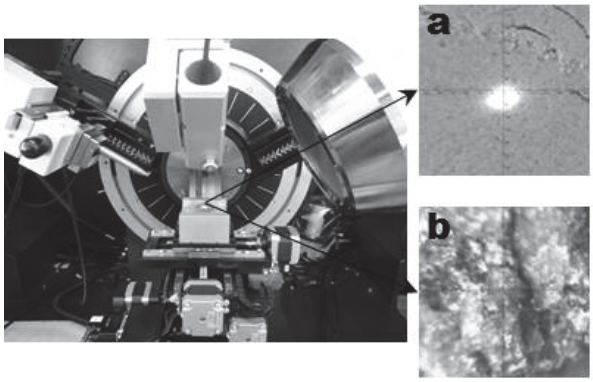


Figure 2. BRUKER D8 DISCOVER XRD^2 diffractometer. (a) μXRD on the milled simple. (b) μXRD on the unmilled sample.
Source: The authors.

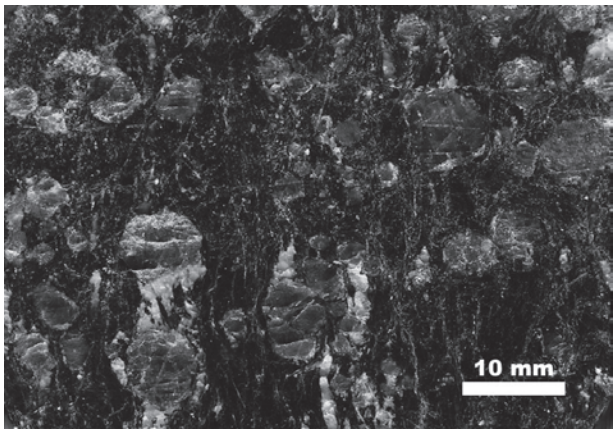


Figure 3. Photograph of the garnet amphibolite's hand specimen.
Source: The authors.

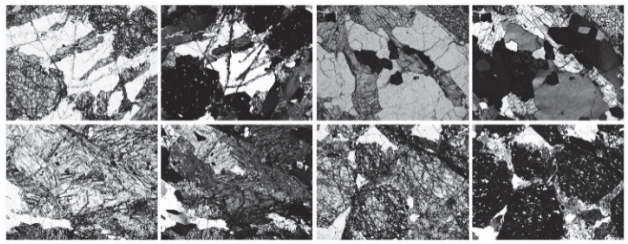


Figure 4. Photomicrographs of the garnet amphibolite observed under a transmitted light microscope. Grt, garnet; Hbl, hornblende; Qtz, quartz; Ilm, ilmenite; Cpy, chalcopyrite; Zrn, zircon; Chl, chlorite.
Source: The authors.

To measure an unmilled sample, the XYZ coordinate system of the quarter-circle Eulerian cradle was adjusted directly. μXRD data on powder and unmilled samples were collected with the BRUKER D8 DISCOVER diffractometer, equipped with the 2D GADDS (Fig. 2), operated in Da Vinci geometry with CuKα1 radiation (λ = 1.5406 Å) at 40 kV and 30 mA; equipped with a Optics Göebel Mirror, a circular divergent slit of 1 mm, a collimator of 1 mm, and a 2-dimensional detector VÅNTEC-500 (with an angular aperture of 42º in the 2θ range and a work distance of 150 mm). Data collection was carried out in the 2θ range of 25-45° in two intervals, with an increase per interval of 20° (2θ), a step scan, and a pixel size of 2048 x 2048. The two dimensional (2D) image GADDS were obtained in the spherical coordinate system of the Debye-Scherrer rings by using the DIFFRACT PILOT program. The diffraction patterns were obtained from the integration of several sections of the Debye-Scherrer rings by using a 0.01º (2θ) step. Phase identification was performed using the crystallographic database Powder Diffraction File (PDF-2) from the International Centre for Diffraction Data (ICDD) and the Crystallographica Search-Match software.

3. Results

3.1. Petrographic analysis

Fig. 3 shows a photograph of the garnet amphibolite of interest in the present study's hand specimen, with the garnet being typically dark reddish brown in color with a vitreous luster and xenoblastic character. It develops large porphyroblasts in a dark green nematoblastic matrix and is mainly composed of hornblende.

Fig. 4 illustrates the corresponding photomicrographs with large xenoblastic garnet porphyroblasts in a matrix mainly composed of hornblende.

3.2. Scanning electron microscopy (SEM)

The backscatter electron (BSE) images Figs. 5-6 show the textural relationships observed between spessartine-type garnet and associated mineral phases with semiquantitative energy dispersive spectrum (EDS) analysis at different points. This allowed particular elements and their relative proportions in the mineral phases that constitute the garnet-amphibolite sample of interest in this study to be identified. EDS analysis reveals that the main mineral phases in the analyzed rocks correspond to spessartine-type garnet and hornblende. The main accessory mineral phases are quartz, zoisite (epidote-group mineral), rutile, ilmenite, pyrite, chalcopyrite and zircon. Chlorite is recognized as a replacement product of hornblende.

Fig. 5a illustrates examples of the matrix mineral phases, with hornblende (1) as the main associated mineral phase, which is commonly replaced by chlorite (2). Rutile (3) developing reaction rims of titanite (4) occurs as the main accessory mineral phase. Very fine-grained zircon (5) crystals are scarcely distributed in the study samples. EDS analysis reveals the following mass ratios: Si:Al:O:Fe:Ca:Mg:Na:K:Ti for hornblende (23.64:6.98:26.91:11.44:6.90:6.73:1.59:0.29:0.29), Si:Al:O:Fe:Mgfor chlorite (14.11:10.91:31.58:19.18:9.91), Ti:O for rutile (64.40:26.49), Si:Al:O:Ca:Ti:Fe for titanite (15.91:0.51:22.39:22.10:28.68:0.33), and Si:O:Zr for zircon (15.62:18.38:45.25). Fig. 5b displays the occurrence of numerous randomly distributed mineral inclusions, such as quartz (6), zoisite (7), hornblende (8), chalcopyrite (9) and ilmenite (10), in spessartine-type garnet. EDS analysis reveals the following mass ratios: Si:O for quartz (47.34:35.62), Si:Al:O:Ca:Fe for zoisite (19.86:15.3726.96:17.11:6.38), Si:Al:O:Mg:Fe for hornblende (13.36:11.11:33.09:11.14:16.00), Fe:Cu:S for chalcopyrite (24.91:28.72:27.65), and Ti:Fe:O for ilmenite (30.93:37.65:20.75).

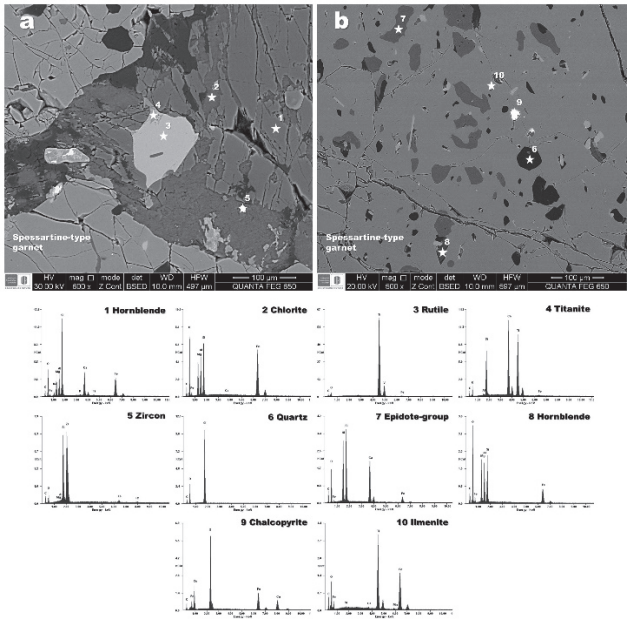


Figure 5. SEM photomicrographs and EDS spectra at the marked stars on the images of the mineral phases associated with spessartine-type garnet. The appearance of C element was attributed to the carbon coating on the sample before SEM analysis.
Source: The authors.

Figs. 6a and 6b illustrate the occurrence of pyrite (1) in the matrix, which contains chalcopyrite (2) inclusions and is partially replaced by leucoxene (3). Figs. 6c and 6d illustrate the occurrence of ilmenite (4) in the matrix, which contains zoisite (5) inclusions and is partially replaced by rutile (6). EDS analysis reveals the following mass ratios: Fe:S for pyrite (38.38:41.25), Fe:Cu:S for chalcopyrite (24.91:28.72:27.65), Fe:O for leucoxene (66.22:19.87),

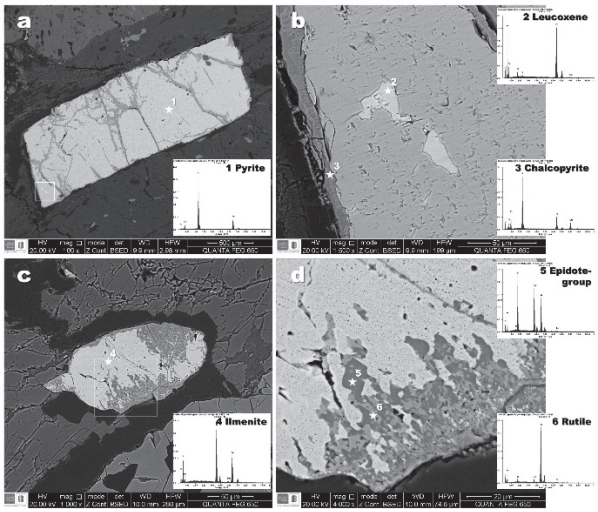


Figure 6. SEM photomicrographs and EDS spectra at the marked stars on the image of the Fe-Ti oxide mineral phases associated with spessartine-type garnet. The appearance of C element attributed to the carbon coating on the sample before SEM analysis.
Source: The authors.

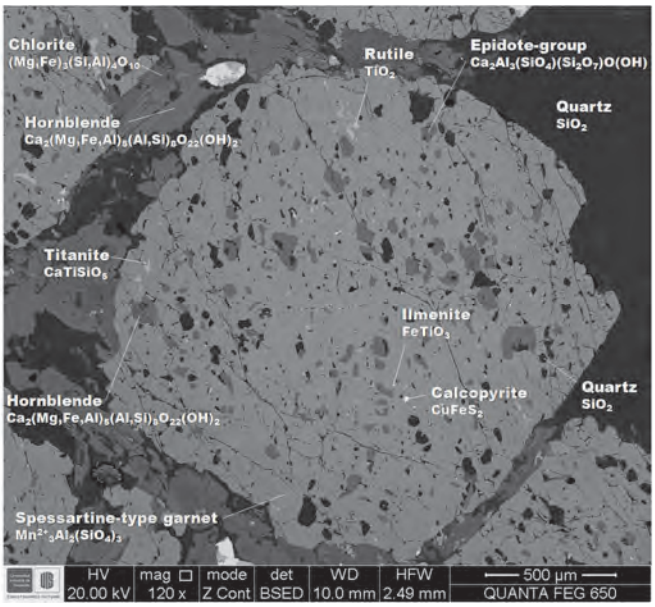


Figure 7. BSE image of spessartine-type garnet. Variation in intensity of grey scale color shows a rough variation in mineral chemistry.
Source: The authors.

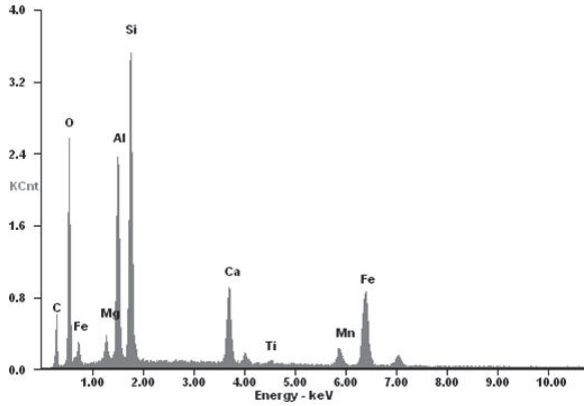


Figure 8. EDS spectrum of spessartine-type garnet. The appearance of C element is attributed to the carbon coating on the sample before SEM analysis.
Source: The authors.

Ti:Fe:O for ilmenite (30.93:37.65:20.75),Si:Al:O:Ca:Ti:Fe for zoisite (16.92:0.44:23.90:21.36:25.47:0.93), and Ti:O for rutile (64.40:26.49).

The backscatter electron (BSE) image Fig. 7 shows the textural relationships observed between spessartine-type garnet and associated mineral phases. Note the high-density Fe-Ti oxide (on the upper and lower parts), the randomly oriented inclusion-rich spessartine-type garnet and associated mineral phases, and the reaction textures from the BSE image contrast.

EDS analysis reveals that the mass ratios (Si:Al:O:Fe:Ca:Mn:Mg) of the spessartine-type garnet are 18.96:11.63:25.49:18.50:6.32:3.67:1.56. The EDS spectrum of spessartine-type garnet (Fig. 8) reveals that it mainly consists of Si, Al, O, Fe and Ca, with a minor amount of Mn and Mg.

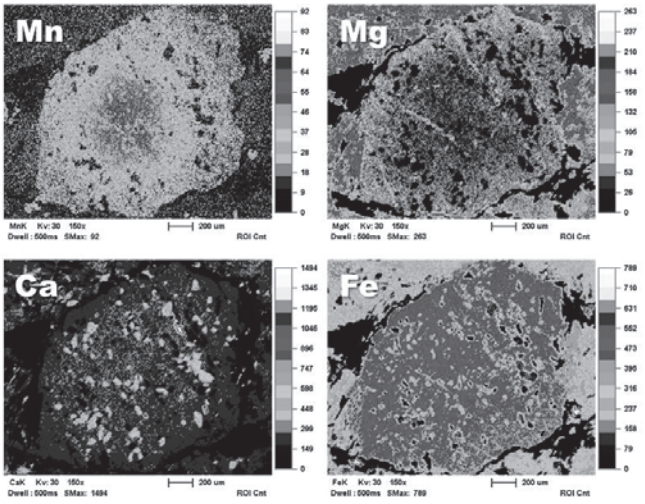


Figure 9. (a) Mn, (b) Mg, (c) Ca and (d) Fe compositional maps of spessartine-type garnet and associated mineral phases. Light colors show areas of high concentration while dark colors represent areas of low concentration (black is very low concentration).
Source: The authors.

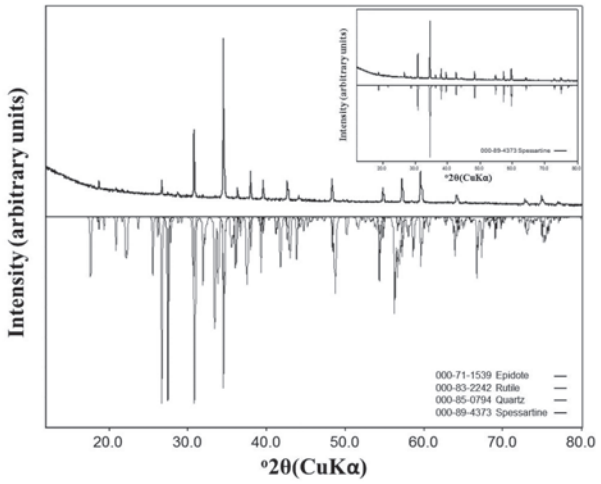


Figure 10. XRPD pattern of spessartine-type garnet. In the lower part shows the diffraction patterns of the mineral phases identified in the crystallographic database Powder Diffraction File (PDF-2) from the International Centre for Diffraction Data (ICDD).
Source: The authors.

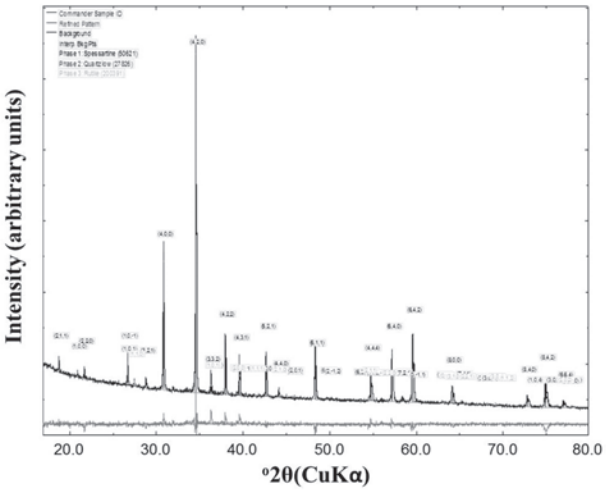


Figure 11. Observed (cross-hatches) and calculated (continuous lines) profiles and the corresponding difference diagram of spessartine-type garnet (λ = 1.5406 Å).
Source: The authors.

X-rays generated by scanning the electron beam across the sample can be used to produce EDS mapping, which provides an image of a meaningful picture of the elemental distribution of a mineral phase, in addition to the BSE. In Fig. 9, the different phases shown on the BSE image (Fig. 9a) can be identified by elemental mapping (Figs. 9b-9d); however, this will only give a qualitative image of the distribution of elements. The elemental mapping shows an oscillatory pattern with respect to Mn (Fig. 9b, X_{sps}), which has a decrease of X_{sps}. However, this cannot be attributed to the abundant epidote-group inclusions, because there is no relationship between zoning and the distribution pattern of inclusions. No chemical zoning can be identified from the core to the rim with regards to Mg (Fig. 9c, X_{grs}) and Ca (Fig. 9d, X_{grs}).

Elemental maps provide valuable information even without μXRD. It is very important to highlight that a high Mn content is mirrored by a low Mg content. Note the high content of Ca, representing zoisite (a Ca-rich mineral phase) inclusions in the spessartine-type garnet. Fe content is very uniform in spessartine-type garnet; however, the low Fe content of the mineral inclusions should be noted.

3.3. X-ray powder diffraction (XRPD)

Fig. 10 shows the XRPD pattern of the spessartine-type garnet, in which the high intensity reflections correspond to spessartine-type garnet (PDF No. 000-89-4373), and the low intensity reflections reveal the presence of quartz (PDF No. 000-89-0794), rutile (PDF No. 000-83-2242), and epidote (PDF No. 000-71-1539).

The Rietveld refinement of the XRPD pattern allows the percentage of the rock-forming minerals to be determined. Fig. 11 illustrates the observed, calculated and difference profiles for the XRPD garnet data after refinement. The crystalline phase of interest in this study corresponds to a spessartine-type garnet structure, although a few extra lines reveal the occurrence of other mineral phases that are described above. The difference between the observed and calculated profiles shows a good fit for the refined parameters (cell-unit constants, atomic positions, factors of peak broadening and phase concentrations).

The Rietveld refinement details (including the parameters of profile broadening U, V, W, m), obtained for the spessartine-type garnet, are presented in Table 1. The refinement of the XRPD data reveals a small preferential orientation in the reflection (400) for the spessartine-type garnet framework, which crystallized in the cubic space

Table 1.
The Rietveld refinement details obtained for the spessartine-type garnet.

Molecular formula	$(Mn_{2.21}Fe_{0.79})Al_2(SiO_4)_3$
Molecular weight (g/mol)	495.03
a = b = c (Å)	11.6086
α = β = γ (º)	90
V (Å^3)	1564.37
Z	8
Space group	Ia$\bar{3}$d
ρ (g/cm^3)	4.208
U	0.146049
V	-0.08092
W	0.016211
m	1.35655
R_{WP} (%)	19.47
R (%)	12.02
R_{exp} (%)	3.4
χ^2	3.2

Source: The authors.

Table 2.
Atomic positions obtained for the spessartine-type garnet.

Atom	Ion	X	Y	Z
Mn1	Mn^{+2}	0	1/4	1/8
Fe1	Fe^{+2}	0	1/4	1/8
Al1	Al^{+3}	0	0	0
Si1	Si^{+4}	0	1/4	3/8
O1	O^{-2}	0.038071	0.043327	0.657748

Source: The authors.

group Ia$\bar{3}$d (nº 230), and has unit cell parameters a = b = c = 11.6086 Å; α = β = γ = 90º, V = 1564.37 Å^3, Z = 8 and ρ =4.208 g/cm^3. ICSD = 50621 was the number used to refine the spessartine-type garnet. These results fully agree with those obtained by Sawada [15].

The concentrations of the refined phases and their corresponding (standard deviation) obtained were: spessartine-type garnet $(Mn_{2.21}Fe_{0.79})Al_2(SiO_4)_3$ = 94.0% (0.3), quartz (SiO_2) = 4.9% (0.1), rutile (TiO_2) = 1.1% (0.1). The concentration and corresponding standard deviation of the epidote cannot be quantified due to its low concentration.

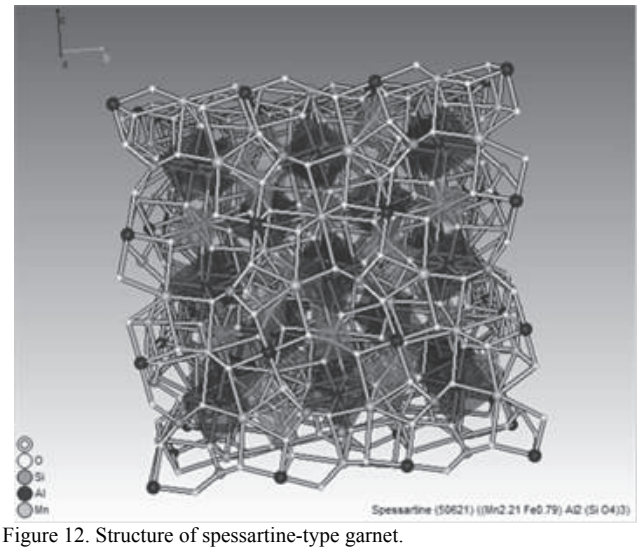


Figure 12. Structure of spessartine-type garnet.
Source: The authors.

The final agreement values for the refinement are R_{wp} = 19.47%, R = 12.02%, R_{exp} = 3.40%, and χ^2 = 3.2.

In the refinement model used in this study, only the atomic coordinates for oxygen were refined with a standard deviation in the three crystallographic axes that was lower than 10% with respect to the values reported in the database. The other elements show fixed positions and, therefore, they were not refined. The atomic coordinates for the spessartine-type garnet are depicted in Table 2.

Fig.12 shows the spessartine-type garnet structure obtained with the software MDI RIQAS5.

3.4. Micro X-Ray Diffraction (μXRD)

The two dimensional image (2θ,γ) from the GADDS detector shows how the spessartine-type garnet appears as homogeneous Debye rings (Fig. 13) of constant intensity, displaying 21 diffraction cones of 21 different planes in its crystalline structure. The diffraction spots and the Debye rings match this structure. The continuous and irregular rings reveal that the sample is polycrystalline, however, its particle size is very coarse.

The crystalline accessory minerals appear as a lower pixel diffraction cone of the quartz crystalline plane (011), and small crystalline spots correspond to rutile and epidote, which are shown in the 2D GADDS image of Fig. 14. The internal rings are characterized by the presence of rutile and epidote. This provides useful information on the scarce number of crystallites in these phases and their low concentration. Microcrystalline quartz appears to produce very thin lines, mostly appearing as a series of tiny connected dots, with some larger point reflections.

On each one of the Debye rings, the structure planes diffracted by the spessartine-type garnet are indicated and the reference area and arrow represent the place where the 1D integration diffraction pattern (Intensity vs. 2θ) was obtained (Fig. 15). The identification of the spessartine-type garnet (PDF No. 000-89-4373) was performed by comparing the observed diffraction pattern with the PDF-2 data base patterns.

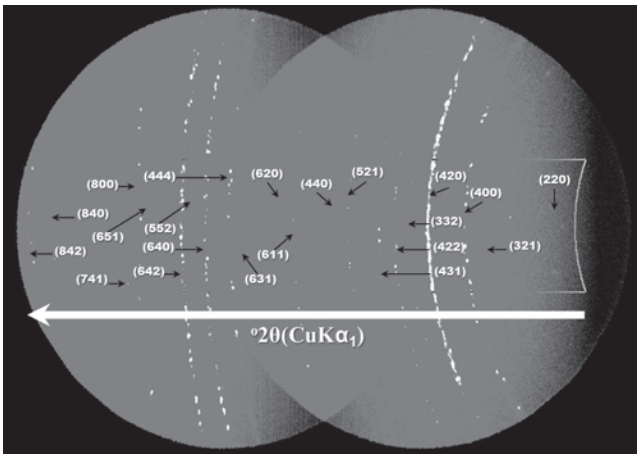


Figure 13. 2D GADDS image of the spessartine-type garnet in the milled sample. The reference area and arrow represent the integration of the 2D GADDS image.
Source: The authors.

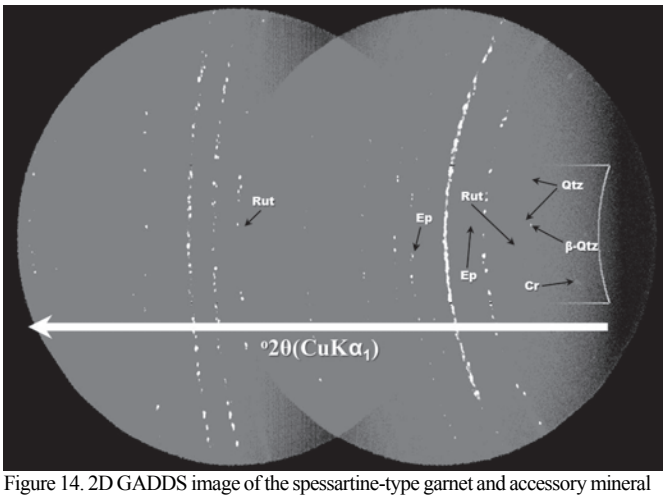


Figure 14. 2D GADDS image of the spessartine-type garnet and accessory mineral phases in the milled sample. The reference area and arrow represent the integration of the 2D GADDS image. Qtz, quartz; Cr, crystoballite; Rut, rutile; Ep, epidote.
Source: The authors.

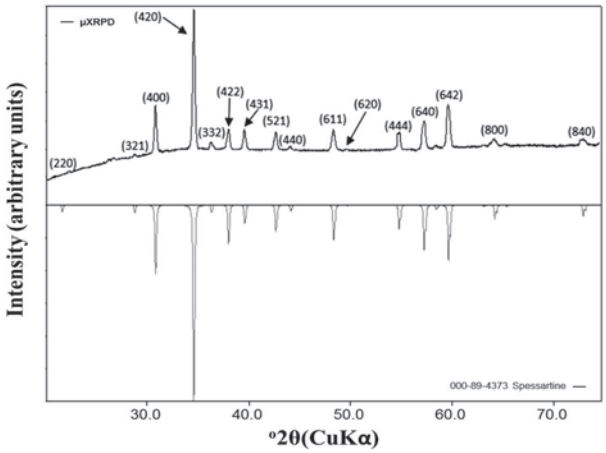


Figure 15. µXRD profile of spessartine-type garnet (λ = 1.5406 Å) in the milled sample, including reflection peaks of mineral inclusions (Qtz, quartz; Cr, crystoballite; Ep, epidote). In the lower part, the XRD diffraction pattern of the spessartine-type garnet, reported by the International Centre for Diffraction Data (ICDD), is shown.
Source: The authors.

µXRD on the unmilled simple reveals the occurrence of spots and discontinue diffraction rings (Fig. 16), which reveals that there is a mixture of crystallites with several sizes and some strong orientations (reflection (840)). The hkl planes observed correspond to the characteristic reflections of the spessartine-type garnet.

In order to observe numerous reflections, it is necessary to rotate the sample at different angles; however, in our case with the eight (8) reflections that were observed, the identification of mineral phases (Fig. 17) was easy.

4. Conclusions

Both scanning electron microscopy and X-ray diffraction techniques are very useful for mineral characterization. Novel analytical methods such as the µXRD technique provide

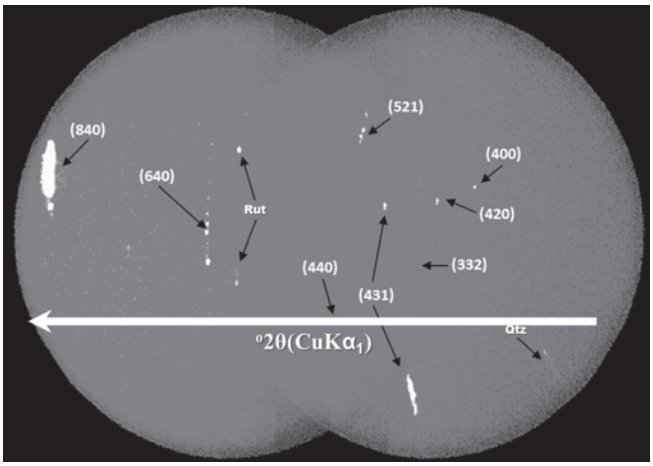


Figure 16. Integration of 2D GADDS image (1D Intensity vs. 2θ plot) and 2D GADDS image of the spessartine-type garnet and accessory mineral phases in the unmilled sample. The diffraction spots and the Debye rings match this structure. Qtz, quartz; Rut, rutile.
Source: The authors.

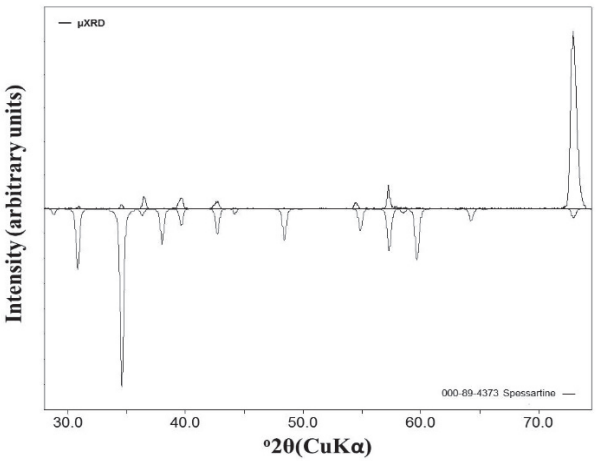


Figure 17. The corresponding µXRD profile of spessartine-type garnet (λ = 1.5406 Å) in the unmilled sample. In the lower part, the diffraction pattern of the spessartine-type garnet is shown, which is reported in the crystallographic database Powder Diffraction File (PDF-2) from the International Centre for Diffraction Data (ICDD). There is a maximum diameter of irradiated sample of 1 mm.
Source: The authors.

the possibility of non-destructive in situ characterization of geological materials with micro spatial resolution. The results in this paper demonstrated the importance of combining such advanced techniques to accurately characterize the distribution of key elements and rock-forming minerals. In this case, analyses were made using (1) petrographic analysis, (2) SEM/EDS analysis, (3) 2-D µXRD (GADDS) representation (4) 1-D µXRD representation. To the best of our knowledge, this is the first study that analyzes garnet in such a way, and the results should prove valuable in the characterization of metamorphic rocks; in particular, to

distinguish the accessory crystalline mineral phases present in garnet.

Acknowledgments

We gratefully acknowledge the Vicerrectoría de Investigación y Extensión of the Universidad Industrial de Santander for the use of their research facilities (scanning electron microscopy and X-ray diffraction). We also thank the Microscopy and X-Rays laboratories at the Universidad Industrial de Santander - Guatiguará Technology Park, and their staff for the analytical service provided for data acquisition. The authors would also like to acknowledge the anonymous referees for their critical and insightful reading of the manuscript and are most grateful to the above-named people and institutions for support.

References

[1] Tissot, R.G., Microdiffraction applications utilizing a two-dimensional proportional detector. Powder Diffraction, 18(2), pp 86-90, 2003. DOI: 10.1154/1.1577354

[2] Flemming, R.L., Salzsauler, K.A., Sherriff, B. L. and Sidenko, N.V., Identification of scorodite in fine-grained, high-sulfide, arsenopyrite mine-waste using micro X-ray diffraction (μXRD). The Canadian Mineralogist, 43, pp. 1243-1254, 2005. DOI: 10.2113/gscanmin.43.4.1243

[3] Harwood, B.P., Flemming, R.L. and Stachel, T., Mapping a mantle xenolith using micro X-ray diffraction. Abstract #MA73B-09, American Geophysical Union, Spring Meeting, 2009.

[4] Catalano, J.G., Heald, S.M., Zachara, J.M. and Brown, Jr., G.E., X-ray microdiffraction study of Uranium speciation in contaminated vadose zone sediments from the Hanford Site, Washington. US Department of Energy Publications, [Online]. Paper 226, 2004. Available at http://digitalcommons.unl.edu/usdoepub/226

[5] Flemming, R.L., Micro X-ray diffraction (μXRD): A versatile technique for characterization of Earth and planetary materials. Canadian Journal of Earth Sciences, 44(9), pp. 1333-1346, 2007. DOI: 10.1139/E07-020

[6] Flemming, R.L., McCausland, P.J.A., Izawa, M.R. and Jacques, N., Reconnaissance micro-XRD studies of meteorites: Rapid in situ mineral identification and textural information. Abstract #2363, 38th Lunar and Planetary Science Conference, Houston, TX, 2007.

[7] Izawa, M.R.M., Flemming, R.L., Banerjee, N.R. and McCausland, P.J.A., Micro X-ray diffraction (μXRD) assessment of shock stage in enstatite chondrites. Meteoritics and Planetary Science, 46, pp 638-651, 2011. DOI: 10.1111/j.1945-5100.2011.01180.x

[8] Dähn, R., Popov, D., Schaub, Ph., Pattison, P., Grolimund, D., Mäder, U., Jenni, A. and Wieland, E., X-ray micro-diffraction studies of heterogeneous interfaces between cementitious materials and geological formations. Physics and Chemistry of the Earth, Parts A/B/C, 70-71, pp. 96-103, 2014. DOI: 10.1016/j.pce.2013.10.010

[9] Pereira, E., Ortíz, F. and Prichard, H., Contribución al conocimiento de las anfibolitas y dunitas de Medellín (Complejo Ofiolítico de Aburrá). DYNA, 73(149), pp. 17-30, 2006.

[10] Ríos, C.A., Castellanos, O.M., Ríos, V. y Gómez, C., Una contribución al estudio de la evolución tectono-metamórfica de las rocas de alta presión del Complejo Arquía, Cordillera Central, Andes Colombianos. Geología Colombiana, 33, pp. 3-22, 2008.

[11] Ruíz-Jiménez, E.C., Blanco-Quintero, I.F., Toro, L.M., Moreno-Sánchez, M., Vinasco, C.J., García-Casco, A., Morota, D. and Gómez-Cruz, A., Geochemistry and petrology of metabasites of the Arquía Complex (Santa Fe de Antoquia and Arquía River, Colombia): Geodynamic implications. Boletín de Ciencias de la Tierra, 32, pp. 65-80, 2012.

[12] Valencia-Morales, Y.T., Toro-Toro, L.M., Ruiz-Jiménez, E.C. and Moreno-Sánchez, M., Pressure-Temperature path of Arquía Group rocks (NW Colombia): A petrographic analysis from mineral assemblages. Earth Science Research Journal, 17(2), pp. 141-149, 2013. DOI: 10.1127/0077-7749/2013/0338

[13] Pardo-Trujillo, A. y Moreno-Sánchez, M., Estratigrafía del occidente colombiano y su relación con la evolución de la provincia ígnea Cretácea del Caribe Colombiano, VIII Congreso Colombiano de Geología, Manizales, Colombia, Agosto, 2001.

[14] Kerr, A., Marriner, G., Tarney, J., Nivia, A., Saunders, A., Thirlwall, M. and Sinton, C., Cretaceous basaltic terranes in Western Colombia: Elemental, chronological and Sr-Nd isotopic constraints on petrogenesis. Journal of Petrology, 38, pp. 677-702, 1997. DOI: 10.1093/petrology/38.6.677

[15] Kretz, R., Symbols for rock-forming minerals. American Mineralogist, 68, pp. 277-279, 1983. DOI: 10.2138/am.2010.3371

[16] Sawada, H., Electron density study of garnets: $Z_3Al_2Si_3O_{12}$ (Z = Mg, Fe, Mn, Ca) and $Ca_3Fe_2Si_3O_{12}$. Journal of Solid State Chemistry, 142(2), pp. 273-278, 1999. DOI: 10.1006/jssc.1998.7983

J.D. Bonilla-Jaimes, received his BSc. in Chemistry in 2010 from the Universidad Industrial de Santander, Bucaramanga, Colombia. He has been working in the X-Ray Laboratory at the Guatiguará Technological Park since 2012. He is member of the Research Group in Structural Chemistry at the School of Chemistry of the Universidad Industrial de Santander, Colombia and is a specialist in X-ray fluorescence and X-ray diffraction techniques and has extensive research experience in characterization of materials.
ORCID: 0000-0002-2079-8043

J.A. Henao-Martínez, received his BSc. in Chemistry in 1982 and MSc. in Chemistry in 1990 from the Universidad Industrial de Santander, Bucaramanga, Colombia. The Universidad de Los Andes, Mérida, Venezuela, awarded him a PhD. in Applied Chemistry in 1996. He has been working as a full-time Lecturer at the School of Geology in the Universidad Industrial de Santander, Colombia, since 1991 where he developed his university teaching over the past 23 years in the field of crystallochemistry. He undertook postdoctoral research at the Instituto Zuliano de Investigaciones (Venezuela) in 2006. He is currently the director of the Research Group in Structural Chemistry at the School of Chemistry of the Universidad Industrial de Santander and the director of the X-Ray Laboratory of the Guatiguará Technological Park. He is a specialist in material characterization in several X-ray diffraction techniques and has extensive research experience in the characterization of materials.
ORCID: 0000-0002-2887-1532

C. Mendoza-Luna, received her BSc. in Physics in 2012 from the Universidad Industrial de Santander, Bucaramanga, Colombia. She is currently undertaking a MSc. in Physics. She has been working in the Microscopy Laboratory at the Guatiguará Technological Park since 2012. She is member of the Research Group in Optics and Signal Processing at the School of Physics at the Universidad Industrial de Santander, and is a specialist in scanning electron microscopy and atomic force microscopy techniques. Her research areas of interest include nanometrology, nanolithography and optoelectronics.
ORCID: 0000-0002-7461-3585

O.M. Castellanos-Alarcón, received his BSc. in Geology in 1999 from the Universidad Industrial de Santander, Bucaramanga, Colombia. He was awarded an MSc. in Geology from the Shimane University, Matsue, Japan, in 2001. He has been working as a full-time Lecturer on the Geology Program (Universidad de Pamplona) since 2003, where he has developed his professional university teaching over the past 11 years in the field of Mineralogy, Metamorphic Petrology and fieldwork on crystalline basement complexes in different areas of Colombia. He is

member of the Research Group on Basic and Applied Geology at the School of Geology, Universidad Industrial de Santander. He is specialist in mineralogy, experimental geology, petrology and the geochemistry of metamorphic rocks, and has extensive research experience in geological mapping, experimental and environmental mineralogy and metamorphic petrology.
ORCID: 0000-0003-0620-0540

C.A. Ríos-Reyes, received his BSc. in Geology in 1989 and his Post-graduate Diploma in University Teaching in 1995 from the Universidad Industrial de Santander, Bucaramanga, Colombia. He was awarded an MSc. in Geology from the Shimane University, Matsue, Japan, in 1999. He was awarded a PhD. in Applied Sciences from the University of Wolverhampton, Wolverhampton, England, in 2008. He has been working as a full-time Lecturer of the School of Geology (Universidad Industrial de Santander) since 1992, and has been teaching in the fields of Mineralogy, Metamorphic Petrology and fieldwork on crystalline basement complexes in different areas of Colombia for the past 22 years. He is currently the director of the Research Group in Basic and Applied Geology at the School of Geology of the Universidad Industrial de Santander and the director of the Microscopy Laboratory of the Guatiguará Technological Park. He is specialist in mineralogy, experimental geology, petrology and geochemistry of metamorphic rocks and has extensive research experience in geological mapping, experimental and environmental mineralogy and metamorphic petrology.
ORCID: 0000-0002-3508-0771

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS

DYNA
http://dyna.medellin.unal.edu.co/

Simultaneous dual true random numbers generator

Josué Aarón López-Leyva [a] & Arturo Arvizu-Mondragón [b]

[a] *Departamento de Ingeniería, CETYS Universidad México, Ciudad de Ensenada, México. josue.lopez@cetys.mx*
[b] *División de física aplicada, Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE) Ensenada, Baja California, México. arvizu@cicese.mx*

Received: October 21th, 2014. Received in revised form: March 13th, 2015. Accepted: December 10th, 2015.

Abstract
This paper details the design and implementation of a simultaneous dual true random numbers generator using only one laser and a digital signal processing system with a DE0 Nano FPGA. We implemented the random generator in such a way that a vacuum optical field will exist in our system. Taking advantage of the inherently random nature of the field, simultaneously quadrature components are measured in order to generate a truly random voltage signal. Also, we used a dynamical system of statistical analysis to eliminate any residual component of direct current on output voltage signal due to an (unavoidable) optical power imbalance in the optical system that was implemented. Finally, were measured the parameters of the auto-correlation and bias probability with values of 0.00010, 0002, respectively, which means that our system can be considered as a true random sequence generator capable of producing two sequences in an independent manner with a bit rate of up to 25 MHz.

Keywords: random sequences; vacuum states; quantum noise.

Generador dual simultáneo de números verdaderamente aleatorios

Resumen
Se presenta el diseño e implementación de un generador dual simultáneo de números verdaderamente aleatorios usando solamente un láser y un sistema de procesamiento digital de señales con FPGA Nano DE0. Implementamos el generador aleatorio de manera que exista un campo óptico de vacío en el sistema; aprovechando la naturaleza inherentemente aleatoria del campo, se miden simultáneamente sus componentes en cuadratura para generar una señal de voltaje verdaderamente aleatoria. Usamos un sistema dinámico de análisis estadístico cuyo objetivo es eliminar cualquier componente residual de corriente continua en la señal de voltaje, ocasionado por un (inevitable) desequilibrio de potencia óptica en el sistema óptico implementado. Se obtuvieron valores de los parámetros de la auto-correlación y el offset de probabilidad de 0,0001 y 0,0002, respectivamente, concluyendo que el sistema puede ser considerado como un verdadero generador de dos secuencias independientes aleatorias a una velocidad de transmisión de hasta 25 MHz.

Palabras clave: secuencias aleatorias; estados del vacío; ruido cuántico.

1. Introduction

Currently, it is very useful to have systems capable of generating random signals and random numbers for diverse fields of knowledge. There are numerous applications that require a Random Number Generator (RNG), such as cryptographic systems, wireless communication simulations, sweepstakes, statistical simulations, etc. Usually, the RNG systems require the configuration of a computer system (or digital control systems) to generate a final "random" digital sequence based on a complex mathematical function [1-4], like those used in various programming languages such as Java, PHP, Matlab, etc. Formally, these kinds of generators are called Pseudo Random Number Generators (PRNG) because they do not produce completely random sequences. Other generators are called True Random Number Generators (TRNG), also known as Hardware Random Number Generators (HRNG), which use a natural noisy signal that exists on the same system or may come from other systems as the noise is obtained from a particular electronic device (hardware) such as resistor, Zener diode, avalanche diode, etc. These generators use various noises present in the systems in order to obtain random sequences and random numbers. Another version of the same RNG systems is called

Free Running Oscillator Random Number Generator (FRO-RNG), in which a random oscillatory signal serves as a source of truly random continuous values from which a final random binary signal will be obtained. Furthermore, there are also systems that generate truly random numbers by means of quantum noise and/or using the phase measurement of an optical signal; these systems fall into the following categories TRNG, PRNG- RNG and FRO-RNG [5]. However, there are some important considerations that must be taken into account in the generators mentioned above, for example the systems that use the Johnson noise as an excitation source are affected by temperature changes; different temperature values produce noisy signals with different amplitudes, therefore an accurate temperature control is required. In the case of the optical systems used to generate random sequences and signals, a control system for the optical intensity (i.e. the number of photons per observation time) is required as this information will be used in the photoreceiver stage [6-9]. If the optical intensity is not controlled it would be required to design and implement other schemes on the photoreceiver side in order to be able to work with a different number of photons. Furthermore, commonly in the TRNG systems only one sequence is generated; therefore, if two or more different random sequences are required, it is necessary to choose one or more of the following options:

a) The more obvious alternative (and maybe the more expensive one) is to use two or more (as required) independent TRNG systems.
b) Use an electronic demultiplexer (DEMUX) in combination with one TRNG system to generate as many sequences as desired; however, there is a trade-off between the rate and the number of sequences to be generated.
c) Obtain the data from a randomized single input stream and store them in a large memory. After this a digital system should be used, such as a high speed FPGA (Field Programmable Gate Array), in order to generate as many independent random sequences, which are required through the different available ports on the FPGA. Obviously, the number of possible sequences generated will depend on the number of available ports and on the processing speed on the FPGA.

When optical systems are used in order to generate random sequences by means of their respective optical noises (such as the phase noise, the amplitude noise and/or the quantum noise), in combination with the optical coherent detection (specifically, the Balanced Homodyne Detection (BHD)), the optical power balance on the different paths of the BHD system is very important (although usually hard to get). This is because an inappropriate optical power balance may produce a saturation of its output signal. It is also necessary to control the DC component of the output voltage signal. In some cases, such control is undertaken by means of a capacitor, although this technique may produce a slight distortion over the output signal [10]. It is important to mention that the obvious way of detecting quantum noise is to directly input the light in a single photodetector and then to analyze the resulting photocurrent using an electronic spectrum analyzer. However, there are limitations such as electronic noise, AC response, efficiencies, power saturation, and saturation of the amplifiers etc. [11]. Because of this we have chosen an alternative: a simultaneous dual true random numbers generator based on the use of the simultaneous detection of the optical vacuum state quadratures that uses balanced homodyne detection. This technique takes advantage of the random nature of the vacuum fluctuations, and at the same time avoids and/or reduces the limitations previously mentioned. The complete system that is described in this paper has a dynamic statistical analysis system implemented in a FPGA DE0 Nano in order to control the DC component that appears because of a (hard to avoid in the practice) inadequate optical power balance. It should be noted that some preliminary results of this hybrid technique in the context of quantum cryptography systems with continuous variables (CV-QKD) has been previously reported on [15]; however, in the present paper we additionally present a detailed mathematical description of its operation as well as more general and conclusive results.

2. Generator implemented

The generator that was used consists of two fundamental stages: the optical subsystem and the digital signal processing subsystem that will be described below in more detail.

2.1. Optical subsystem

The generator that was used is shown in Fig. 1a) and 1b). It consists of a laser source and neutral density filters (F1) to adjust the variance of the detected quantum noise using shot noise units (because these units give us information about the quantum noise of the optical state) and to work in an optical power range acceptable for the BHDs. The laser produces an optical coherent state $|\alpha\rangle$ that is described by the following equation:

$$|\alpha\rangle = \widehat{D}(\alpha)|0\rangle \tag{1}$$

where $|0\rangle$ represents the vacuum state and $\widehat{D}(\alpha)$ the displacement operator in order to generate the coherent state from the vacuum state. Roughly speaking, the vacuum state has a probability function that is described by the equation:

$$|0\rangle \implies W_0(q_0, p_0) = \frac{1}{\pi} e^{(-q_0^2 - p_0^2)} \tag{2}$$

where $W_0(q_0, p_0)$ represents the Wigner function of the vacuum state in both quadrature components q_0 and p_0. Therefore, the probability function of the coherent state is:

$$|\alpha\rangle \implies W(q, p) = \frac{1}{\pi} e^{(-(q-q_0)^2 - (p-p_0)^2)} \tag{3}$$

where q and p gives the average value of the number of photons. $n = \sqrt{q^2 + p^2}$ The optical power beam splitter 50/50 (BS1) is then used to produce the two signals (4) and (5) from which the two truly independent random binary sequences are generated.

$$|\alpha\rangle_1 \implies W(q_1,p_1) = \frac{1}{\pi\sqrt{2}} e^{\left(-(q_1-q_0)^2-(p_1-p_0)^2\right)} \quad (4)$$

$$|\alpha\rangle_2 \implies W(q_2,p_2) = \frac{1}{\pi\sqrt{2}} e^{\left(-(q_2-q_0)^2-(p_2-p_0)^2\right)} \quad (5)$$

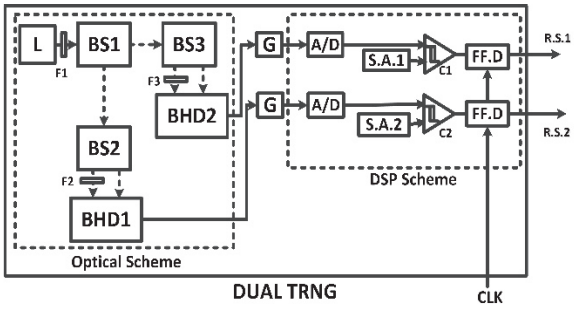
Then the beam splitters (BS2 50/50) and (BS3 50/50) produce two optical signals "balanced", due to the different optical path (fiber length) and optical devices; by adjusting a variable attenuator in one of the optical paths improves optical balance. In this way the marginal distribution functions (6) and (7) are obtained in each quadrature component:

$$W(q_0)\alpha\ \frac{1}{2\pi} e^{(-q_0^2)} \quad (6)$$

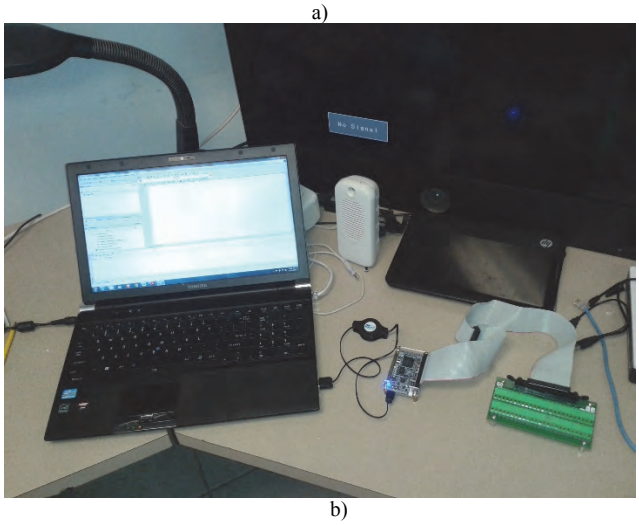
$$W(p_0)\alpha\ \frac{1}{2\pi} e^{(-p_0^2)} \quad (7)$$

Thus, the BHD1 and BHD2 generate two electrical output signals (8) and (9) that represent the quantum noise of the vacuum state in both quadrature components due to the unused port of the beam splitters according to the Q function of the optical quantum state [12].

$$Q(q_0)\alpha\ \frac{1}{2\pi} e^{(-\frac{q_0^2}{2})} \quad (8)$$



a)



b)

Figure 1. DUAL TRNG implemented, a) block diagram with F: neutral density filters, BS: beam splitter, BHD: Balanced Homodyne Detector, G: Electrical Amplifier, A/D: Analog-to-Digital Converter, S.A: Statistical Analysis, C: Digital comparator, FF.D: Flip flop D, R.S: Random sequence, b) Photo of the electronic subsystem of the overall system
Source: The authors.

$$Q(p_0)\alpha\ \frac{1}{2\pi} e^{(-\frac{p_0^2}{2})} \quad (9)$$

These electrical fluctuations are truly random and its bandwidth is determined by the BHDs bandwidth (5MHz for the BHDs used). Therefore, by using the quantum noise, two truly random signals are obtained.

In order to know if the measured noise is really of a quantum nature (which is truly random), we must first determine if the implemented scheme is operating within the standard quantum limit (SQL, Standard Quantum Limit). One way of being sure of this is checking that the variance of the error signal holds a linear relation with the optical power of the laser; additionally it should be above the other noises. If the system is not working in the SQL, the BHDs measure other noises that are not necessarily of a random nature [5]. Fig. 2 shows the measures of the electronics and quantum noises in a time domain, where is possible determinate if the variance of the quantum noise is greater than the electronic noise. Fig. 3 shows the different variances of the quantum noise for different optical power.

These signals are amplified (in our case, each BHD has a maximum gain of 30,000 V / V), and with this gain the noise values shown in Fig 3 and 4 were measured.

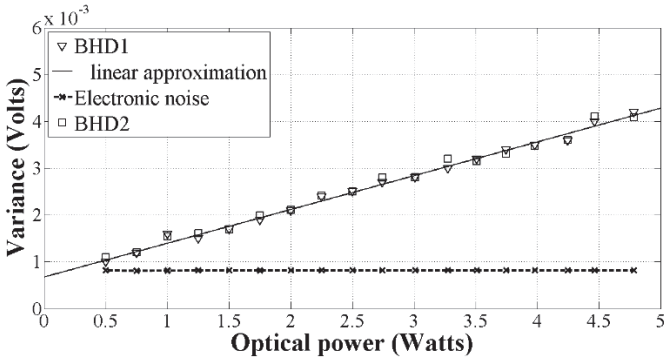


Figure. 2 Measurements of the shot noise in a temporal domain on each BHD.
Source: The authors.

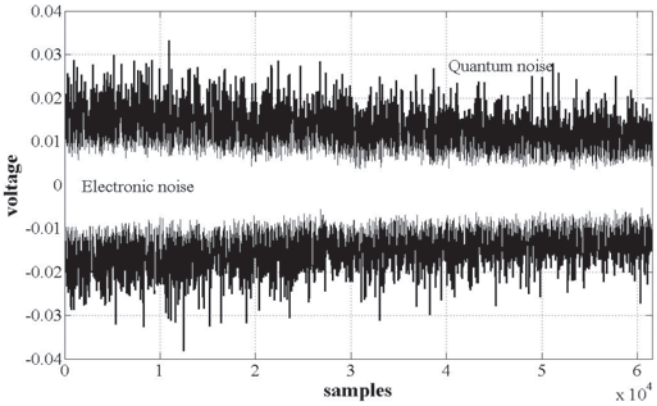


Figure. 3 Measurements of the BHD's electronic and quantum noises.
Source: The authors.

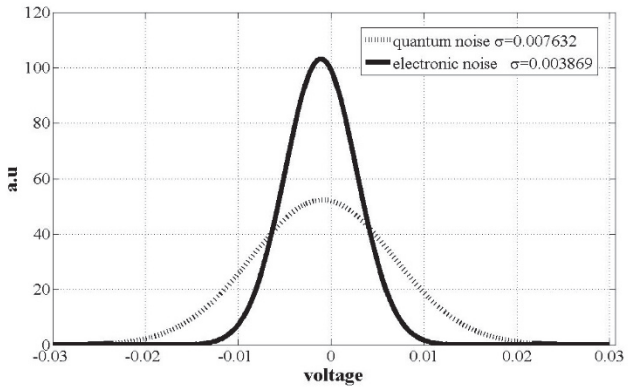


Figure 4. Measurements of the BHD's electronic noise and the quantum noise.
Source: The authors.

2.2. DSP subsystem

A Digital Signal Processing (DSP) subsystem based on the use of a DE0 Nano FPGA is used in order to digitize the electrical signal by means of an analog-to-digital 8-bits converter (ADC). An algorithm for dynamic statistical analysis that determines the average value of the analog random signal (due to a slight and hard to avoid optical imbalance) was implemented in VHDL. By using this analysis the reference signal (the threshold decision) is determined and used in order to generate a true random binary sequence. Furthermore a flip-flop D block (in VHDL) is used in order to vary the bit rate of the final binary sequence from 1 to 10 Mbps. This bit rate is determined by the FPGA clock (up to 50 MHz in our case). The algorithm flowchart is shown in Fig. 5. First, the programs set all the inputs and outputs variables in the *entity* of the VHDL. Next, many *architectures* are used in the same program to perform different tasks such as: a) *ADC input ports* (this stage converts the analog data into digital data for the next digital signal processing), b) *Digital data load in the memory*, (here the converted data are loaded in the internal memory of the FPGA), c) *Digital data average,* (in order to obtain the average value of the digital data that corresponds to the DC component), d) *Threshold comparator,* (in order to generate the final random binary sequence), and e) *Setting of the transmission rate,* (in order to change the bit rate of the final random binary sequence). Roughly speaking, a VHDL program was designed for a frequency divider in order to obtain a different transmission rate for the final sequence.

3. Test and measurement of the performance

As mentioned above for an adequate performance of our scheme, the measurement of the quantum noise is very important. One way to measure the quantum noise in such systems is through the variance of the photoelectric signal in the BHDs. The BS2 and BS3 send classical noise onto both BHD detectors and, thus, the photocurrents are correlated to each other. For quantum noise the effect of the beam splitter is different, i.e. the two resulting photocurrents are not

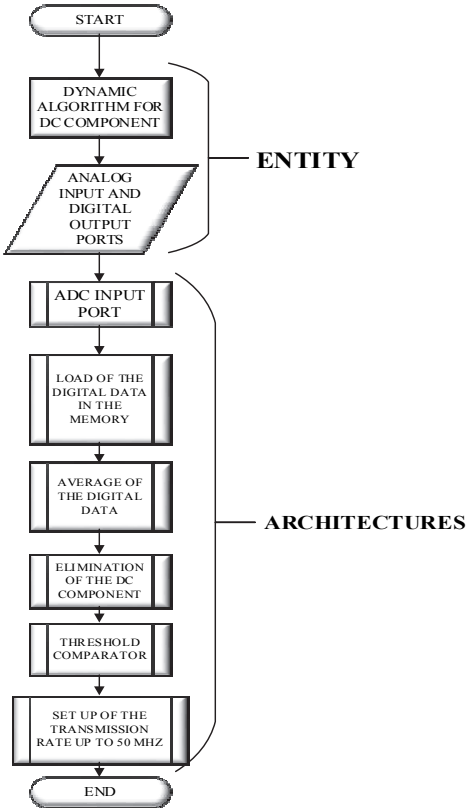


Figure 5. Flowchart of the algorithm implemented in FPGA.
Source: The authors.

correlated [11-13]. Fig. 3 presents the electronic noise measurements (white trace) and quantum noise (black trace) in the temporal domain in which is possible to distinguish the different variance values between them. However, because the setting of an adequate optical power is required, it will be necessary to change the optical power; the variance of the quantum noise also must also be modified. The probability function of the noises measured was shown in the Fig. 4. As can be seen in the figure, the variance (or standard deviation) of the quantum noise is greater than the variance of the electronic noise, which ensures that the experiment is working in the Standard Quantum Limit. In Fig. 2 we present the measurements for the quantum noise variance for different optical powers, and as expected, there is a linear dependence of the latter with the optical power. In this case, we used a laser with an average optical power of 5 mW. Thus, it was determined that the experiment was working at the standard quantum limit 18 dB above of the electronic noise. Measurements of quantum noise variance were performed by means of an oscilloscope with 50,000 samples at $4x10^9$ samples/sec. The above behavior may be modeled using the following function, $y = ax + b$, where y is the total noise in Volts (V), x is the optical power of the local oscillator in watts (W), a is a factor associated with the conversion factor of the photodetectors (in our case a =0.66 V/mW), b =0.8 mV is related to the electrical noise present in Volts. Thus, for an optical power of the local oscillator of 5 mW the r.m.s voltage is 4.35 mV. It is important to mention that the r.m.s voltage in a Gaussian density function with zero expected mean is equal

to its variance. Therefore, an optical power of 5 mW in the local oscillator corresponds to a variance of 4.35 mV.

Regarding the performance of the system, there are fundamentally two parameters: the probability bias (bp) and the auto-correlation of the binary sequences generated. Sometimes, due to imperfections in measurement and hardware, the bits generated by an RNG system will contain a non-zero value of auto-correlation and probability bias, which may be minimized using the Von Neumann method [14]. Thus, the probability bias (bp) is defined as $bp = p(1) - p(0)$ where $p(1)$ and $p(0)$ represent the probability of the binary values, 1 and 0, respectively, after having stored a large amount of bits. The best value would be bp=0 which means that the binary values are equiprobable; however, we must be sure that the value of autocorrelation is close to zero to ensure randomness [8]. The results obtained with our scheme show a value of bp=0.0002. The Fig 6 shows the performance of bp in a considerable observation time. As mentioned above, this value may be improved using the von Neumann method; however, in this case we considered that because we have obtained a very low value, an improvement was not required. Regarding the measurement of the discrete-time autocorrelation $r_{12}(j)$ of the binary sequence, we used the equation (10) in which j is the lag between the sequences x_1 and x_2 that are analyzed with a length N. Thus $r_{12}(j)$ is defined as:

$$r_{12}(j) = \sum_{n=-\infty}^{\infty} x_1(n)x_2(n+j) \qquad (10)$$

Then, the equation (10) is normalized and we obtain the equation:

$$\rho_{12}(j) = r_{12}(j) \Big/ \left\{ \frac{1}{N} \sqrt{\sum_{n=0}^{N-1} x_1^2(n) \sum_{n=0}^{N-1} x_2^2(n)} \right\} \qquad (11)$$

The overall system performance may be obtained through the probability bias and autocorrelation values. In order to do this we performed measurements of up to 10 Mbps with different optical powers. By varying the optical power it is possible to show that the system of statistical analysis is capable of working adequately even if the variance of the quantum noise has variations. The Figs. 6 and 7 show that the values of probability bias and autocorrelation are 0.0002 and 0.0001 respectively to measure ranges from 1 to 5 mW and from 1 to 10 Mbps.

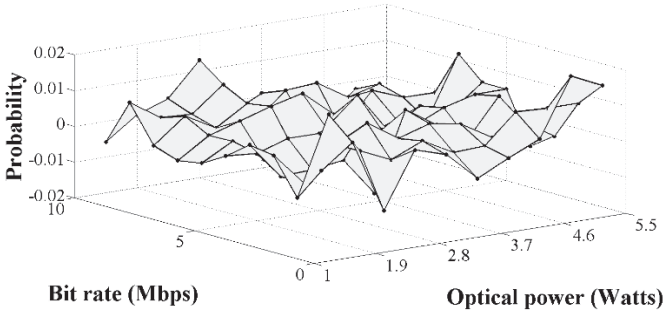


Figure 6. Probability bias measurement for j = 100 and different bit rates (1-10Mbps).
Source: The authors

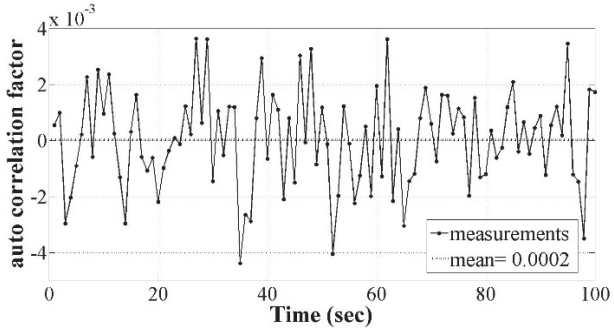


Figure 7. Correlation factor measurement for j = 100 at10 Mbps
Source: The authors.

3. Conclusions

This work presents a mathematical description, and the design and implementation of a True Random Numbers Generator based on a system that simultaneously measures both quadrature components of an optical field (the vacuum state) and also uses a dynamic statistical analysis. As mentioned above, some preliminary results of this hybrid technique in the context of quantum cryptography systems with continuous variables (CV-QKD) have previously been reported on [15]. However, in this paper we have presented more general and conclusive results in terms of the mathematical model and the development of a generator suitable to be used in different applications such as cryptographic systems (already mentioned), simulations, sweepstakes, etc., using the concepts of modern photonics [16]. In the system developed, the transmission rate of the final sequences may be independently configured for different simultaneous applications. In our experimental set-up we were able to get adequate values (to assure the randomness of the signals generated) of probability bias (0.0002) and autocorrelation (0.0001) for a bit rate of up to 10 Mbps. The bit rate obtained depends on the bandwidth of the BHDs and DSP subsystem used. Therefore, implementing a system with a higher rate is possible by means of changing such subsystems (there are currently BHDs capable of operating on the order of the GHz as well as FPGAs and DSPs that have very high processing speeds).

Acknowledgments

The authors wish to thank the support they were given from CETYS University and CICESE Research Center. This work was supported by a CONACYT Basic Science Grant.

References

[1] Blaner, B., Abali, B. Bass, B.M., Chari, S., Kalla, R., Kunkel, S., Lauricella, K., Leavens, R., Reilly, J.J. and Sandon, P.A., IBM

POWER7+ processor on-chip accelerators for cryptography and active memory expansion, IBM Journal of Research and Development, 57(6), pp.1-16, 2013. DOI: 10.1147/JRD.2013.2280090

[2] Francillon, A. and Castelluccia, C., TinyRNG. A cryptographic random number generator for wireless sensors network nodes. 5th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks and Workshops, [Online]. pp. 1-7. Limassol. 2007. Available at: http://10.1109/WIOPT.2007.4480051

[3] Yang M., Guo, Q. and Wang, Z., A random number generator based software channel simulator for land mobile satellite channel. International Conference on Wireless Communications, Networking and Mobile Computing. pp. 1095-1098 Shanghai. 2007. DOI: 10.1109/WICOM.2007.280

[4] Chen, I,-T., Tsai, J.-M. and Tzeng, J., Audio random number generator and its application. International Conference on Machine Learning and Cybernetics (ICMLC), pp.1678-1683. Guilin. 2011. DOI: 10.1109/ICMLC.2011.6017002

[5] Stipcevic, M., Quantum random number generators and their use in cryptography, MIPRO Proceedings of the 34th International Convention, [Online]. pp. 1474-1479, 2011. Available at: http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=1354136

[6] Zhu, Y., He, G. and Zeng, G,. Unbiased quantum random generator based on squeezed vacuum state. Int. J. Quantum Inform., [Online]. 10(1), pp. 1-13, 2012. Available at: http://www.worldscientific.com/doi/abs/10.1142/S0219749912500128

[7] Stipčevića, M. and Medved-Rogina, B., Quantum random number generator based on photonic emission in semiconductors, Review of scientific instrument, [Online]. 78, pp. 1-7, 2007. Available at: http://scitation.aip.org/content/aip/journal/rsi/78/4/10.1063/1.2720728

[8] Fürst, H., Weier, H., Nauerth, S., Marangon, D.G., Kurtsiefer, C. and Weinfurter, H., High speed optical quantum random number generation, Optics Express. 18(12), pp. 13029-13037. 2010. DOI: 10.1364/OE.18.013029

[9] Wayne, M.A. and Kwiat, P.G., Low-bias high-speed quantum random number generator via shaped optical pulses, Optics Express, 18(8), pp. 9351-9357, 2010. DOI: 10.1364/OE.18.009351

[10] Sanford-Williams C.R., Optoelectronic experiments on random bit generators and coupled dynamical systems. PhD thesis. Department of Physics, University of Maryland, USA, 2013.

[11] Bachor, H.A. and Ralph, T.C., A guide to experiments in quantum optics. Wiley VCH., [Online] Chapter 8, 2004. Available at: http://onlinelibrary.wiley.com/book/10.1002/9783527619238

[12] E. Ip, et al., Coherent detection in optical fiber systems, Optics Express, 16(2), pp. 753-791, 2008. DOI: 10.1364/OE.16.000753

[13] Yuen, H.P. and Chan, V.W.S., Noise in homodyne and heterodyne detection. Optics letters. 8(3), pp.177-179, 1983. DOI: 10.1364/OL.8.000177

[14] Abbott, A.A. and Calude, C.S., Von Neumann normalisation of a quantum random number generator, Journal of Computability, [Online]. 1(1), pp. 59-83, 2012. Available at: http://content.iospress.com/articles/computability/com001

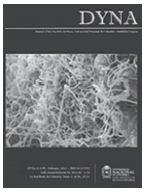
[15] Lopez, J.A., Dual quantum random number generator using a FPGA for QKD-CV systems: Preliminary results. International Journal of Emerging Research in Management &Technology, [Online]. 3(6), pp. 6-8.2014. Available at: http://www.ermt.net/docs/papers/Volume_3/6_June2014/V3N6-138.pdf

[16] Martin-Pereda, J.A., La fotónica: Ayer y mañana. Revista Dyna Ingeniería e Industria, 89, pp.501-503. 2015. DOI: 10.6036/7116

J.A. Lopez-Leyva, obtained his BSc. degree with an emphasis on Telecommunications from the Superior Technology Institute of Cajeme (ITESCA) in Sonora Mexico. From 2006 to 2008 he worked on Networking and Telephony projects. Finally, he obtained his PhD. in quantum communication using satellites and quantum cryptography at the CICESE Research Center in Baja California, Mexico. His current research interests include free space optical communications, Coherent optical communications, optical networks, statistical signal processing and quantum cryptography systems.
ORCID-ID: 0000-0002-3004-5686

Arturo Arvizu-Mondragón, received his BSc. and MSc. degrees in Electronics in 1985 and 1990, respectively, from the Universidad Nacional Autónoma de México, Mexico and his PhD. in Telecommunications in 2000, from the CICESE Research Center, Ensenada, BC, Mexico. In 1987 he joined the Institute of Electrical Research, Cuernavaca, Morelos, México, working on projects relating to the optical and optoelectronics communications systems that are applied in power generation systems, and laboratories to test, measure and characterize electrical systems. In 1992 he joined the CICESE Research Center where he currently works in the fields of quantum communications, optical fiber and optical wireless communications with coherent detection. In 2000 and 2001 worked in a post-doctoral position in the telecommunications department at the École nationale supérieure des télécommunications, Paris, France.
ORCID-ID:
0000-0001-6926-2197

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS



DYNA
http://dyna.medellin.unal.edu.co/

Modeling waste management in a bioethanol supply chain: A system dynamics approach

Danny Waldir Ibarra-Vega [a]

[a] *Escuela de Ciencias Exactas e Ingeniería, Ingeniería Ambiental Universidad Sergio Arboleda, Bogotá, Colombia. danny.ibarra@usa.edu.co*

Received: November 25th, 2014. Received in revised form: March 13th, 2015. Accepted: December 10th, 2015.

Abstract
This paper presents a model and simulations, which was built with a System Dynamics methodology applied to waste management in the bioethanol supply chain in Colombia. The stages of the supply chain used were cane planting, production and the distribution process. The residues taken into account were bagasse and vinasse. Moreover, four simulation scenarios were performed in which the values of variables modified management strategies. The installed bioethanol production capacity was increased to observe the evolution of increased waste generation evaluated. The initial conditions for the simulation were modified to represent a production of about 2,500,000 liters / day. This leads to an increased generation of vinasse of over 400 million liters on average in a month and around more than 1 million tones of bagasse. The results are used to estimate the impact of management strategies on the amount of waste generated in the production of bioethanol.

Keywords: manuscript formatting; camera-ready manuscript.

Modelamiento de la gestión de residuos en la cadena de suministro de bioetanol, enfoque con dinámica de sistemas

Resumen
En este artículo se presenta un modelo y sus simulaciones, construido con la metodología de Dinámica de Sistemas, aplicado a la gestión de residuos en la cadena de suministro de bioetanol en Colombia. Las etapas de la cadena de suministro utilizadas fueron siembra de caña, proceso de producción y distribución. Los residuos que se tuvieron en cuenta fueron el bagazo de caña y las vinazas. Se realizó la simulación de cuatro escenarios en los que se modificó los valores de las variables denominadas estrategias de gestión. Se aumentó la capacidad instalada de producción de bioetanol para ver la evolución temporal del aumento de la generación de los residuos evaluados. Las condiciones iniciales para la simulación fueron modificadas para representar una producción de 2,500,000 litros/día. Esto permite que se incremente la generación de vinazas por encima de 400 millones de litros en promedio en un mes y alrededor de más de 1 millón de toneladas de bagazo. Con los resultados se permite estimar el impacto de las estrategias de gestión en el número de los residuos generados en la producción de Bioetanol.

Palabras clave: Dinámica de sistemas, Modelamiento, Bioetanol, Residuos

1. Introduction

Supply chain management is a rather wide subject. Over recent years researchers have reviewed it from different points of view. [1]. The concept of sustainable supply chain is growing within supply chain management, which is characterized by the integration of environmental and social objectives that address the economic dimension of the chain [2]. The use of quantitative models for decision-making can be found in research conducted to link the social and environmental issues in the supply chain [3]. Models are either a simplified representation or abstraction of reality. They are based on a set of variables and causal relationships [4,5].

The use of models and simulations allows the inclusion of external issues such as environmental variables. These models are very useful in Green Supply Chain Management (GSCM). The GSCM can be defined by the inclusion of environmental thinking in the process. This ranges from the integration of product design, selection and provision of raw

materials, production, distribution and delivery of the final product to consumers to the end of the products' life cycle and reverse logistics [6]. In order to accomplish a green supply chain it is necessary to minimize, or preferably, eliminate the negative effects that the supply chain generates on the environment. This requires the use of environmentally friendly materials and minimizes waste generated. One of the environmental aspects to be considered in the area of sustainability is the management of wastes that are generated in production processes.

Agro-industrial waste management has gained attention in different countries due to improper handling and final disposal in the environment [7]. While handling of waste is defined in terms of reuse and exploitation, this does not reach 100% of the generated waste. As a result, increasing the production of bioethanol (EtOH) will also directly increase the generation of waste, and the problem of waste management.

According to the above, we propose a hypothesis which uses modeling and simulation tools, such as system dynamics, that can be associated with external factors, such as environmental variables, specifically waste management in the supply chain.

The model developed in this study focuses on the fields of biofuels, specifically bioethanol. The residues that were taken into account to build the model were: sugarcane bagasse and vinasses.

2. Description of the system to study

Bioethanol is a type of biofuel produced from the alcohol fermentation of sugars in agricultural crops or crop residues. This is by far the most technologically mature biofuel derived from microorganisms and is a good candidate to replace fossil fuels [8]. In Colombia, bioethanol is produced from sugar cane. The production of this type of plant is well established in the country and has greater energy efficiency compared to other raw materials from which the bioethanol is produced. Its production in Colombia is carried out mainly in the Cauca River Valley, primarily in the departments of Cauca, Valle, Risaralda and Caldas, which encompass 47 municipalities [9].

Information on Colombian installed capacity for bioethanol production (1,250,000 liters/day) was taken as the basis of this study. Variables associated with the supply chain were also considered, as well as the number of planted sugarcane hectares intended for bioethanol, the milling capacity, the performance of cane per hectare, the inventory of bioethanol in factories, bagasse generation, vinasse generation, and the inventory of bioethanol for wholesale, among others.

Since bioethanol in the country was first produced, the generation of waste or by-products such as bagasse and vinasse has been identified. Vinasse is the wastewater generated from the production of ethanol, either from sugar cane, corn, wheat, yucca, or lignocellulosic residues. These are associated with the amount of bioethanol produced. The ratio of generation is on average between 10 and 15 liters per each liter of bioethanol, depending on each factory equipment [10]. Moreover, another by-product of this process that is involved in this system is the bagasse that is generated in the milling of juice production. Bagasse can be used as formation of animal food or as an element for energy cogeneration, and therefore, reduces energy costs [11].

Table 1.
Basic Scenario Simulation.

PARAMETERS	VALUES
Hectares	35000 Ha
Crop Yield	80 Ton/Ha
Milling Yield	70 L/Ton
Bagasse Generation	28,60%
Vinasse Generation	8 L/L ethanol
Sowing Time	11 months

Source: The author

The two main residues, vinasse and bagasse, were considered. Scenarios with different input values obtained from the Federation of Biofuels reports in 2014 [12] were established for these residues. Table 1 shows the information on the basic parameters of the simulation. The simulation time limit is one hundred and fifty days (150), i.e. five months of production.

The following are the main attributes that were identified in obtaining and delimiting the system under study. They describe the component parts of the bioethanol supply chain, which has the most impact on waste generation.

- Hectares of Cane: The number of sugarcane planted hectares that are intended for the production of bioethanol.
- Planting Cane: The planting of the required sugarcane for bioethanol production.
- Cane harvest: The number of hectares harvested and intended for the production of bioethanol.
- Milling to get Juice: Milling process of harvested cane.
- Cane Bagasse: Bagasse quantity generated in the milling stage.
- Fermentable Juice: Quantity of cane juice intended for fermentation.
- Bioethanol Production (EtOH): Production process according to production rate of fermentable juice.
- Bioethanol Inventory: Accumulation in liters of produced EtOH.
- Vinasse Generation: Quantity of vinasse generated in EtOH production.

3. System Modeling

The modeling was undertaken using the System Dynamics methodology. The steps proposed by Forrester [13-14], and Aracil & Gordillo [15] were followed in the process. It also took into account the methodology followed by Callejas et al [16]. Therefore, we started by building a causal loop diagram, and then moved on to the formulation of a Levels and Flow diagram in order to obtain equations representing the model. Finally, the simulation was carried out. We used the Vensim Ple [17] in this study.

3.1. System Dynamics

System Dynamics is a methodology for analysis and problem solving. It was developed by Jay Forrester at the

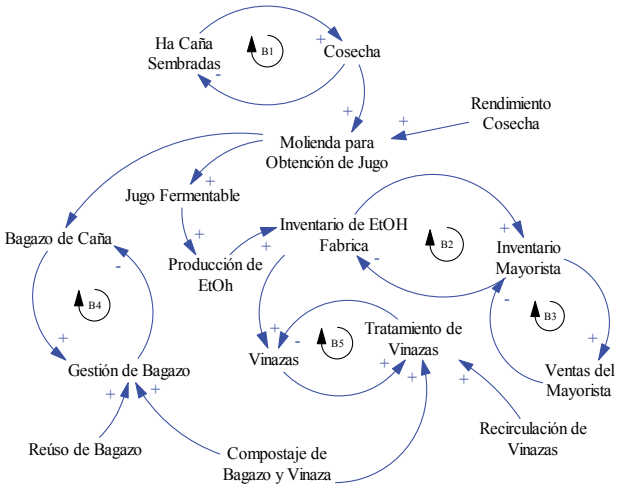


Figure 1 Causal loop Diagram of the Problem.
Source: The author

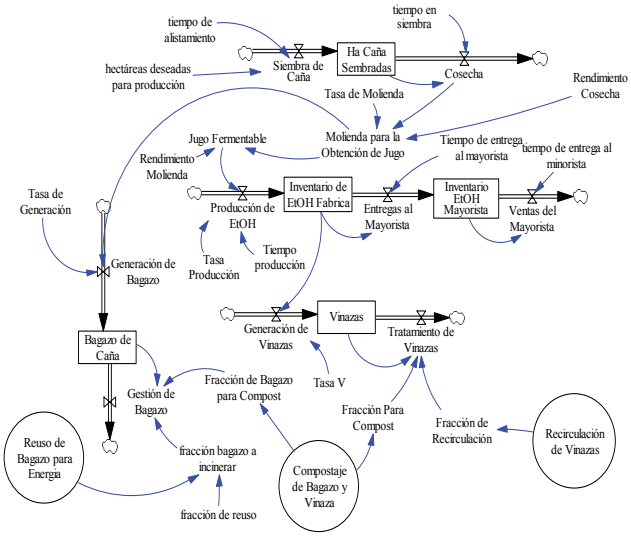


Figure 2. Levels and Flow Diagram of the problem.
Source: The author

Massachusetts Institute of Technology (MIT). He presented this methodology in his research [13-14]. It facilitates the of learning complex systems through formal models and simulation methods [18]. In System Dynamics, any aspect of the world is conceived as a causal interaction between attributes that describe it. Thus, systemic representations with arrows and dots, called causal loop diagrams, are built. They capture all the hypotheses proposed by the modeler. From these hypotheses we can learn about the system in order to take part in the decision making. Subsequently, the Levels and Flow diagram is obtained, in which a quantification of the causal loop diagram is undertaken. We can later obtain a set of equations that allow the decision maker to see the system of interest's behavior through specialized software simulations [19].

It should be noted that the background to this research reveals a growing interest in environmental issues in recent years and, in particular, an interest in environmental issues applied to supply chains management [20].

3.2. System conceptualization

In this stage, causal relationships that can be identified in each one of the problem's attributes are described. Fig. 1 shows the results in the causal loop diagram.

Fig. 1 shows the conceptualization of the problem expressed with feedback loops, which demonstrates the structure of the system.

B1 represents the following relationship: if more hectares of sugarcane are planted there are more hectares harvested; if more are harvested, fewer hectares of sugarcane are planted.

In turn, more harvested cane that signifies more raw materials for the production of juice by milling, which increases the fermentable juice to produce bioethanol (EtOH).

B2 represents EtOH inventory increasing with the production of EtOH. Given the dynamics of this market in Colombia, it is assumed that everything that is produced goes directly to the wholesale inventory.

In B3, a constant sale of EtOH is assumed. Therefore, it can be understood that if there is a higher wholesale inventory, there will be more wholesale sale.

B4 represents cane bagasse generation. This comes from the milling stage in order to obtain juice. It is stated that the more milling there is to obtain juice the higher the production of cane bagasse. Therefore, if there is more bagasse generated, there is a requirement for more management; if there is more management, there will be a lesser quantity of bagasse generated.

B5 represents vinasse generation. It is obtained in the production process and it is associated with the number of liters of EtOH produced. This vinasse requires treatment because of its high pollutant content. The more vinasse that is treated the less the quantity of untreated vinasse.

3.3. Formulation

In this phase, the causal loop diagram shown in Fig. 1 is reworked and it becomes a Levels and Flow diagram, which is known as "the language" of the system dynamics simulation. Fig. 2 shows the model developed in Vensim.

The generation of waste in the production process of bioethanol is one of the most important environmental aspects when considering sustainability. While it is true that these residues are susceptible to use and treat, it is necessary to evaluate what will happen to the amount of waste that is generated when the Colombian government continues to increase the amount of bioethanol in the country.

In Table 2 and Table 2.1 the model equations and arithmetic relations that were considered for the simulations and management strategies are presented.

4. Model Behavior

In this study, four different scenarios which enabled the understanding of the system behavior were established. The scenarios were constructed by changing the values of bagasse reuse strategies for energy and composting bagasse and vinasse (0-1). Similarly, the bioethanol production capacity was increased to twice the current production (installed capacity 1,250,000 liters / day) and the behavior of the amount of managed and unmanaged waste was observed.

Initially, the model was validated with real data, the current bioethanol production capacity in the country and also with an initial simulation for this capacity. As shown in Fig. 3, the Y axis displays the amount of liters of ethanol generated, and the X axis shows the time expressed in 6 months.

In Figs. 4, 5, 6, and 7 the simulation results of the generation and management of the main waste produced in the supply chain of bioethanol can be found. The Y axis is the number of liters of ethanol produced, the number of vinasses liters generated and the amount of cane bagasse in tonnes. The X axis is the simulation time in days.

Table 2. Model equations

Variable Name	Type	Formula	Unit
Sugarcane sown Hectares	Levels	Sugarcane planting – Harvest	Hectares
EtOH Inventory -Factory		EtOH Production –Deliveries to Wholesale	Liters
EtOH Inventory - Wholesale		Deliveries to Wholesale – Sales to Wholesale	Liters
Bagasse		Bagasse Generation – Bagasse Management	Tons
Vinasse		Vinasse Generation – Vinasse Treatment	Liters
Sugarcane planting	Flows	Hectares desired for Production / Induction time	Hectares
Harvest		Sugarcane sown Hectares / Sowing Time	Hectares
EtOH Production (Bioethanol)		Fermentable Juice * Production Rate/Production Time	Liters
Deliveries to Wholesale		EtOH Inventory –Factory /Delivery Times to Wholesale	Liters
Sales to Wholesales		Wholesale EtOH Inventory/Delivery Times to Retail	Liters
Vinasse Generation		Factory EtOH Inventory *V Rate	Liters
Vinasse Treatment		Vinasse*(Recirculation Fraction + Fraction for Compost)	Liters
Bagasse Generation		Milling to Obtain Juice *Generation Rate	Tons
Bagasse Management		Bagasse *(Bagasse for Incineration Fraction + Fraction of bagasse for Compost)	Tons
Hectares desired for Production	Parameter	External Data	Hectares
Induction time		External Data	Months
Sowing Time		External Data	Months
Crop Performance		External Data	Ton / Hectare
Milling Performance		External Data	Liter / Ton
Production Time		External Data	day
Production Rate		External Data	Dimensionless
Delivery Times		External Data	day
V Rate (Vinasse Generation Rate)		External Data	Liters of Vinasse /Liter of EtOH
Milling to Obtain Juice	Auxiliary	Crop Performance * Crop)* Milling Rate	Tons
Fermentable Juice		Milling to Obtain Juice * Milling Performance	Liters

Source: The author

Table 2.1.
Equations of Management Strategies

Variable Name	Type	Formula	Unit
Reuse of bagasse for energy	Auxiliary	0= Do not implement, 1= Implementation	Dimensionless
Fraction of bagasse to incinerate	Auxiliary	Reuse fraction * Reuse of bagasse for energy	%
Bagasse and Vinasse Compost	Auxiliary	0= Do not implement, 1= Implementation	Dimensionless
Fraction of Vinasse for compost	Auxiliary	(Bagasse and Vinasse compost)*0.3	%
Fraction of Compost to reuse	Parameters	External Data	%
Reuse Fraction	Parameters	External Data	%

Source: The author

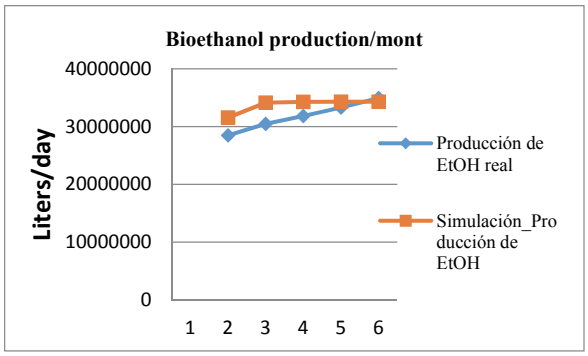


Figure 3. Model validation with historical production data.
Source: The author

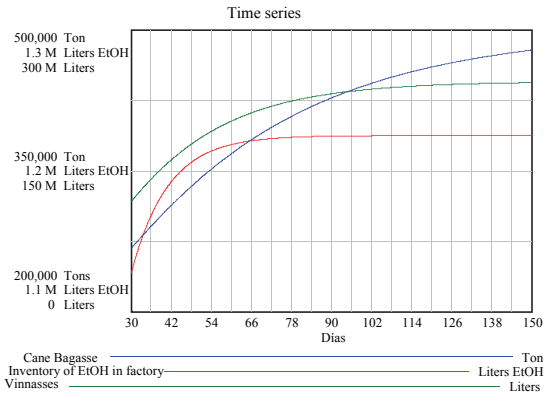


Figure 4. Scenario 1: initial conditions
Source: The author

4.1. Behavior Analysis

Fig. 4 shows that the amount of waste generated is actually much higher compared with the production of bioethanol in an average month. This first scenario shows a simulation with no management strategies for cane bagasse reuse, composting of bagasse and cane, and vinasses recirculation to the process. These strategies had an initial value = 0.

Fig. 5 shows a simulation with the initial conditions of the first scenario, but with waste management strategies having been implemented. These had a value = 1; this is shown in Table 2.

Behavior shows waste reduction, of both vinasse and bagasse. The proportion of bagasse to be incinerated is 70% as the other 30% can be destined to marketing.

Fig. 6 shows a third scenario following the country's interest in increasing production. Initial conditions for the simulation were modified to represent a production of about 2,500,000 liters / day. This leads to increased generation of vinasse of over 400 million liters on average per month and around more than 1 million tonnes of bagasse. In this scenario, management strategies had a value = 0. It is worth mentioning that the System Dynamics models are not used to forecast but to observe and evaluate behavioral tendencies. The major concern in this scenario what would happen if the vinasses that was generated was not treated. Where would it be discharged?

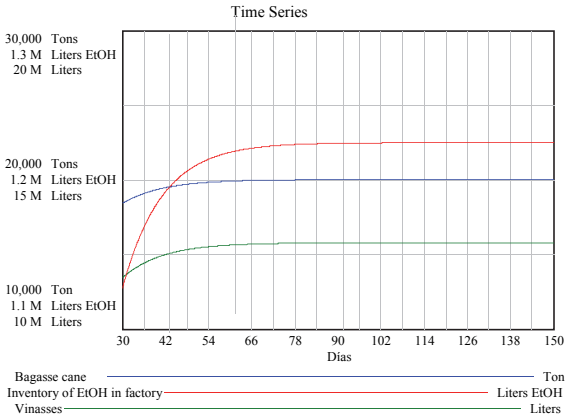


Figure 5. Scenario 2, Implementation of Management Strategies
Source: The author

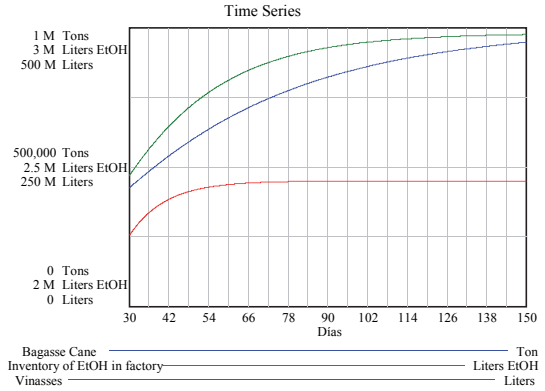


Figure 6. Scenario 3, increased hectares of sugarcane and bioethanol production.
Source: The author

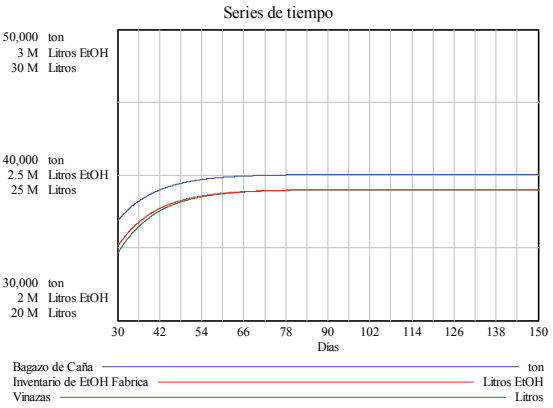


Figure 7. Scenario 4, implementation of strategies with a higher production of bioethanol.
Source: The author

Fig. 7 shows a simulation with the initial conditions of the third scenario. However, this time waste management strategies are employed. As in scenario 2, strategies have a value = 1. The trend shows a significant waste reduction in

comparison to Fig. 6. However, there are still much untreated waste that may cause environmental hazards.

5. Conclusions and Future Work

Taking into consideration the Colombian policy of expanding the supply of biofuels, it is necessary to focus on the environmental importance of the bioethanol supply chain. The higher the production capacity the greater the environmental changes that would put pressure on the environment. The Colombian government is currently seeking the expansion of bioethanol production plants.

In bioethanol models, it is necessary to link the energy efficiency from the bagasse combustion process with its calorific value/ power and the performance of the steam generated. This is necessary in order to give the model more accuracy and to be able to evaluate more scenarios prospectively. Similarly, it is necessary to link other waste generated in the supply chain such as *chachaza* (the residue left after filtering the pressed cane juice)**,** with the oils used.

The model illustrates different scenarios or future trends in waste generation in the bioethanol industry. This is due to the change of parameters and the initial conditions that vary, depending on the information that is to be used in the model.

It was possible to vary the behavior of bioethanol production represented by the model with real information, which was taken from the Federation of Biofuels. Even though a correlation between the data is perceived, it is necessary to continue measuring the model to find a more accurate trend.

With the system dynamics methodology it is possible to create models that can be linked to environmental variables and simulate future scenarios of potential impacts. With his methodology, differential equations that show the behavior of the model variables are constructed. These equations can be integrated with other methodologies, such as Dynamical Systems, Life Cycle Analysis and Material and Energy Balances, in order to have a rigorous methodology to assess the sustainability of bioethanol supply chains or other biofuels.

References

[1] Beske, P., Land, A. and Seuring, S., Sustainable supply chain management practices and dynamic capabilities in the food industry: A critical analysis of the literature. International Journal of Production Economics. 152, pp. 131-143, 2014. DOI: 10.1016/j.ijpe.2013.12.026

[2] Seuring, S., Müller, M., From a literature review to a conceptual framework for sustainable supply chain management. Journal of Cleaner Production. 16(15), pp. 1699-1710, 2008. DOI: 10.1016/j.jclepro.2008.04.020

[3] Branderburg, M., Govindan, K., Sarkis, J. and Seuring, S., Quantitative models for sustainable supply chain management: Developments and directions. European Journal of Operational Research 233, pp. 299-312, 2014. DOI: 10.1016/j.ejor.2013.09.032

[4] Bertrand, J.W.M. and Fransoo, J.C., Operations management research methodologies using quantitative modeling. International Journal of Operations & Production Management, 22(2), pp. 241-264, 2002. DOI: 10.1108/01443570210414338

[5] Meredith, J., Theory building through conceptual methods. International Journal of Operations & Production Management, 13(3), pp. 3-11, 1993. DOI: 10.1108/01443579310048182. DOI: 10.1108/01443579310028120

[6] Srivastava, S., Green supply-chain management: A state-of-the-art literature review. International Journal of Management Reviews, 9(1), pp. 53-80, 2007. DOI: 10.1111/j.1468-2370.2007.00202.x

[7] Christofoletti, C., Escher, J., Correia, J., Urbano, J. and Fontanetti, C., Sugarcane vinasse: Environmental implications of its use. Waste Management, 33, pp. 2752-2761, 2013. DOI: 10.1016/j.wasman.2013.09.005

[8] Zerva, A., Savvides, A.L., Katsifas, E.A., Karagouni, A. and Hatzinikolaou, D., Evaluation of Paecilomyces variotii potential in bioethanol productionfrom lignocellulose through consolidated bioprocessing. Bioresource Technology 162, pp. 294-299, 2014. DOI: 10.1016/j.biortech.2014.03.137

[9] CUE. Consorcio,. Sostenibilidad de biocombustibles en Colombia. Banco Interamericano de Desarrollo (BID) Ministerio de Minas y Energía. Medellín. 2012.

[10] Cortez, L., Magalhães, P. and Happi, J., Principais subprodutos da agroindústria canavieira e sua valorização. Revista Brasileira de Energia. 2, pp. 111-146, 1992.

[11] Ojeda, K., Ávila, O., Suárez, J. and Kafarov, V., Evaluation of technological alternatives for process integration of sugarcane bagasse for sustainable biofuels production-Part 1, Chemical Engineering Research and Design, 89(3), pp. 270-279, 2011. DOI: 10.1016/j.cherd.2010.07.007

[12] Federación Nacional de Biocombustibles de Colombia, Cifras Informativas del Sector Biocombustibles Etanol Anhidro de Caña, [Online]. 2014. Disponible en: http://www.fedebiocombustibles.com/v3/estadistica-mostrar_info-titulo-Alcohol_Carburante_(Etanol).htm

[13] Forrester, J., Industrial Dynamics. Pegasus Communications. Inc. Waltham. ISBN 978-1614275336, 1999, 482 P.

[14] Forrester, J., Urban Dynamics. Pegasus Communications. Inc. Waltham. ISBN 978-0262060264, 1969, 286 P.

[15] Aracil, J. y Gordillo, F. Dinámica de Sistemas. Madrid: Alianza Editorial S.A, 1997.

[16] Callejas, M., Valero, H.A. and Alarcón, A.C., Simulation based on system dynamics for evaluating the quality of transport service in a complex social system. DYNA. 80(180), pp 33-40, 2013.

[17] VENSIM. Simulation software. [Online]. Disponible en: http://vensim.com/free-download/ Julio 2014.

[18] Ramírez, S., Escalante, D.y Pineda, M., Impacto en las decisiones de la cadena de suministros de una empresa de confección de prendas deportiva acuáticas con la utilización de dinámica de sistemas. Revista Ingeniería Industrial, 9(1), pp. 67-85, 2010.

[19] Redondo, J.M., Modelado de mercados de electricidad. Tesis Dr. Facultad de Ingeniería, Universidad Nacional de Colombia, Sede Manizales, Manizales, Colombia, 2013.

[20] Sarache, W., Costa, Y. and Martinez, J., Environmental performance evaluation under a green supply chain approach. DYNA, 82(189), pp. 207-215, 2015. DOI: doi. 10.15446/dyna.v82n189.48550

D. Ibarra-Vega. Received his BSc. in Biotechnological Engineering, and Sp. in Environmental Management. He has a MSc. Degree in Environmental Management and Assessment. He is a PhD student in Engineering at the Universidad Nacional de Colombia in Manizales, Colombia. He is currently the Dean and a Professor at the Universidad Sergio Arboleda on the Environmental Engineering program. His research field of interest is the first, second and third generation of biofuel sustainability, and environmental modeling.
ORCID: 0000-0002-3339-6430

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS

DYNA
http://dyna.medellin.unal.edu.co/

Fine material effect on kaolin suspensions rheology

Moisés Oswaldo Bustamante-Rua [a], Néstor Ricardo Rojas-Reyes [b] & Gali Ronel Quitian-Chila [c]

[a] *Instituto de Minerales CIMEX, Facultad de Minas, Universidad Nacional de Colombia, Medellín, Colombia. mobustam@unal.edu.co.*
[b] *Instituto de Minerales CIMEX, Facultad de Minas, Universidad Nacional de Colombia, Medellín, Colombia. nrrojasr@unal.edu.co*
[c] *Instituto de Minerales CIMEX, Facultad de Minas, Universidad Nacional de Colombia, Medellín, Colombia. grquitianc@unal.edu.co.*

Received: February 2nd, 2015. Received in revised form: April 9th, 2015. Accepted: Mayo 8th, 2015

Abstract
A good rheological characterization can be used as a control parameter within the industrial processing of kaolin. The kaolin used was characterized by SEM, XRD, XRF and particle size. Also it was classified and separated in three sizes of fine material, which was introduced in suspensions with three different size distributions. The analysis was based on a rheological study of the fine particles influence, on the suspension viscosity. The results show that it is possible to modify the viscosity by altering the fines content without changing the solid fraction of the suspension. Suspensions of kaolin with 40% content of fines tend to decrease its viscosity value. Suspensions with quantities of fine greater than 60 %, increase the value of its viscosity. In the research are also presented the proposed mechanisms by which the presence of fine increases or decreases the value of the viscosity of a suspension.

Keywords: Kaolin, Rheology, Fine material, Viscosity.

Efecto del material fino en la reología de suspensiones de caolín

Resumen
Una buena caracterización reológica puede ser usada como parámetro de control dentro del procesamiento industrial del caolín. El caolín de trabajo se caracterizó mediante SEM, DRX, FRX y tamaño de partícula. Tres tamaños de material fino, clasificado y separado, se introdujeron en suspensiones con tres distribuciones de tamaño diferente; el análisis reológico se basó en estudiar la influencia del material fino sobre la viscosidad de la suspensión. Los resultados muestran que es posible la modificación de la viscosidad alterando el contenido de finos sin cambiar la concentración en peso de una suspensión. Suspensiones de caolín con contenidos de finos alrededor de 40% tienden a disminuir su valor de viscosidad. Suspensiones con cantidades de finos mayores a 60%, aumentan el valor de su viscosidad. En la investigación también se presentan los mecanismos propuestos por los cuales la presencia de finos aumenta o disminuye el valor de la viscosidad de una suspensión.

Palabras clave: Caolín, Reología, Material fino, Viscosidad.

1. Introducción

Pastas naturales, industriales y suspensiones son materiales granulares compuestos generalmente por un gran número de elementos como gotas, burbujas, minerales, polímeros o partículas de diferente clase, tamaños y formas, todas sumergidas en un medio líquido o gas. Una forma de estudiar el comportamiento de estas pastas y suspensiones es mediante la caracterización reológica, donde se analiza la deformación y el flujo, además de lo concerniente a la interacción entre esfuerzo de cizalla, tasa de cizalladura y tiempo de prueba [1]. Se esperaría que con esta gama de factores y características de los materiales, existieran muchos comportamientos reológicos, sin embargo, estos sistemas compuestos exhiben características de flujo dentro de ciertos rangos, debido principalmente a la presencia de cualidades similares entre las interacciones que ocurren entre el medio continuo y el medio discreto [2]. El comportamiento reológico de suspensiones minerales indica el nivel de [interacción inter-partícula o agregación, y por lo tanto éste puede ser usado como parámetro de control dentro de su procesamiento industrial [3]. La caolinita [$Al_4Si_4O_{10}(OH)_8$], corresponde estructuralmente a un mineral de la arcilla del tipo 1:1, es decir, formado por una capa de tetraedros de sílice

unida a otra capa de octaedros de alúmina a través de grupos hidroxilo que están compartidos entre las dos capas [4]. Las arcillas caoliníticas son valoradas comercialmente debido al uso directo en la industria, mientras que los caolines son procesados previamente para incrementar las propiedades deseadas que residen en las partículas [5]. El caolín, posee múltiples aplicaciones industriales, las más conocidas son: como recubrimiento y relleno para papel, imprenta, lubricación, adhesión, industria cerámica, pinturas, plásticos, refractarios, cementos, muy importante en la industria de cosméticos y farmacéutica, construcción, en cauchos y hules, química y forraje [6]. Esta variedad de usos derivan de sus propiedades fisicoquímicas entre las que destacan: blancura, poca reactividad ante agentes químicos, no toxicidad, poder cubriente, resistencia mecánica después de calcinación y elevada refractariedad [7]. Los procesos productivos para la fabricación de productos cerámicos por vía húmeda necesitan la preparación de suspensiones acuosas de arcillas con elevado contenido de sólidos y baja viscosidad [8].

Johnson et al. [9] señalan que la relación entre la reología y la química superficial en suspensiones de caolín es complicada, principalmente por la presencia de caras y bordes heterogéneamente cargados en cada partícula además de la forma laminar en que se presentan todas las partículas. En las pruebas llevadas a cabo, ellos reportan un aumento en el esfuerzo de fluencia en relación directa con la fracción volumétrica; En el modelo experimental reológico, determinan que es necesario una cuantificación de las propiedades estructurales de la suspensión de caolín, con respecto a tamaño de partícula, forma, número de coordinación y orientación neta. Omland et al. [10] analizaron la gran influencia sobre la viscosidad al agregar partículas sólidas (arcillas y arenas) a un fluido estándar. Las partículas adicionadas presentaron un tamaño promedio de 63 a 75 μm con una distribución de tamaño normal. Los autores cuantificaron el efecto en un diagrama ternario partículas finas - medianas - gruesas con una distribución de tamaño constante para todos los tamaños. El efecto de las partículas adicionadas sobre la viscosidad, aunque dependía de la cantidad, era más considerable con las partículas más finas y mayor aún si las partículas eran arcillosas. Genovese [11] escribió una revisión histórica de la reología de sistemas dispersos, analizando entre otros puntos, el efecto de la distribución de tamaño de partícula sobre la viscosidad. El autor determinó que las partículas finas ayudan a aumentar la fracción de empaquetamiento máximo, pues durante el flujo se colocan en los espacios que dejan las partículas grandes, y en el mismo sentido, esta localización preferencial ayuda a la disminución de la viscosidad pues debido a su tamaño pequeño hacen el papel de lubricante. Olhero y Ferreira [12] analizaron el tamaño de partícula, la distribución de tamaño y el tiempo de molienda de suspensiones de sílice y su efecto sobre su reología. Dentro del estudio, que se llevó a cabo con partículas gruesas (entre 1 y 100 μm) y partículas finas (entre 1 y 10 μm) el autor determinó que para partículas gruesas, el comportamiento reológico fue cizallo-adelgazante y para partículas finas, su comportamiento reológico cambió a cizallo-espesante. Luckham y Ukeje [13] trabajaron con partículas de poli-estireno (diámetros promedio de 400 nm), tres polidispersidades (0,085; 0,3 y 0,485), y fracciones volumétricas entre 25 y 60%. Ellos determinaron que la viscosidad de la suspensión aumenta con el incremento en la fracción volumétrica con una relación exponencial y que los valores más altos de esfuerzo de fluencia se encontraron con las partículas menos poli-dispersas (0,085) para las mismas fracciones volumétricas. Un estudio hecho por Geldart [14] demuestra que suspensiones de partículas esféricas con un porcentaje óptimo (hallado experimentalmente) de los componentes de la suspensión y con finos de tamaño 22% del tamaño de los gruesos, producirá una reducción en la viscosidad.

El objetivo del presente estudio es determinar la relación entre el tamaño de partícula, el porcentaje de finos agregados y la concentración de sólido sobre el comportamiento la viscosidad aparente de las suspensiones de caolín fabricadas.

2. Experimental

El material usado en el presente trabajo proviene de una mina de caolín ubicada en el departamento de Antioquia – Colombia. Las partículas de caolín se clasificaron y separaron en diferentes tamaños, con las cuales se fabricaron suspensiones a diferentes fracciones de sólidos y diferentes porcentajes de fino, para posteriormente realizar la caracterización reológica de cada suspensión.

Se separaron seis tipos de tamaño de partícula, tres tamaños gruesos denominados: +20 μm, +30 μm y +40 μm y tres de tamaño finos denominados: –20 μm, –30 μm y –40 μm. Los tamaños gruesos se obtuvieron por un proceso de tamizado normal y los tamaños finos por un proceso de clasificación basado en la velocidad de sedimentación [15]. Cada material grueso se mezcló con su material fino [16], (ejemplo: +20 μm con –20 μm), y los porcentajes agregados de fino fueron 0, 5, 10, 20, 40, 80 y 100%, con los cuales se fabricaron las suspensiones para obtener 3 concentraciones en peso (*Cp*) que fueron 0,2; 0,3 y 0,4. Los valores de *Cp* elegidos para el estudio son valores similares a los que se trabajan en la industria del procesamiento de minerales arcillosos [17]. Las suspensiones se fabricaron con agua desionizada y microfiltrada, con aplicación de ultrasonido para disgregar las partículas y una agitación mínima de cinco minutos para homogenizar la mezcla preparada, luego cada suspensión fue caracterizada reométricamente.

Para analizar la morfología del mineral que compone el caolín se utilizó un microscopio electrónico de barrido marca Phenom Pro X. El análisis por difracción de rayos X (DRX) se realizó en un equipo Panalytical Xpert-Pro con una lámpara de cobre de longitud de onda 1,548 Å, a una velocidad de barrido de 2°/min, utilizando el método de polvo de Debye-Scherrer. Para la caracterización del tamaño de partícula se utilizó un equipo Mastersizer 2000 de la casa Malvern Instruments.

Dentro del procesamiento normal de un material cerámico, las suspensiones de caolín están sometidas a una serie de etapas tales como mezclado, agitación y bombeo, por lo que a menudo es importante combinar adecuadamente las propiedades reológicas del sistema en valores muy dispares de velocidad de cizalla [18], por esta razón se seleccionó un barrido de la tasa de cizalladura, para cubrir diferentes escenarios de flujo. La evaluación reológica se realizó en un

reómetro rotacional Bohlin Instruments ***C-VOR*** 200, con la geometría plato-plato de 40 mm de diámetro, fabricado en titanio, con un gap de 1000 µm, en la celda Peltier a temperatura controlada de 25°C (± 0,1) y a una tasa de cizalladura desde 0,1 a 1000 1/s en 300 s [19, 20].

3. Resultados y discusión

3.1. Caracterización morfológica

Los diferentes análisis hechos a las partículas de caolín en estudio mostraron que es un material para uso industrial, lo que significa que no es esférico, ni coloidal, ni posee un monotamaño. La Fig.1a muestra una vista general del caolín como se recibe de la mina, en donde se observa la gran heterogeneidad tanto en el tamaño como en forma de las partículas de mineral. En las Figs. 1b y 1c, se observa el detalle de la zona que se señala en la Fig.1a, estas figuras muestran la estructura típica laminar que presenta el caolín.En la Fig.1d se muestran las partículas luego de ser mezcladas con agua, de aplicarle ultrasonido y secarlas, donde se observa la disgregación de las partículas de caolín en forma de láminas y además, que estas partículas poseen áreas basales irregulares.

El análisis de las anteriores figuras demuestra que las partículas gruesas se componen de partículas laminares ordenadas en forma de "libros", las cuales al tener contacto con el agua, y con ayuda del ultrasonido, se disgregarán para formar suspensiones de partículas individuales de forma laminar, como es normal en materiales arcillosos. Se puede suponer que la reología de estas suspensiones tenderá a cambiar en función de la tasa de cizalladura, desde un sistema desordenado con alta viscosidad, hasta un sistema ordenado donde todas las partículas tengan una dirección preferencial de flujo.

3.2. Caracterización mineralógica

La Fig.2 muestra los 3 difractogramas para el material de partida (– 40 micras), para el material de tamaño intermedio (– 30 micras), para el material con tamaño más pequeño (– 20 micras). Un análisis por FRX reportó 89% de caolinita, 8% de moscovita y 3% de sílice para el material de partida (– 40 micras). A medida que el tamaño de partícula disminuye van desapareciendo los picos representativos de la moscovita y de la sílice, de tal forma que el proceso de tamizado puede ser utilizado también como un tipo de concentración física de caolín.

3.3. Caracterización granulométrica.

En la Fig.3 se observa la distribución del tamaño de partícula (DTP) para el caolín de partida (proveniente de mina) y para los 6 tipos de tamaño clasificado y separado. El caolín de partida posee una distribución 80% menor a 42 µm, 70% menor a 30 µm y 50% menor a 18 µm. Los tamaños de partícula denominados + 40 µm, + 30 µm y + 20 µm reportan un d(10) = 40, 30 y 20 µm respectivamente, mientras que los tamaños – 40 µm, –30 µm y –20 µm reportan un d(90) = 40, 30 y 20 µm respectivamente, lo cual demuestra el grado de confiabilidad en la separación de cada fracción de tamaño.

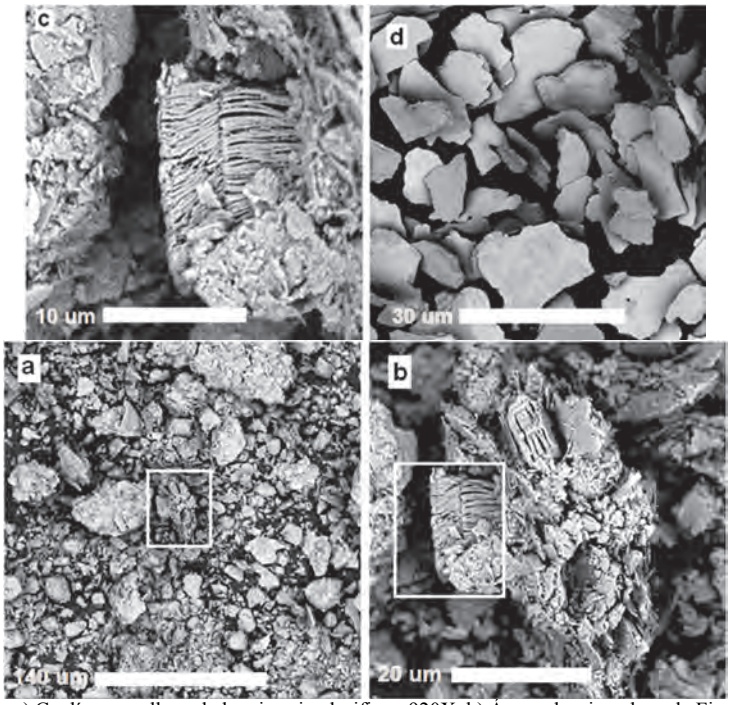


Figura 1. Micrografías por SEM de: a) Caolín como llega de la mina sin clasificar, 920X. b) Área seleccionada en la Fig.1a. 5200X c) Área seleccionada en la Fig.1b, 11000X. d) Caolín luego de ser mezclado con agua, sonicado, probado y secado, 4000X
Fuente: Los autores.

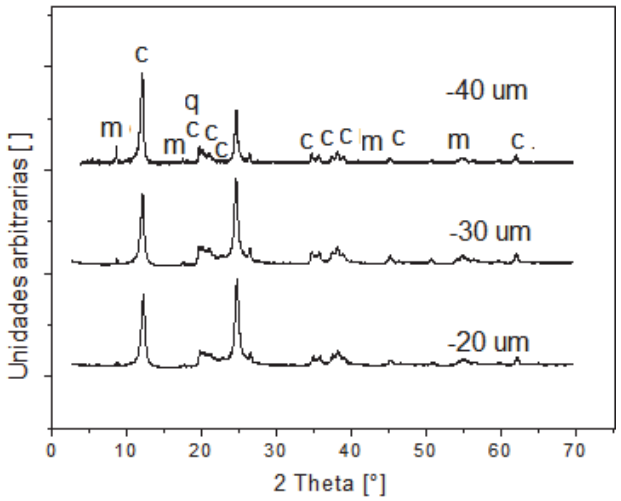


Figura 2. Difractogramas para los tres tipos de tamaños analizados; caolín (c), moscovita (m) y cuarzo (q).
Fuente: Los autores.

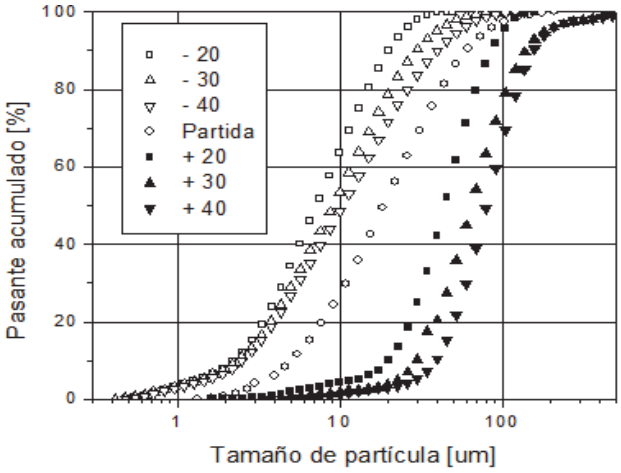


Figura 3. DTP para el caolín de partida y los 6 tamaños clasificados. Fuente: Los autores.

3.4. Caracterización reológica

En la Fig.4 se presentan 9 superficies de respuesta, donde se analiza la tasa de cizalladura en el eje *x*, el porcentaje de finos en el eje *y*, y la viscosidad en el eje *z*, que son el resumen de 54 reogramas realizados. La primera fila (Figs. 4 a, b y c) muestra los resultados del material más fino (+/– 20 µm) a las tres concentraciones en peso analizadas (0,1; 0,2 y 0,3), la segunda fila (Figs. 4 d, e y f) para el material intermedio (+/– 30 µm) y la tercera fila (Figs. 4 g, h e i) para el material más grueso estudiado (+/– 40 µm).

- Análisis de las superficies de respuesta (Fig. 4): a) Influencia del contenido de finos sobre la viscosidad: en general se observa que los mayores valores de viscosidad se presentan para suspensiones con altos contenidos de finos, superiores al 60%. Las suspensiones con contenidos de finos entre 20 y 60% entregan valores de viscosidad relativamente más bajos, y para suspensiones con contenidos de finos menores a 20% se observan valores de viscosidad intermedios. Rangos intermedios de porcentaje de finos disminuyen la viscosidad, tal vez debido a una organización en grupos o paquetes, donde las partículas más pequeñas se colocan entre las medianas y grandes y todas fluyen como familias en una dirección y con velocidad similar. b) Influencia de la tasa de cizalladura sobre el comportamiento reológico: todas las suspensiones mostraron un comportamiento cizallo-adelgazante, lo cual es típico en suspensiones arcillosas, sin embargo, este comportamiento cizallo-adelgazante es más evidente en suspensiones con altos contenidos de finos (> a 60%), y menos evidente en suspensiones con contenidos de finos entre 20 y 60%, lo cual es un fenómeno que coincide en los mismos rangos con respecto a la influencia del contenido de finos sobre la viscosidad.
- Análisis entre gráficas (Fig. 4): a) Influencia del contenido de sólidos (*Cp*) sobre la viscosidad: se observa que el aumento en el valor de viscosidad de cada suspensión es proporcional con el aumento del *Cp* e independiente del tamaño de partícula, lo cual, en este caso, puede ser debido al corto intervalo entre los tamaños, pues éstos varían desde 20 µm hasta 40 µm. b) Influencia del contenido de sólidos (*Cp*) sobre la estabilidad de las gráficas: suspensiones con un *Cp* = 0,1 muestran inestabilidad en las curvas de flujo. Lo anterior se comprobó al evaluar la muestra a una sola tasa de cizalladura, la cual entregó el mismo comportamiento inestable en la curva de flujo. Las pequeñas ondulaciones de estas curvas desaparecen con el aumento en el *Cp*, posiblemente porque aumentando la cantidad de partículas disminuyen los fenómenos de desequilibrio entre la fase continua y la fase discreta, como por ejemplo la sedimentación. c) Influencia del tamaño de partícula sobre la viscosidad: es de notar que a medida que aumenta el tamaño de partícula disminuyen los valores promedio de viscosidad. Si se tiene en cuenta que para el mismo peso de sólidos existe mayor cantidad de partículas pequeñas en comparación con las grandes, entonces es posible que la mayor viscosidad se vea reflejada por el mayor gasto energético requerido para mover éstas partículas pequeñas.

Aunque ya se observó que el valor de la viscosidad aumenta con el valor de *Cp*, la Fig.5 muestra una proporción de aumento logarítmica, efecto posiblemente debido a 3 elementos: a) reducción de la distancia entre las partículas a introducir nuevas partículas; b) aumento de las interacciones inter-partículas y c) mayor demanda de energía para mover mayor cantidad de material. También se observa que la estabilidad reológica de las curvas (fenómenos de desequilibrio entre la fase continua y la fase discreta) mejora a medida que aumenta el valor de *Cp*.

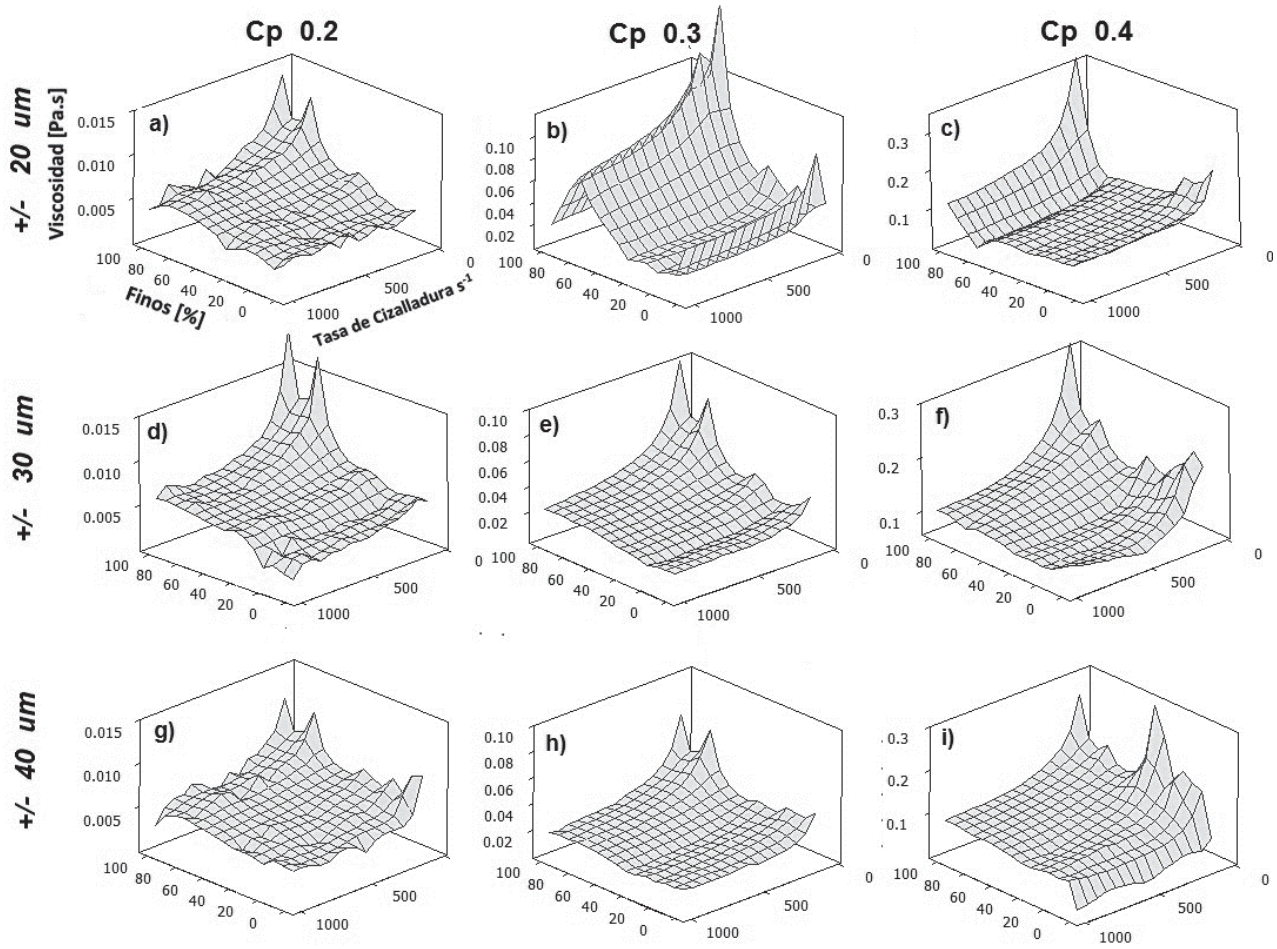


Figura 4. Superficies de flujo para cada una de las suspensiones fabricadas.
Fuente: Los autores.

La Fig. 6 presenta el comportamiento general de la viscosidad de las suspensiones de caolín a una tasa de cizalla constante de 500 1/s, el cual es el valor medio del barrido que se analizó. En general para todas las gráficas se observa una disminución del valor de viscosidad a medida que aumenta el porcentaje de finos, hasta un valor alrededor de 40% de finos. En la literatura se reporta que la adición de cantidades relativamente pequeñas de finos reducirá la viscosidad de la suspensión sustancialmente [21], lo anterior se observa solamente para las suspensiones con tamaño de +/– 20 µm. Para los otros 2 tamaños donde no se cumple lo anterior, una posible explicación es que la fricción entre las partículas es la principal causa de la variación de la viscosidad de una suspensión con altas fracciones de sólidos; la adición de material fino reducirá la fricción entre las partículas gruesas por medio de mecanismos de lubricación, donde las partículas finas recubren cada partícula gruesa [22]. La cantidad de finos depende de la relación de diámetros característicos, ya que al presentar diámetros grandes, como sucede en las suspensiones +/– 40 µm, se tendrá un área superficial mayor que se deberá recubrir con partículas pequeñas, lo cual cambia para partículas +/– 20 µm, pues el área para recubrir será menor, haciendo entonces que varíen los porcentajes necesarios para reducción de viscosidad por agregación de finos para un tamaño u otro.

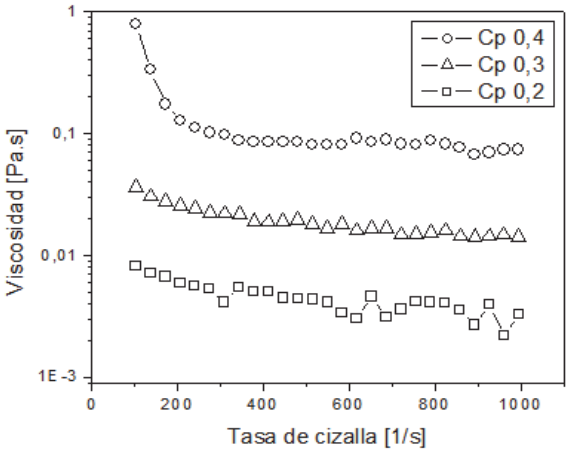


Figura 5. Curva de viscosidad en función de la tasa de cizalla para caolín + 40 µm con 20% de finos – 40 µm, para los tres *Cp* analizados.
Fuente: Los autores.

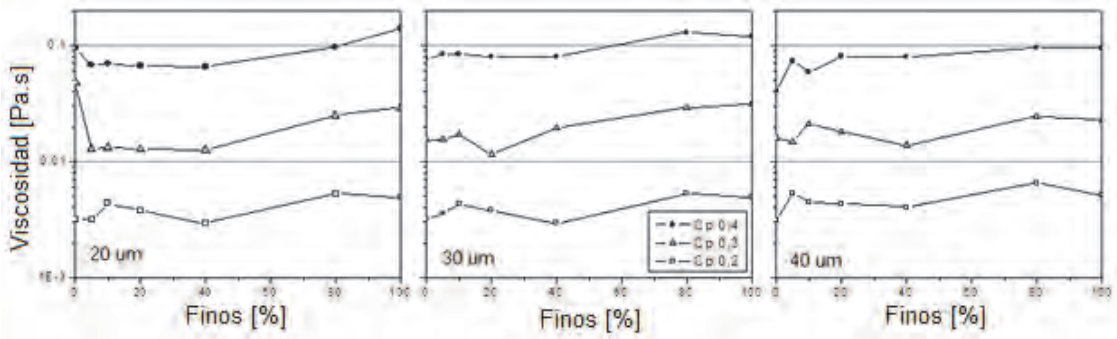


Figura 6. Viscosidad en función de los finos agregados para *Cp* de 0,2; 0,3 y 0,4; y tamaños +/– 20, 30 y 40 µm, a una tasa de cizalla constante de 500 1/s.
Fuente: Los autores.

Otro mecanismo de variación de la viscosidad será la disminución de área de contacto entre partículas gruesas, ya que al agregar un porcentaje de finos adecuado, éstos cumplirán la función de mantener cada partícula gruesa separada de las demás generando así que sólo interactúen por la pequeña área superficial de los finos, con base en lo anterior se pueden observar disminuciones de viscosidad en las suspensiones nuevamente en determinados rangos. El aumento inicial de la viscosidad en las suspensiones de caolín con 5 a 10% de finos se podría explicar indicando que para éstos porcentajes de finos no existe cantidad suficiente para formar completamente un recubrimiento sobre las partículas gruesas y que por el contrario los finos actuarían como "trabas mecánicas" entre las partículas de mayor tamaño, produciendo dificultad para su flujo. Otra posible causa es que cierta cantidad de finos agregados comenzarán a ocupar los intersticios entre las partículas grandes, es decir, el conjunto de partículas gruesas y finas pueden fluir como paquetes ordenados [23], disminuyendo el esfuerzo necesario para su flujo y por lo tanto la viscosidad que presenta la suspensión. Este empaquetamiento no se puede producir ni con 0% ni con 100 % de finos. Esta última teoría se ve respaldada por la Fig.7, donde se puede observar que las suspensiones tendientes a distribuciones mono-modales (5% finos) poseen menor viscosidad que las suspensiones con distribuciones poli-modales (80% finos).

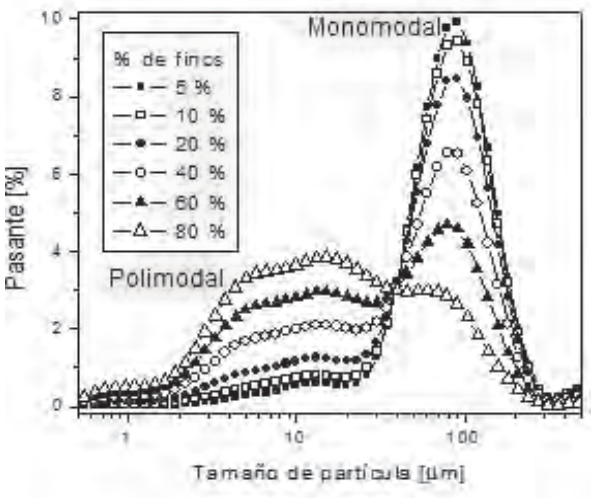


Figura 7. DTP para las mezclas de material grueso +40 µm con finos –40 µm.
Fuente: Los autores.

4. Conclusiones

Los resultados presentados y discutidos en este estudio muestran que es posible la modificación de la viscosidad alterando el contenido de finos sin cambiar la concentración en peso de la suspensión. Suspensiones de caolín con contenidos de finos alrededor de 40% tienden a disminuir su viscosidad. Los posibles mecanismos para la disminución de la viscosidad son los siguientes: a) las partículas finas recubren cada partícula gruesa lo que permitirá una disminución de la fricción entre éstas. b) disminuye el área de contacto entre partículas gruesas, ya que al agregar un porcentaje de finos adecuado, éstos cumplirán la función de mantener cada partícula gruesa separada de las demás. c) el conjunto de partículas gruesas y finas pueden fluir como paquetes ordenados, disminuyendo la energía necesaria para su flujo y por lo tanto la viscosidad que presenta la suspensión.

La viscosidad de una suspensión aumenta en una proporción logarítmica al aumentar el contenido de sólidos de la suspensión, lo cual es un comportamiento esperado, sin embargo, para suspensiones con contenidos de sólidos mayores a 60%, también aumenta el valor de su viscosidad sin importar el contenido de sólidos. La posible explicación de este aumento es debido a que los espacios disponibles entre las partículas gruesas se sobresaturan de partículas finas entregando como resultado: a) reducción de la distancia entre las partículas a introducir nuevas partículas; b) aumento de las interacciones inter-partículas y c) mayor demanda de energía para mover mayor cantidad de material.

Agradecimientos

Los autores expresan sus agradecimientos a COLCIENCIAS por la financiación del proyecto con código 1118-479-22017, a la Universidad Nacional sede Medellín y a la empresa CECOLTEC S.A.S por el apoyo recibido.

Referencias

[1] Macosko, C., Rheology. Principles, measurements and applications. Wiley-VCH. United States of America. 1994.

[2] Coussot, P., Rheometria of pastes, suspensions, and granular materials. A John Wiley & Sons, United States of America. 2005.

[3] Farrokhpay, S., The importance of rheology in mineral flotation: A review. Minerals Engineering, (36-38), pp. 272-278. doi:10.1016/j.mineng.2012.05.009, 2012.

[4] Bartolomè, J.F., El Caolín: Composición, estructura, génesis y aplicaciones. Boletin Sociedad Española de Cerámica y Vidrio, 36(20). 1997.

[5] Torres, J., Gutierrez, R., Castello, R. y Viscayno, C., Análisis comparativo de caolines de diferentes fuentes para la producción de metacaolín. Rev. LatinAm. Metal. Mater., 31, pp. 35-43, 2011.

[6] Teh, E., Leong, Y., Lui, Y., Fourie, A. and Fahey, M., Differences in the rheology and surface chemistry of kaolin clay slurries: The source of the variations. Chem. Eng. Sci. 64, pp. 3817-3825, 2009.

[7] Murray, H., Traditional and new applications for kaolin, smectite and palygorskite: A general everview. Applied Clay Science, 17, pp 207-221, 2000.

[8] Acevedo, P., Carreño, A., Macias, J. y Pedraza. Preparación de suspensiones acuosas coloidales dispersas de arcillas montmorilloníticas. Rev. LatinAm. Metal. Mater. S1 (2), pp. 859-868. 2009.

[9] Stephen, B., Johnson, A. and Russell, P., Volume fraction effects in shear rheology and electroacoustic studies of concentrated alumina and kaolin suspensions. Colloids and Surfaces A: Physicochemical and Engineering Aspects 141, pp. 119-130. 1998.

[10] Tor, H., Arild, S, Svanes, and Amundsen, P., The influence of particle type and size distribution on viscosity in a non-Newtonian drilling fluid. Annual Transactions of the Nordic Rheology Society, 13, pp. 107-110, 2005.

[11] Genovese, D., Shear rheology of hard-spheres, dispersed, and aggregated suspensions, and filler-matrix composites. Advances in Colloid and Interface Science, (171-172), pp. 1-16, 2012.

[12] Olhero, S.M. and Ferreira, J.M., Influence of particle size distribution on rheology and particle packing of silica-based suspensions. Powder Technology 139, pp. 69-7, 2004.

[13] Luckham, P. and Ukeje, M., Effect of particle size distribution on the rheology of dispersed systems. Journal of Colloid and Interface Science 220, pp. 347-356, 1999.

[14] Geldart, D., The effect of particle size and size distribution on the behavior of gas-fluidized beds. Powder Technol., 6, pp. 201-205, 1972.

[15] Webb, P. and Orr, C., Analytical methods in fine particle technology. micromeritics instruments corporation, Norcross, GAUSA, 1997.

[16] Rojas, N., Garrido, P. y Bustamante, O., Metodología para el diseño del transporte hidráulico de suspensiones de relaves de cobre. VI Congreso Internacional de Materiales. Bogotá – Colombia, Noviembre de 2011.

[17] Redd, J., Principes of ceramics processing. John Wiley & Sons. Second Edition. Canada, 1995.

[18] Murray, H.H., Applied clay mineralogy. Occurrences, processing and application of kaolins, bentonites, palygorskite-sepiolite, and common clays. Developments in Clay Science, 2. Elsevier. First Edition. 2007.

[19] Gebhard, S., A practical approach to rheology and rheometry. Second Edition. Thermo Haake Rheology. Gebrueder HAAKE GmbH, Karlsruhe, Federal Republic of Germany, 1998.

[20] Rojas, N, Modelación de la ecuación constitutiva de suspensiones de caolín en función de la energía libre superficial del mineral. Tesis Dr., Facultad de Minas, Universidad Nacional de Colombia, Medellín, Colombia, 2013.

[21] Matheson, G.L., Herbst, W.A. and Holt, P.H., Ind. Eng. Chem. Characteristics of fluid–solid systems, 41, pp. 1099-1104, 1949.

[22] Murray, H. and Kogel, J., Engineering clay products for the paper industry. Applied Clay Science, 29, pp. 199-206, 2005.

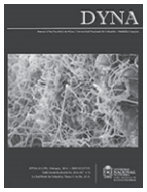
[23] Bustamante, O., Modelación matemática del tensor de esfuerzo y de la viscosidad de una suspensión mineral. Tesis Dr., Universidad de Concepción, Concepción, Chile, 2002.

N.R, Rojas-Reyes, es Ing. en Metalurgia en 1999 de la Universidad Pedagógica y Tecnológica de Colombia, Tunja, Colombia, MSc. en Ingeniería en 2003 en la Universidad de Concepción, Concepción, Chile, y Dr. en Ingeniería en 2014 en la Universidad Nacional de Colombia, Medellín, Colombia. Entre 2004 y 2006 trabajó en la Universidad Pedagógica y Tecnológica de Colombia en el área de la metalurgia extractiva y modelado de procesos y SICE 2007 para la Universidad Nacional de Colombia. Actualmente, es profesor titular de la Facultad de Minas, de la Universidad Nacional de Colombia. Sus intereses de investigación incluyen: la metalurgia extractiva, reología de minerales y físico-química de los minerales.
ORCID: 0000-0002-1644-471X

M.O. Bustamante-Rúa, es Ingeniero de Minas y Metalurgia de la Universidad Nacional de Colombia, MSc. en Ciencias de la Ingeniería con Mención en Metalurgia Extractiva - Mención - Mineralurgia de la Universidad de Concepción, Chile, Dr. en Ciencias de la Ingeniería con Mención en Metalurgia Extractiva – Minería, de la Universidad de Concepción, Chile. Es profesor titular de la Universidad Nacional de Colombia, director e investigador del Instituto de Minerales CIMEX, de la Facultad de Minas, Universidad Nacional de Colombia, sede Medellín, Colombia, ha dirigido e integrado proyectos de investigación en las áreas de concentración y beneficio de minerales, modelación y optimización matemática de procesos y operaciones mineras, reología de suspensiones y pastas, ha publicado artículos en revistas científicas y participado como expositor en congresos y simposios nacionales e internacionales. Además ha sido director de diversas tesis de Maestría y Doctorado.
ORCID: 0000-0002-1692-991X

G.R Quitian-Chila, es Ing. Mecánico y estudiante de maestría en ingeniería de materiales y procesos de la Univerisidad Nacional de Colombia, investigador del Instituto de Minerales CIMEX-Facultad de Minas, integrante del grupo de investigación de micro y nanoanálisis-CEMMI-ASOCM, labora en la industria como soporte científico en laboratorios de I+D de diferentes industrias. Sus investigaciones las ha desarrollado en áreas de modelamiento por método de elementos finitos, reología de suspensiones y pastas, elaboración de nanofibras y nanoencapsulado por técnicas de electrospinning y electrospray. Sus intereses incluyen: Nanotecnologia y materiales inteligentes (Smart materials)
ORCID: 0000-0001-6833-4777

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS



DYNA
http://dyna.medellin.unal.edu.co/

Numerical simulation to assess the elastic-strain energy distribution in a silicon rubber disk subjected to a punch shear test (PST)

Adrián Lopera-Valle, Fabio A. Suárez-Bustamante & Juan P. Hernández-Ortiz

Universidad Nacional de Colombia, Sede Medellín. Medellín, Colombia, adloperav@unal.edu.co, fasuarez@unal.edu.co, jphernandezo@unal.edu.co

Received: February 10th, 2015. Received in revised form: August 20th, 2015. Accepted: September 14th, 2015

Abstract
Finite element method simulations were implemented to understand how the strain energy is distributed in a disk-like sample during a punch shear test. Material's Young modulus can be estimated from this test; however, there is not enough available information about the distribution of the strain energy inside the sample during the deformation process. The proposed methodology seeks to give insight into the deformation process. Experimental results for a cured silicon rubber sample were used to validate the simulation results. It was found that the estimation of the Young modulus with the punch shear test depends on the ratio between the span-to-punch diameters. This conclusion applies to the simulated results, following Timoshenko's theory for the deformation of thin plates. Understanding how energy is accumulated during a punch shear test is an important and useful characteristic in terms of the design of armor systems.

Keywords: Punch-Shear Test (PST), Finite Element Method (FEM), hyper-elasticity, silicon rubber, strain-energy, ballistic armor design.

Simulación numérica para evaluar la distribución de energía de deformación en un disco de caucho siliconado sometido a un ensayo de corte por punzonado (PST)

Resumen
Simulaciones por el método de elementos finitos fueron empleadas para comprender como se distribuye la energía de deformación al interior de muestras con forma de disco durante un ensayo de punzonado. El módulo de Young puede calcularse a partir de este ensayo; sin embargo, no hay suficiente información disponible sobre la manera en la cual se distribuye la energía de deformación en una muestra durante el proceso de deformación. La metodología propuesta busca dar luces alrededor del proceso de deformación. Resultados experimentales obtenidos con muestras de cauchos siliconados curados fueron usados para validar los resultados de las simulaciones. Se encontró que el valor estimado para del módulo de Young depende de la relación entre el diámetro de la muestra y el diámetro del punzón. Entender como la energía se acumula durante el ensayo de punzonado es un aspecto importante y útil para el diseño de sistemas de protección balística.

Palabras Clave: Ensayo de Corte por Punzonado (PST), Método de Elementos Finitos, híper-elasticidad, caucho siliconado, energía de deformación, diseño de materiales balísticos.

1. Introduction

The Punch Shear Test (PST) has been widely implemented as a material characterization technique for many applications. PST has been useful to assess elastic and plastic properties of materials [1-4], their creep behavior [5-7], ductile-brittle transition [8], fracture resistance [9-10] and wear resistance [11-12]. Information from the PST has been successfully applied in applications such as polymeric biomaterials design, artificial joint replacement [1,2,11-14], manufacturing of steels for turbines [15-17], welding processes [5], nuclear power generation [3,8], and materials design [8], among others. Due to the clear difference between the elastic, plastic and failure regimens, PST has been

commonly performed in polymeric and metallic materials such as PMA [2], UHMWPE [1,11,12,14], AISI-316 [3] and Ti [18]. PST has also served as a technique to compare classic material families and composites to evaluate the effect of reinforcements [19,20-24].

The PST can establish correlations between materials' static and dynamic mechanical response [24-27]. For classic solid mechanics, quasi-static and dynamic mechanic responses have been considered as nonrelated phenomena in which material behavior depends on load rate. However, some recent works have suggested that it is possible to define a correlation between the dynamic behavior and quasi-static mechanical properties through common mechanical tests such as the uniaxial tensile test, compression test, planar tensile test and the PST, among others [21,22,24-27]. Marsavina et al. [27] presented a correlation between static and dynamic fracture toughness of polyurethane rigid foams of different densities. A quadratic relation between the static and dynamic responses of rigid foams was proposed in that work using a static three point bend and a pendulum impact test, respectively. Moreover, K.T. Chau et al. [25] proposed theoretic and experimental models to establish relations between static and dynamic energies required for the fragmentation of plaster spheres of different diameters and strengths. Besides these works, more research groups focused on the study of static-dynamic relationships [24-27].

Proper relations between materials' static and dynamic responses through the PST are valuable in areas of ballistic protection and impact mechanics, and provide new avenues in designing and assessing novel ballistic protection materials. Recently, Quasi-Static (QS) PST has been implemented in order to understand impact and penetration of a bullet into a material (target) [20,21]. The QS-PST simulates the different phases of the dynamic penetration process through the variation of geometrical parameters. In addition, the PST has been used to study the ballistic performance of composite materials such as a carbon fiber reinforced nanotube–epoxy composite [23] and a glass fiber reinforced thermoplastic composite [22]. Similarly, other researchers have applied the PST for ballistic materials design [28,29].

Gama et al. [20,22] have used the PST to characterize the progressive damage and delamination of a glass fiber reinforced thermoplastic composite. They proposed a ballistic penetration model in which the bullet impact and penetration are simulated by varying the Span to Punch Ratio ($SPR = R/b$) [21]; Fig. 1. Their model quantifies the energy at each penetration phase and provides information about the material effect and sample thickness on the ballistic performance of composites. This work shows the feasibility of the PST in terms of the design of armor systems and other impact applications [23,28,29].

A PST assembly consists of a disk-like sample fixed and guided by removable annular components, which allows the modification of sample boundary conditions, Fig. 1. A punch applies a localized load at the center of the disk to deform it. This generates a disk displacement, which mostly occurs at its center. Loads and displacements are measured using electronic sensors, i.e. a load cell and a Linear Variable Differential Transformer (LVDT), respectively. The load-

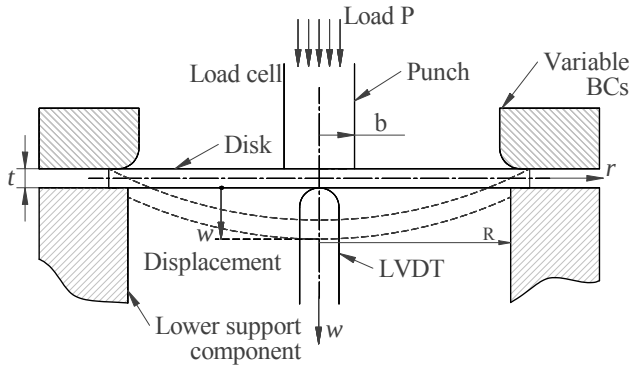


Figure 1. Punch Shear Test (PST) assembly.
Source: The Authors

displacement curve provides information on the mechanical properties of the sample and its material [1-3, 5, 11-17]. Other types of assembly can also be used; previous studies have changed the punch shape (rounded [1,2,11-14], flat [4, 19,20-22] and ring-like [18]), the boundary conditions (clamped [4,14], simply supported [18]), the scale of the test size (from 0.25 mm - microscale [4,18] to around 200 mm – mesoscale [20-22]), and the measurement techniques. These changes provide a wide array of possibilities to characterize materials by PST techniques.

Fig. 2 shows a sketch of a typical PST load-displacement curve for a ductile response material, like a thermoplastic, an elastomeric polymers or a reinforced polymer composite. Initially, the punch faces the sample and while the load rises, a linear elastic response is presented with an initial slope given by the system stiffness. A transition zone, depicted by points between *a* and *b* in Fig. 2 appears where the material releases energy by a yielding mechanism for ductile materials or a matrix cracking mechanism for composites. Linear behavior is again observed after the transition, the points between *b* and *c* in this figure, with a different slope called the Hardening Stiffness. This linear response is due to a hardening stiffness process and the material elastic-yield response. Finally, a failure load is reached and the punch breaks through the lower

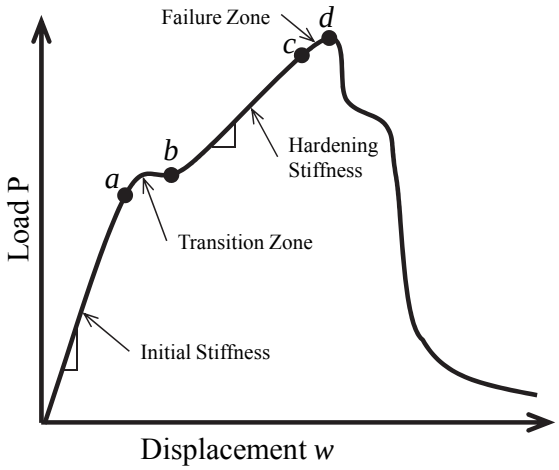


Figure 2. Sketch of a typical QS-PST curve.
Source: The Authors

sample surface, point *d* in Fig. 2. After failure, friction between the punch and sample surfaces is also sensed and plotted. The area under the curve provides important information about stored energy during each identified PST stage [12,14,20-22].

Although the PST has been widely implemented as a valuable tool for the development of new composites and ballistic protection materials, there is not enough information about the distribution of the strain energy inside the sample during the deformation process. Therefore, this work attempts to gain insight into the deformation process through a Finite Elements Method (FEM) simulation. The main idea is to understand how strain energy is accumulated inside a disk-like plate made of a hyper-elastic material (silicone rubber). Mechanical properties are estimated from the FEM simulations and compared with previous PST experimental results by Bing-Feng Ju et al. [4]. Understanding the way in which the strain energy is stored is an important aspect for the design of composite materials. The results provide selection parameters for each constituent of a composite material - matrix and reinforcements -, their configuration - fiber, particles, coating, etc. - and their partial or relative quantities. Some main factors to consider are the way the energy is stored and the specific energy dissipative mechanisms. Results from the current work could be used to develop novel methodologies to design material for body and structural armors.

This manuscript is organized in 4 sections that are described as follows: Section 2 gives a brief overview of the material model considered in the current work. The material parameters for the model were obtained from the results presented by Meunier et al. (2008) [30] to characterize hyper-elastic materials from common mechanical tests. The geometrical parameters and load conditions for the Finite Elements Simulation carried out in the current work are shown in Section 3. They were defined by taking into account previous PST experiments published by Bing-Feng Ju et al. (2005) [4]. These were used as a reference point from which a comparison was made with the results obtained in the current work; also, the geometrical constants and variables, model simplifications and measurement points are illustrated in this section. Section 4 presents a procedure to validate the results presented in current work and the three main results are also shown: the load-displacement behavior of the simulated systems, the effect of the geometry on the accuracy to estimate the Young's Modulus and its effect on the distribution of stored energy inside the material. Finally, Section 5 provides a summary of the main conclusions and questions that will motivate future research.

2. Model and material properties

Silicon rubber was the material selected for the analysis and the validation of the proposed methodology. This type of hyper-elastic material requires an appropriate model to describe its mechanical behavior. Opposite to a linear response of an elastic material, a hyper-elastic material, such as rubber, biological tissue and other vulcanized elastomers [31-35, 45-48], present a non-linear stress-strain behavior under the yield point. Hyper-elastic stress-strain responses can be described by a strain energy function that is represented by several models. Mooney, Haines and Wilson, and Genthave endeavored to describe the strain energy density, U, as a function of the two first invariants of the deformation tensor and other temperature dependent constants [30-37]. Moreover, Ogden's model expresses the strain energy density in terms of the vector of principal extensions, $\lambda = (\lambda_1, \lambda_2, \lambda_3)$, as follows [31, 38, 48-50],

$$U(\lambda_1, \lambda_2, \lambda_3, J) = \sum_{i=1}^{N} \frac{2\mu_i}{\alpha_i}\left(\lambda_1^{\alpha_i} + \lambda_2^{\alpha_i} + \lambda_3^{\alpha_i} - 3\right) + \sum_{i=1}^{N} \frac{1}{D_i}(J-1)^{2i} \quad (1)$$

where $\mu_i[Pa], \alpha_i$ and $D_i[Pa^{-1}]$ are temperature dependent material parameters and J is the elastic volume ratio, which relates the deformed geometry with the non-deformed geometry [31,38,47]. Noting that, for incompressible materials, such as the one considered in current work, D_i parameters are equal to 0 and the second summation term in Equation 1 vanishes in order to avoid an indeterminate value [31,38,48-50].

The material parameters used for current simulations were taken from Meunier et al.'s work (2008) [30]. The authors used the models that were previously mentioned for silicon-rubber. To model that material, they performed simple tensile, plain strain tensile, simple compression and plain strain compression tests. For the current work, those results and Ogden's model with $N = 3$ were used. Past studies showed that this model is the best for hyper-elastic materials, such as silicone rubber [30-32,35,37].

Experimental results [30] were implemented in ABAQUS [38,51], an evaluation material module used to solve $N = 3$ Ogden's Model. Parameters μ_i, α_i and D_i were then calculated and are listed in Table 1. These estimations were obtained for a non-compressible material [30-37] with a Poisson's ratio of $v = 0.5$ [39-41]. These material parameters were used to validate the FEM simulation results. The constants for Ogden's model are related to the Young modulus through the shear modulus, $\mu_T = \sum_{i=1}^{N} \mu_i$ [31], as follows [14],

$$E_{REF} = 2 * \mu_T(1 + v) = 3\mu_T \quad (2)$$

3. Finite Element Method (FEM) simulation

Bing-Feng Ju et al. [4] performed QS-PST experiments and related their results with Timoshenko's low-deformed circular plates theor , in order to characterize the mechanical properties of silicone rubber disks. They proposed a new method for assessing mesoscale disks. An Atomic Force Microscope (AFM) was used to measure the load and displacement in the sample. Their samples were MRTV1 silicone rubber membranes with different radii (R= 1, 1.5, 2, 2.5, 3, 4, 5 mm) and a constant thickness (t= 120 ± 2 μm). A 250 μm rigid cylindrical graphite punch attached to a force transducer was used to apply the compressive load and an

Table 1.
Estimated material parameters for Ogden's model with $N = 3$.

$\mu_1 = 460000\ Pa$	$\mu_2 = 270\ Pa$	$\mu_3 = -7400\ Pa$
$a_1 = 1.4$	$a_2 = 10$	$a_3 = -3.3$
$D_1 = 0\ [Pa^{-1}]$	$D_2 = 0\ [Pa^{-1}]$	$D_3 = 0\ [Pa^{-1}]$

Source: [30]

LVDT actuator sensed the displacement. Using these measurements, they estimated the flexural rigidity $\mathcal{D}$ as a function of the load and displacement using the Timoshenko's solution for symmetric flexion in circular plates, i.e. [4,42],

$$\mathcal{D} = \frac{P}{16\pi w_{max}}\left[\frac{3+v}{1+v}R^2 + b^2 ln\frac{b}{R} - \frac{7+3v}{4(1+v)}b^2\right] \quad (3)$$

where v is the Poisson's ratio, P is the applied load, R is the sample radius, b is the punch radius and $w_{r=o} = w_{max}$ is the axial axis displacement (see Fig. 1). The Young modulus for a sample with thickness t is then calculated as follows,

$$E = 12(1-v^2)\mathcal{D}/t^3 \quad (4)$$

Table 2 summarizes the main results from Bing-Feng Ju et al.'s study [4]. The measured mean elastic modulus and its standard deviation are shown in Fig. 3.

The experimental results from Young's Modulus obtained by Bing-Feng Ju et al. [4] are scattered around a mean value of 1.69 MPa and, apparently, they do not depend on the sample size. The results were close to the mean value for tests carried out using samples with a radius between 1 mm and 3 mm. However, results exhibited significant deviations for samples with a radius between 1.5 and 2.5 mm. This behavior suggests that there is no relation between the sample size and the measured modulus, and that the deviations are associated with practical aspects during the tests [4]. Finally, an important conclusion of this work is the feasibility of characterizing mechanical properties of rubber-like materials by using QS-PST.

Table 2.
Experimental results extracted from

R, mm	P, μN	$w_{max.}$, μm	$\mathcal{D}$, μJ	E_{EXP}, MPa
1.0	203.0	27.42	0.325	1.693
1.5	147.1	42.67	0.352	1.828
2.0	130.3	67.54	0.352	1.834
2.5	84.90	83.22	0.293	1.526
3.0	75.40	98.05	0.319	1.661
4.0	50.40	110.2	0.338	1.661
5.0	32.00	118.2	0.313	1.631

Source: [4].

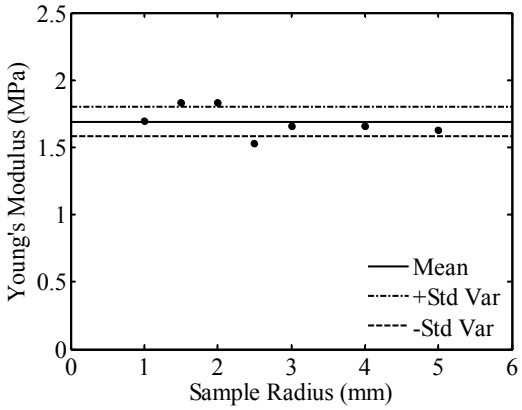


Figure 3. Experimental estimation of Young's modulus reported by Bing-Feng Ju et al. [4] and plotted by the current authors.
Source: The Authors

In the current work, punch tests by Bing-Feng Ju et al. [4] were reproduced using FEM simulations. ABAQUS software [38], version 6.10, was used to represent an axisymmetric loaded circular plate as is illustrated in Fig. 4. Elements were of the CAX8RH type: an axi-symmetrical second order element with 8 nodes [38,51]. The volume partition process involved 400 elements for the smallest disk and 2800 elements for the biggest disk. The selection of these amounts of elements was the result of a convergence analysis of the number of elements in the mesh that were used to define an adequate element size required to achieve convergence in the results. The mesh used in the solution to the problem was refined by increasing the number of elements in a ratio of 1.5 nodes per step until the solutions converged within 3% of variation. The disk size was varied between 0.275 and 5 mm to get several Span to Punch Ratio ($SPR = R/b$ [21]), from 1.1 to 20. The applied load and the punch radius were kept constant at 100 μN and 0.25 mm, respectively. The results of displacement and strain energy were measured at 0.013, 0.05, 0.1, 0.2, 0.4, 0.7 and 1 normalized radius position ($r_n = r/R$) at the top and bottom faces and the middle cross-section of the plate. There is a higher density of measuring points around the center of the disk due to the expected higher variations in the application zone, $0 < r < b$ (Fig. 4). Displacement at the center of the plate and on the applied load was used in Equations (3) and (4) to obtain a FEM-estimated Young Modulus, E_{FEM}.

A flow diagram summarizing the numerical methodology is shown in Fig. 5. The FEM was implemented to simulate the deformation of a disk-shape plate made of a hyperelastic material under QS-PST. The material parameters for the hyper-elasticity model were taken from the experimental work of Meunier et al. [30]. The geometrical and load conditions were defined from the QS-PST experiments reported by Bing-Feng Ju et al. [4]. Thereafter, the simulation results were validated by comparing the E_{EXP} and E_{FEM} values against the E_{REF} and also by comparing their relative errors. Finally, from the FEM simulations, normalized displacement and normalized stored strain energy results were also obtained and analyzed regarding the SPR variable.

4. Results

To validate the numerical methodology and to be able to extract additional information from the FEM simulation, a reference modulus, E_{REF}, calculated using the Ogden's model, was used. From the simulations, an E_{FEM} Young modulus was calculated and compared to the E_{REF} and E_{EXP}

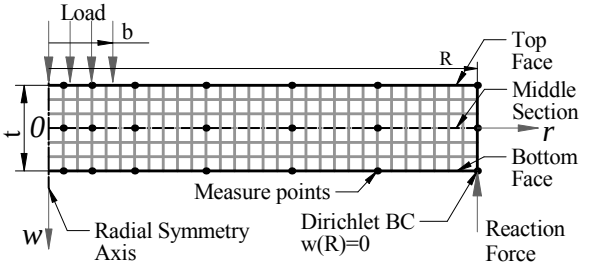


Figure 4. Geometry and typical mesh for FEM simulations.
Source: The Authors

moduli; the E_{EXP} is the experimental modulus reported by Bing-Feng Ju et al. [4]. Once the E_{REF} and the E_{FEM} are equal, displacements and strain-energy results from the FEM simulations are used to build strain-energy curves as function of SPR and r_n.

Table 3 shows the comparison between the moduli from the simulated and measured PST experiments. The E_{FEM} modulus is listed at each simulated condition, while the E_{EXP} modulus is listed for the available data. The reference value of the Young Modulus calculated from Equation (2) was $E_{REF} = 1.36\,MPa$. The relative errors, ε_{FEM} and ε_{EXP}, are also shown in Table 3. These errors are deviations of the FEM and experimental moduli from the reference modulus. These results are also shown in Fig. 6.

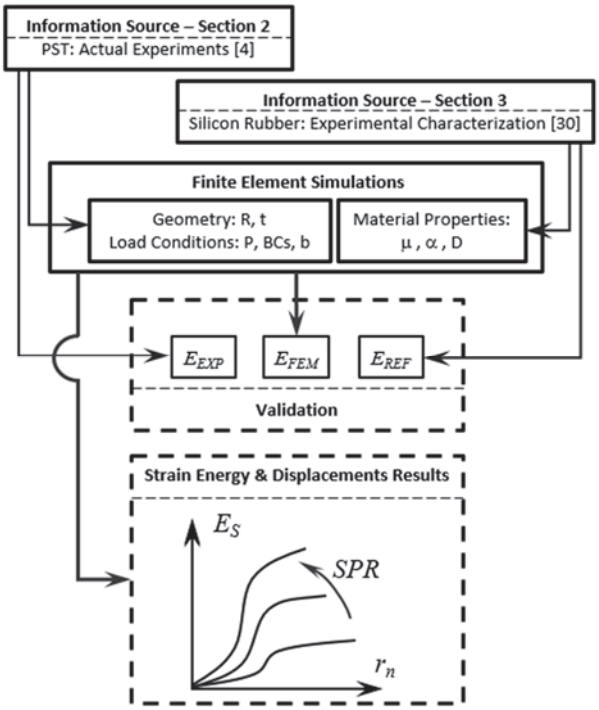


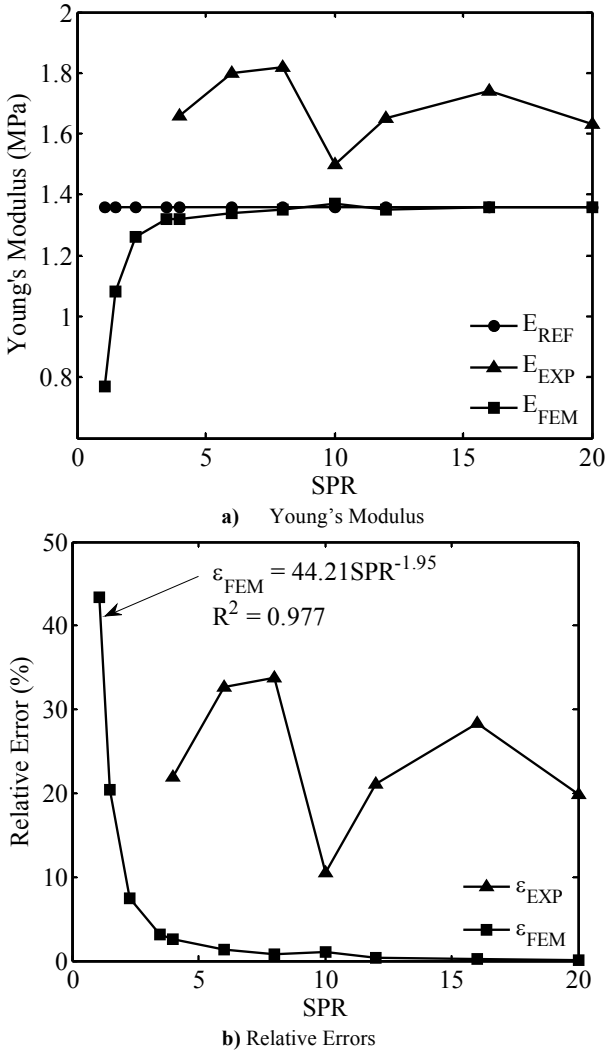
Figure 5. Validation methodology.
Source: The Authors

Table 3.
Young's modulus from PST simulations, by the Authors, and measurements from [4]. The reference modulus is $E_{REF} = 1.36\,MPa$.

R mm	SPR	E_{EXP} MPa	E_{FEM} MPa	ε_{EXP} (%)	ε_{FEM} (%)
0.275	1.1	–	0.77	–	43.4
0.375	1.5	–	1.08	–	20.4
0.575	2.3	–	1.26	–	7.5
0.875	3.5	–	1.32	–	3.1
1.0	4	1.66	1.32	21.9	2.6
1.5	6	1.80	1.34	32.7	1.3
2.0	8	1.82	1.35	33.8	0.7
2.5	10	1.50	1.37	10.5	1.1
3.0	12	1.65	1.35	21.1	0.3
4.0	16	1.74	1.36	28.3	0.2
5.0	20	1.63	1.36	19.9	0.1

Source: The Authors

The relative difference for the FEM modulus ε_{FEM}, decreases monotonically as the disk radius increases, equivalent to an increase in the SPR. Once the SPR is higher than 6, the relative difference is almost zero, lower than 1%. Therefore, SPR>6 upwards should be considered as a critical condition where the Young Modulus estimation starts to be in complete agreement with the reference value. Considering that the disk thickness and the punch diameter were held constant during simulations, bigger disks will behave like bending or stretching plates. In this situation the shear effect tends to be negligible as the SPR increases; thus, reducing this noise provides a better estimation of the elastic modulus. The higher values of ε_{FEM} obtained for samples with a small radius or low SPR values must be related to the strong effect of the shearing. In conclusion, the accuracy of the E_{FEM} strongly depends on the SPR and the sample thickness. From a practical point of view, using an SPR higher than 3.5 provides a relative error that is lower than 5% could be considered to be a good estimation of the material Young modulus.



a) Young's Modulus

b) Relative Errors

Figure 6. Experimental, reference and FEM Young's Modulus with their corresponding relative errors.
Source: The Authors

However, the experimental relative difference, ε_{EXP}, does not follow the same behavior as the ε_{FEM}. For the different evaluated conditions, ε_{EXP} is independent of SPR and it varies between 10% and 35%. A possible explanation is that some experimental issues may have screened this effect. According to the results, E_{FEM} values are similar to E_{REF}, but they are slightly lower than the reference value for all conditions. Conversely, E_{EXP} is always higher than the reference. However, these values for E_{REF}, E_{EXP} and E_{REF} have the same magnitude order and they are in the typical reported range for this these kinds of materials [39-41].

Fig. 7 shows the relation between the normalized displacement (w/t) and the normalized radial position (r/R). Disks with SPR= 12, 16 and 20 present a greater displacement than their thickness; this implies that these disks would not be covered by the theory of plates and shells mechanics that was defined by Timoshenko [42]. However, the results shown in Table 3 indicate that the Young modulus determination of these disks is as accurate as higher $SPRs$, once Timoshenko's plates and shells mechanics theory is applied. Although both disk material and applied load were the same for all simulations, stiffness of the disk decreased as the radius was increased, which lead to a higher deflection in bigger disks. Thus, the geometry of the bent disks changes as the SRP increases. Fig. 8 shows SPR=1.1 and SPR=20 displacement profiles at the plate top (face upon load), middle section and bottom (support disk face). There is a noticeable difference between the scales associated with a disk of 1.1 and 20 SPR. For SPR=1.1, there are significant differences in the correspondent corresponding displacements of the three evaluated surfaces. At the boundary these differences reached the highest values. From a practical point of view, and taking into account the magnitude of the equivalent displacements obtained for 20 SPR disk, these displacements could be assumed to be nearly zero. For SPR=20, the results suggest that normalized displacements are almost constant in the load zone where they reached the maximum value (about 3.5).

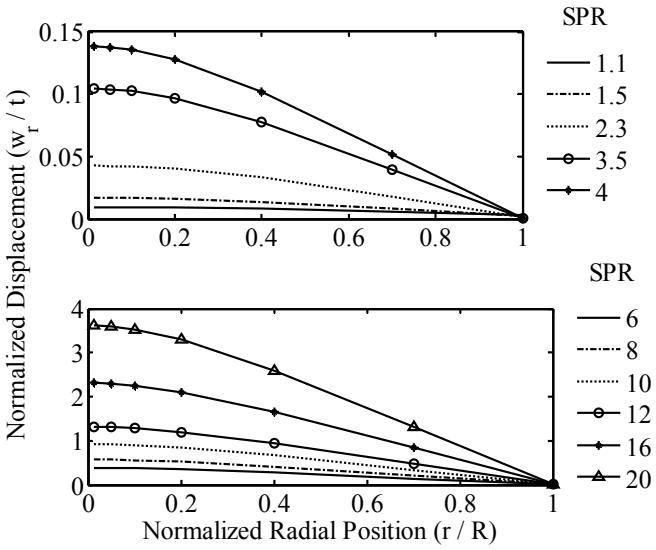


Figure 7. Disk displacement results at the top surface for several SPR. Source: The Authors

Out of this zone, the normalized displacements tend to exhibit linear behavior. Contrary to what was mentioned for the 1.1 SPR disk case, there are no relevant differences for the magnitudes of the correspondent normalized displacements of the three evaluated surfaces. The load conditions (magnitude of the load, size and shape of the loaded area) and the disk geometry (its radius and thickness) modify the behavior of the system to make it similar to membrane behavior where the effect of the thickness is negligible. This happens when SPR is higher than 6 about. In these cases, the load would seem to be applied in a small area (like a point load or a concentrated load at the limit case) and the real size of the disk would be by far larger than in the SPR=1.1case. Therefore, the system could store more energy and the summation of the local material displacements would generate a higher maximum displacement at the center of the disk compared with its equivalent in the 1.1 SPR system, as is shown in Fig. 7.

Fig. 9 shows the Normalized Stored Strain Energy ($NSSE = SSE_r/SSE_R$y) as a function of the normalized radius position (r/R) for different SPR values. It can be observed that when the SPR is higher than 2.3, the energy distribution curves have the same behavior (S-like shape). Additionally, for $SPR > 8$, there are no relevant changes in the NSSE distribution. For lower SPR values ($1 < SPR < 2.3$), curvature and concavity of the curves exhibit significant changes. Curves tend to adopt a J-like shape, and for the values near to 1 they apparently have only one concavity, i.e. the inflection point has disappeared.

The way to store energy in the disk mass (energy distribution curves) strongly depends on SPR, especially for lower values. Fig. 9 clearly shows an example: a 20 SPR system can store a high quantity of strain energy at a low volume ratio. For this system, 70% of the strain energy is accumulated at 25% of its normalized volume, which is equivalent to 50% of its normalized radius. A system with $1.1 \leq SPR \leq 2.3$, on the other hand, requires a higher volume ratio to store the same amount of strain energy. According to Fig. 9, a 1.1 SPR system would need 78.3% of the normalized volume ratio to store 70% of strain energy, meaning 88.5% of its normalized radius. Thus, for the 20 SPR, the strain energy is stored more heterogeneously than for the 1.1 SPR. Shear-Bending transition occurs at SPR values from 1.1 up to 2.3 for the same load conditions and the same thickness.

Finally, Fig. 10 shows the Normalized Stored Strain Energy Density ($NSSED = SSE\rho_r/SSE\rho_R$) as a function of the normalized radial position (r/R) for different SPR values. It provides information about homogeneities of the strain energy density within the disks. For systems with $SPR \geq 3.5$, the stored strain energy is preferentially accumulated in a small material volume around the center of the disk; therefore, the energy density is at its maximum at the disk's center and quickly decreases to its minimum value at the disk's edge. For systems with SPR from 1.1 to 2.3, the strain energy reaches two maximum values. The higher energy concentration is also at the center of the disk. However, due to the fact that the load application area is similar to the area of the entire disk, the energy density slowly decreases from its center to around 0.7 of the its normalized radius. Thus, it

starts growing as far as its edge where it is supported. When SPR decreases near to one, this second maximum value sharply increases, becoming bigger than the first maximum. This is because the load state is reaching a pure shear condition and the bending effect turns insignificant.

The behavior of the curves in Fig. 9 and Fig. 10 is strongly related to the relative size of the load application zone in the sample size. The results suggest that the way in which energy is stored around the boundary of the loaded zone is basically affected by a concentrated shear effect, the action of which is greater when the SPR is near to one and insignificant for SPR>6. In Fig. 9, the slope of the curves (the strain energy gradient) reaches its maximum at the edge of the load zone due to the main role of the shear effect in this area (the deviatory component of the strain tensor must be maximized in this localized region). This explains the fast rise of the strain energy around the edge of the loaded area ($r_n \rightarrow b/R$) for all the evaluated SPR.

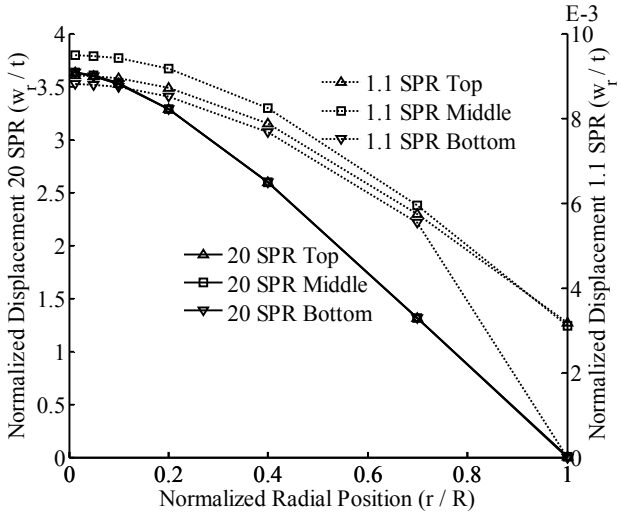


Figure 8. Displacement comparison for 1.1 and 20 SPR considering the three evaluated surfaces in each case.
Source: The Authors

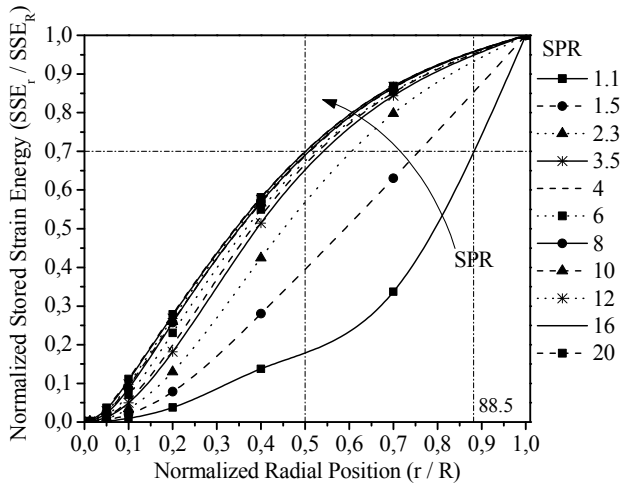


Figure 9. Normalized Stored Strain Energy (NSSE) as function of SPR and r/R.
Source: The Authors

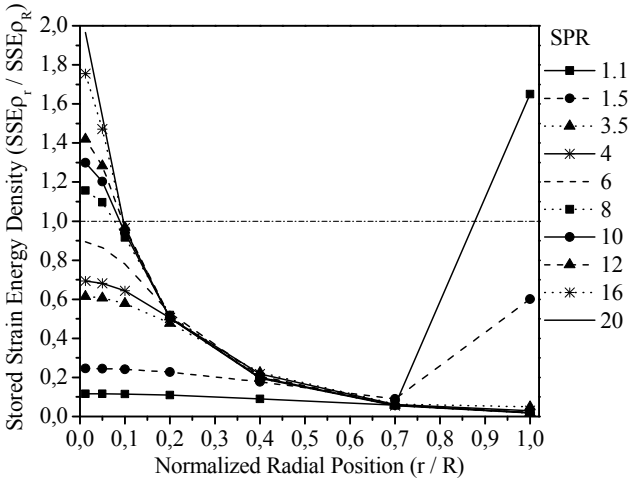


Figure 10. Normalized Stored Strain Energy Density (NSSED) as function of SPR and r/R.
Source: The Authors

The curves of the density of strain energy showed in Fig. 10 indicate that for those conditions in which shear stresses have a more significant effect on the system behavior (low SPR values), the material points that reach higher energy levels are between the load zone edge and the sample boundary. In this region the deviatory stresses are higher than central portion of the disk, and the effective volume that stores the energy is small; therefore, the stored strain energy density increases quickly. In this case, it is important to note that the average stored strain energy per unit of volume of the disk (i.e. the $SSE\rho_R$) varies with the SPR ratio for all the systems analyzed; however, when SPR is increased, it behaves asymptotically to reach a maximum value at infinite. So, when $NSSED = SSE\rho_r/SSE\rho_R$ is calculated for low values of SPR, its value would tend to be greater. However, due to the better load distribution in this case, the local strain energy is lower than in the case of high SPR values (where the load acts concentrated) and its value is lower inside the load. In the region between the punch edge and the sample boundary, its value is greater than in the last region because of the strong shear effect. Finally, in the case of systems with high SPR values, the load is concentrated in a small area and the stresses in that zone and its neighborhood are higher than at the sample boundary, as is reported in Fig. 10.

5. Conclusions

FEM simulations to model the Punch Shear Test (PST) using a hyper-elastic material (silicon rubber) have been reported. Material parameters were taken from a previous work undertaken by Meunier et al. (2008) [30] to estimate a reference Young Modulus. Geometrical and load conditions were obtained from the punch shear experiments carried out by Bing-Feng Ju et al [4].

The estimation of the Young Modulus by Punch Shear Tests (PSTs) depends on the SPR in a plate with a constant thickness. For SPR values between 1.1 and 2.3, this dependence is strong; while for SPR values greater than 6 or 8, it is weak. The predicted Young Modulus that was

determined from simulations based on the PST tends to the reference value, reaching relative errors lower than 1% at higher SPR values. A potential decay of the relative error of the predicted Young Modulus was found, suggesting that the best predictions are for SPR values greater than 6. These results apply to the simulated punch test and Timoshenko's theories, and also, it could be extended to the experimental PST.

The feasibility of the PST to characterize mechanical properties in materials strongly depends on the assembly geometry (punch and sample) when the SPR is lower than 3. When properties, such as Young Modulus, tensile yielding stress and ultimate tensile stress are going to be determined by PST, it is recommended to use samples with an SPR higher than 6 in order to improve the results accuracy. On the other hand, properties such as shear modulus, shear resistance or shear yield could be estimated more accurately by using samples with a low SPR value [44].

The current work has provided an in-depth understanding of strain energy accumulation and its radial distribution in circular plates under punch tests. The simulations provide information about the way in which shear and bending effects are influenced by the disk geometry and the load conditions. The effect of the geometry of the systems was also evaluated in great detail using disk-like-samples.

Future work can use the results presented for two purposes: (i) To guide the design of a punch shear test device, which would help to define the magnitude order of the dimensions of the devices and forces to be managed. (ii) To use the information given by punch shear tests to develop models to assess the ballistic performance of materials as a complement and extension of the work started by the researchers Gama and Gillespie [20-23].

Acknowledgements

This research was supported by COLCIENCIAS (Departamento Administrativo de Ciencia, Tecnología e Inovación), ThoR S.A. and TermalTec S.A. The authors would like to thank Anderson Valencia (ThoR S.A.), Jairo Montes (TermalTec S.A.), the Colombian Army and COTECMAR (Corporación de Ciencia y Tecnología para el Desarrollo de la Industria Naval, Marítima y Fluvial) for their constant support and helpful discussions.

References

[1] Kurtz, M., Foulds, R., Jewett, W., Srivastav, S. and Edidin, A., Validation of a small punch testing technique to characterize the mechanical behavior of UHMW-PE, Biomaterials 18, pp. 1659-1663, 1997. DOI: 10.1016/S0142-9612(97)00124-5

[2] Giddings, V., Kurtz, M., Jewett, W., Foulds, R. and Edidin, A., A small punch test technique for characterizing the elastic modulus and fracture behavior of PMMA bone cement used in total joint replacement Biomaterials 22, pp. 1875-1881, 2001. DOI: 10.1016/S0142-9612(00)00372-0

[3] Byun, T., Lee, E., Hunn, J., Farrell, K. and Mansur, L., Characterization of plastic deformation in a disk bend test. J. Nuclear Materials 294, pp. 256-266, 2001. DOI: 10.1016/S0022-3115(01)00484-6

[4] Ju, B., Ju, Y., Saka, M., Liu, K. and Wan, K., A systematic method for characterizing the elastic properties and adhesion of a thin polymer membrane, Int. J. Mechanical Sciences 47, pp. 319-332, 2005. DOI: 10.1016/j.ijmecsci.2005.02.006

[5] Kim, B.J., Sim, Y.B., Lee, J.H., Kim, M.K. and Lim, B.S., Application of small punch creep test for Inconel 617 alloy weldment, Procedia Engineering 10, pp. 2579-2584, 2011. DOI: 10.1016/j.proeng.2011.04.425

[6] Zhou, Z., Zheng, Y., Ling, X. and Zhou, R., A study on influence factors of small punch creep test by experimental investigation and finite element analysis, Materials Science and Engineering A527, pp. 2784-2789, 2010. DOI: 10.1016/j.proeng.2011.04.425

[7] Stratford, G., Persio, F. and Klaput, J., Miniaturized creep testing using the small punch (sp) test technique, in 11th International Conference on Fracture, March 20-25, 2005, Turin, Italy.

[8] Kim, M., Oh, Y. and Lee, B., Evaluation of ductile–brittle transition temperature before and after neutron irradiation for RPV steels using small punch tests, Nuclear Engineering and Design 235, pp. 1799-1805, 2005. DOI: 10.1016/j.nucengdes.2005.05.014

[9] Chang, Y., Kim, J., Choi, J., Kim, Y., Kim, M. and Lee, B., Derivation of ductile fracture resistance by use of small punch specimens, Eng. Fracture Mechanics 75, pp. 3413-3427, 2008. DOI: 10.1016/j.engfracmech.2007.06.006

[10] Abendroth, M. and Kuna, M., Identification of ductile damage and fracture parameters from the small punch test using neural networks, Engineering Fracture Mechanics 73, pp. 710-725, 2006. DOI: 10.1016/j.engfracmech.2005.10.007

[11] Edidin, A.A., Rimnac, C.M., Goldberg, M. and Kurtz, M., Mechanical behavior, wear surface morphology, and clinical performance of UHMWPE acetabular components after 10 years of implantation, Wear 250, pp. 152-158, 2001. DOI: 10.1016/S0043-1648(01)00616-0

[12] Ge, S., Kang, X. and Zhao, Y., One-year biodegradation study of UHMWPE as artificial joint materials: Variation of chemical structure and effect on friction and wear behavior, Wear 271, pp. 2354-2363, 2011. DOI: 10.1016/j.wear.2010.11.048

[13] Jaekel, J., McDonald, W. and Kurtz, M., Characterization of PEEK biomaterials using the small punch test, J. Mechanical behavior biomedical materials IV, pp. 1275-1282, 2011. DOI: 10.1016/j.jmbbm.2011.04.014

[14] Kurtz, M., Jewett, C.W., Bergstrom, S., Foulds, R. and Edidin, A., Miniature specimen shear punch test for UHMWPE used in total joint replacements Biomaterials 23, pp. 1907-1919, 2002. DOI: 10.1016/S0142-9612(01)00316-7

[15] Foletti, S., Madia, M., Cammi, A. and Torsello, G. Characterization of the behavior of a turbine rotor steel by inverse analysis on the small punch test, Procedia Engineering 10, 3628-3635, 2011. doi:10.1016/j.proeng.2011.04.597

[16] Fleury, E. and Ha, J., Small punch tests to estimate the mechanical properties of steels for steam power plant: I. Mechanical strength, International Journal of Pressure Vessels and Piping 75, pp. 699-706, 1998. DOI: 10.1016/S0308-0161(98)00074-X

[17] Fleury, E. and Ha, J., Small punch tests to estimate the mechanical properties of steels for steam power plant: II. Fracture toughness, International Journal of Pressure Vessels and Piping 75, pp. 707-713, 1998. DOI: 10.1016/S0308-0161(98)00075-1

[18] Hoffmann, M. and Birringer, R., Quantitative measurements of Young's modulus using the miniaturized disk-bend test, Materials Science and Engineering A202, pp. 18-25, 1995. DOI: 10.1016/0921-5093(95)09817-8

[19] León, C. and Drew, R., Small punch testing for assessing the tensile strength of gradient Al/Ni–SiC composites, Materials Letters 56, pp. 812-816, 2002. DOI: 10.1016/S0167-577X(02)00619-5

[20] Xiao, J.R., Gama, B.A. and Gillespie, J.W. Jr., Progressive damage and delamination in plain weave S-2 glass/SC-15 composites under quasi-static punch-shear loading, Composite Structures 78, pp. 182-196, 2007. DOI: 10.1016/j.compstruct.2005.09.001

[21] Gama, B.A. and Gillespie, J.W. Jr., Punch shear based penetration model of ballistic impact of thick-section composites, Composite Structures 86, pp. 356-369, 2008. DOI: 10.1016/j.compstruct.2007.11.001

[22] Erkendirci, F. and Gama, B., Quasi-static penetration resistance behavior of glass fiber reinforced thermoplastic composites, Composites: Part B, pp. 3391-3405, 2012. DOI: 10.1016/j.compositesb.2012.01.053

[23] Tehrani, M., Boroujeni, A., Hartman, T., Haugh, T., Case, S. and Al-Haik, M., Mechanical characterization and impact damage assessment of

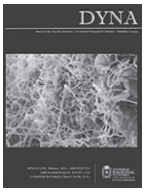
a woven carbon fiber reinforced carbon nanotube–epoxy composite, Compos Sci Technol 75, pp. 42-48, 2013. DOI: 10.1016/j.compscitech.2012.12.005
[24] Mäder, E., Gao, S. and Plonka, R., Static and dynamic properties of single and multi-fiber/epoxy composites modified by sizings, Composites Science and Technology 67, pp. 1105-1115, 2007. DOI: 10.1016/j.compscitech.2006.05.020
[25] Chau, K., Wei, X., Wong, R. and Yu, T., Fragmentation of brittle spheres under static and dynamic compressions: experiments and analyses, Mechanics of Materials 32, pp. 543-554, 2000. DOI: 10.1016/S0167-6636(00)00026-0
[26] Tronskar, J., Mannan, M. and Lai, M., Correlation between quasi-static and dynamic crack resistance curves, Engineering Fracture Mechanics 70, pp. 1527-1542, 2003. DOI: 10.1016/S0013-7944(02)00148-0
[27] Marsavina, L., Linul, E., Voiconi, T. and Sadowski, T., A comparison between dynamic and static fracture toughness of polyurethane foams, Polymer Testing 32, pp. 673-680, 2013. DOI: 10.1016/j.polymertesting.2013.03.013
[28] Jenq, S., Jing, H. and Chung, C., Predicting the ballistic limit for plain woven glass/epoxy composite laminate, Int. J. Impact Engng 15(4), pp. 451-464, 1994. DOI: 10.1016/0734-743X(94)80028-8
[29] Potti, S. and Sun, C., Prediction of impact induced penetration and delamination in thick composite laminates, Int. J. Impact Engng 19(1), pp. 31-48, 1997. DOI: 10.1016/0734-743X(94)80028-8
[30] Meunier, L., Chagnon, G., Favier, D., Orgéas, L. and Vacher, P., Mechanical experimental characterization and numerical modeling of an unfilled silicone rubber, Polymer Testing 27, pp. 765-777, 2008. DOI: 10.1016/j.polymertesting.2008.05.011
[31] Korochkina, T., Jewell, E., Claypole, T. and Gethin, D., Experimental and numerical investigation into nonlinear deformation of silicone rubber pads during ink transfer process, Polymer Testing 27, pp. 778-791, 2008. DOI: 10.1016/j.polymertesting.2008.06.003
[32] Shergold, O., Fleck, N. and Radford, D., The uniaxial stress versus strain response of pig skin and silicone rubber at low and high strain rates, Int. Journal of Impact Engineering 32, pp. 1384-1402, 2006. DOI: 10.1016/j.ijimpeng.2004.11.010
[33] Podnos, E., Becker, E., Klawitter, J. and Strzepa, P., FEA analysis of silicone MCP implant, Journal of Biomechanics 39, pp. 1217-1226, 2006. DOI: 10.1016/j.jbiomech.2005.03.019
[34] Horgan, C. and Murphy, J., Compression tests and constitutive models for the slight compressibility of elastic rubber-like materials, Int. Journal of Engineering Science 47, pp. 1232-1239, 2009. DOI: 10.1016/j.ijengsci.2008.10.009
[35] Sasso, M., Palmieri, G., Chiappini, G. and Amodio, D., Characterization of hyperelastic rubber-like materials by biaxial and uniaxial stretching tests based on optical methods, Polymer Testing 27, pp. 995-1004, 2008. DOI: 10.1016/j.polymertesting.2008.09.001
[36] Charlton, D., Yang, J. and Teh, K., A review of methods to characterize rubber elastic behavior for use in finite element analysis, Journal of Rubber Chemistry and Technology 67, pp. 481-503, 1994. DOI: 10.5254/1.3538686
[37] Behroozi, M., Olatunbosun, O. and Ding, W., Finite element analysis of aircraft tyre – Effect of model complexity on tyre performance characteristics, Materials and Design 35, pp. 810-819, 2012. DOI: 10.1016/j.matdes.2011.05.055
[38] Hibbitt, Karlsson. ABAQUS Theory Manual, V. 6.3, Sorensen Inc., Pawtucket. USA, 2002.
[39] Wood, L. and Martin, G., Compressibility of natural rubber at pressures below 500 KG/CM^2, Rubber Chem. Technol., 37, 850 P., 1964. DOI: 10.6028/jres.068A.022
[40] Holownia, B., Effect of Carbon Black on Poisson's Ratio of Elastomers, Rubber Chem. Technol., 48, 246 P., 1975. DOI: 10.5254/1.3547450
[41] Rightmire, G.K., An experimental method for determining Poisson's ratio of elastomers, J. Lubrication Technol., 381, 1970. DOI: 10.1115/1.3451416
[42] Timoshenko. S. and Woinowsky-Krieger, S., Theory of plate and shells, 2nd ed., McGraw-Hill, Singapore, 1970.
[43] Manzella, A., Gama, B. and Gillespie J., Jr., Effect of punch and specimen dimensions on the confined compression behavior of S-2 glass/epoxy composites, Comp. Structures 93, pp. 1726-1737, 2011. DOI: 10.1016/j.compstruct.2010.11.006
[44] Hankin, G., Toloczko, M., Hamilton, M. and Faulkner., R., Validation of the shear punch tensile correlation technique using irradiated materials, Journal of Nuclear Materials 258-263, pp. 1651-1656, 1998. DOI: 10.1016/S0022-3115(98)00203-7
[45] Ogden, R.W., Non-Linear elastic deformations, Dover Publications Incorporated. New York, 1997.
[46] Kaden, J., Computational simulations of fiber reinforced hyperelastic materials, McGill, University, Quebec, 2007,
[47] Ta, A., Labed, N., Holweck, F., Thionnet, A. and Peyraut, F., A new invariant-based method for building biomechanical behavior laws-Application to an anisotropic hyperelastic material with two fiber families, International Journal of Solids and Structures 50, pp. 2251-2258, 2013. DOI: 10.1016/j.ijsolstr.2013.03.033
[48] Kaster, T., Sack, I. and Samani, A., Measurement of the hyperelastic properties of ex-vivo brain tissue slices, Journal of Biomechanics 44, pp. 1158-1163, 2011. DOI: 10.1016/j.jbiomech.2011.01.019
[49] Feng, Z., Peyraut, F. and He, Q., Finite deformations of Ogden's materials under impact loading, Int. J. of Non-Linear Mechanics 41, pp. 575-585, 2006. DOI: 10.1016/j.ijnonlinmec.2006.02.003
[50] Agostiniani, V. and DeSimone, A., International Journal of Non-Linear Mechanics 47, pp. 402-412, 2012.
[51] Isaza, J., Mariaka I. y Ramirez, J., Caracterización de propiedades mecánicas mediante análisis inverso del método de los elementos finitos combinado con ensayo de indentación, DYNA 79(179), pp. 126-133, 2013.

Lopera-Valle, received his BSc. in Mechanical Engineering in 2012 from the Universidad Nacional de Colombia in Medellin, Colombia, where he has been member of the Tribology and Surfaces research group since 2010. He has also worked on topics related with Thermal Barrier Coatings (TBC) to improve the performance of combustion turbines. In 2015, he received his MSc.in Mechanical Engineering from the University of Alberta in Canada. The use of flame-sprayed metallic coatings as heating elements for polymer-based structures that are exposed to cold environments was the main topic of his thesis dissertation. Currently, he is taking a PhD. in Chemical and Materials Engineering at the University of Alberta, Canada and working on the development and simulation of thin-film sensors for medical and biological applications.
ORCID ID: 0000-0001-6806-7541.

F.A. Suárez-Bustamante, received his BSc. in Mechanical Engineering degree in 2002 from the Universidad Nacional de Colombia (UN) – Medellín, Colombia, where he has been a member of the Tribology and Surfaces Group from 2000 and has worked on topics related with mechanical contact, surface engineering and wear modeling. From 2002 until 2005 he worked as an engineering analyst in a hydro-electrical power generation plant that belongs to ISAGEN E.S.P. In 2008, he obtained his MSc.of Science degree in Engineering from the same university and worked on the effect of surface conditioning on hydrodynamic lubrication of sliding bearings and their tribological performance. He is currently a Dr. candidate at UN and works on the design of a composite material for body armor. He is also an entrepreneur and owner of a technology based company, Faro Tecnológico S.A.S.
ORCID ID: 0000-0002-9810-6641.

J.P. Hernandez-Ortiz, received a BSc. in Mechanical Engineering degree in 1998 from the Universidad Pontificia Bolivariana, Medellin, Colombia, where he worked as a research assistant at the Energy and Thermodynamic Institute until 2000. He obtained his PhD degree in 2004, from the University of Wisconsin-Madison, USA, in the Department of Mechanical Engineering and minored in Chemical Engineering. From 2004 to 2007, he undertook postdoctoral research in the Department of Chemical and Biological Engineering at the University of Wisconsin-Madison, USA. Currently, he is a full professor in the Department of Materials, Facultad de Minas, at the Universidad Nacional de Colombia, Medellín, Colombia. He has published more than 50 papers and holds honorary positions at the University of Wisconsin-Madison and the Institute for Molecular Engineering at the University of Chicago, USA. His research interests are based on multi-scale modeling of complex systems for biological and structured materials applications.
ORCID ID: 0000-0003-0404-9947.

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS



DYNA
http://dyna.medellin.unal.edu.co/

Optimal estimating the project completion time and diagnosing the fault in the project

M. T. Hajali-Mohamad [a], M. R. Mosavi [b] & K. Shahanaghi [c]

[a] *Facultad de Ingeniería Industerial, Irán Universidad de Ciencia y Tecnología, Irán. hajialinajar@iust.ac.ir.*
[b] *Facultad de Ingeniería Eléctrica, Irán Universidad de Ciencia y Tecnología, Irán. m_mosavi@iust.ac.ir*
[c] *Facultad de Ingeniería Industerial, Irán Universidad de Ciencia y Tecnología, Irán. shahanaghi@iust.ac.ir.*

Received: Juli 5th, 2014. Received in revised form: December 10th, 2015. Accepted: January 10th, 2016

Abstract
The main objective of the project management team is to implement the project taking into consideration the Budget, schedule and constraints. In addition, project accomplishment, especially with large projects, requires the project to be correctly envisaged. Earned value (EV) management is a valuable technique for analyzing and controlling the performance of the project and predicting the total cost before its completion. Thus, fuzzy systems such as Adaptive Network based on the Fuzzy Inference System (ANFIS) and Parallel Structure based on the Fuzzy System (PSFS) are used to predict the project completion time. In this paper, the plan value diagram is used to predict the earn value diagram using three methods. These three methods are based on the PSFS and Neural Networks (NNs), which help with the implementation of the projects in organizations. The results of these three methods decreased the prediction error of the EV diagram by up to 2%.

Keywords: Earned Value, Plan Value, ANFIS, PSFS, Neural Network.

Estimación optima de terminación el tiempo del proyecto y diagnóstico de la falla en el proyecto

Resumen
El objetivo principal del equipo de gestión del proyecto es poner en práctica el proyecto con respecto al presupuesto, horario y cubrir las limitaciones del proyecto. Además, la realización del proyecto (especialmente los grandes proyectos,) requiere la predicción adecuada del proyecto. Gestión del valor ganado es una técnica valiosa para el análisis y control de la ejecución del proyecto y predice el costo total del proyecto antes de la finalización del proyecto. Por lo tanto, Fuzzy sistemas tales como la Red de adaptación basada en Fuzzy sistema de inferencia (ANFIS) y Estructura paralela basada en el Fuzzy Sistema (PSFS) se utilizan para predecir el tiempo de finalización del proyecto. En éste artículo, el diagrama de valor del plan se utiliza para predecir el diagrama valor en tres métodos. Estos tres métodos se basan en los PSFS y Redes Neuronales que son capaces de ser poner en práctica los proyectos en las organizaciones. Los resultados de estos tres métodos disminuyeron el error de predicción del valor obtenido diagrama de hasta 2%.

Palabras clave: Valor Ganado, Valor del Plan, ANFIS, PSFS, NN.

1. Introduction

Earned Value Management (EVM) is a project management technique used to measure project progress in an objective manner. It is a technique used to control and manage projects by combining dynamic project data and expressing it usefully, highlighting risks and measuring the project [1]. The EVM method provides early indications of project performance to highlight the need for eventual corrective actions [2]. Nowadays, implementing Earned Value (EV) techniques has many benefits and would enhance the project's cost and scheduled performances. However, the research on the EV is very restricted. Fig. 1 shows Plan Value (PV), EV and AC curves in EVM.

There are two basic approaches in the literature: model-based and statistical methods to solve the EVM problem. The

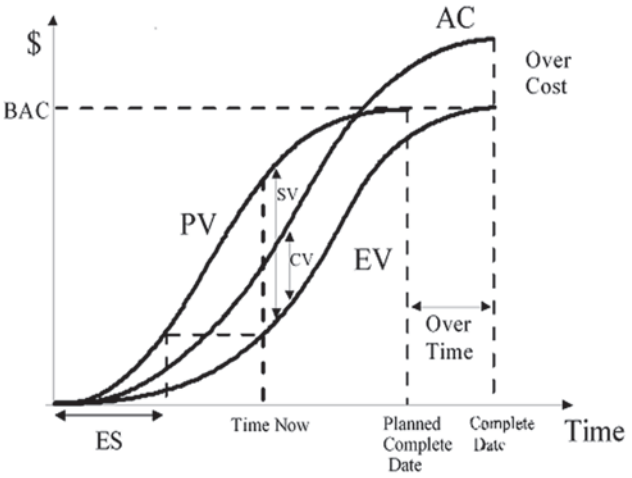


Figure 1. PV, EV and AC curves in EVM.
Source: the authors.

model-based approach is based on the assumption that there is enough prior information to build an accurate model.

Naeni et al. developed a new fuzzy EV model which had the advantage of developing and analyzing the EV indices, as well as the time and the cost estimates when there is uncertainty [3]. The developed model is very useful in evaluating the progress of a project where uncertainty arises. Vandevoorde and Vanhoucke concluded that the most reliable method for estimating time at completion is the Earned Schedule (ES) method [4].

They were the first authors who widely compared the three time prediction methods and tested them with a simple one activity project and a real life data set. A reliable forecasting method of the final cost and duration is proposed by Lipke et al. to improve the capability of project managers for making informed decisions. They use data from 12 real life projects to improve the capability of project managers to be able to make informed decisions by providing a reliable forecasting method of the final cost and duration. Hunter et al. focused on the implementation of EVMS on the Radiation Belt Storm Probes (RBSP) project, explained EV processes and the implementation's cost and analyzed the benefits of EVMS to provide insight into cost/benefit considerations for other projects regarding EVMS implementation [5]. Acebes et al. proposed an innovative and simple graphical framework for project control and monitoring in order to integrate the dimensions of project cost and schedule with risk management, thus extending the EV methodology [6]. They build a graphical methodology to know when a project remains "out of control" or "within expected variability" during the project lifecycle. With this aim, they defined and represented new control indexes and new cumulative buffers. Cioffi presented a new notation for the EV analysis to make EV mathematics more transparent and flexible [7].

Plaza and Turetken present a spreadsheet-based decision support tool that automates the calculations and analyses in EVM/LC [8]. Von Wangenheim presents an educational board game to reinforce and teach the application of EVM concepts in the context of undergraduate computing programs, complementing expository lessons on EVM basics [9]. The game has been developed based on project management fundamentals and teaching experience in this area. The results point out a very positive effect of the game on social interaction, engagement, immersion, attention and relevance to the course objectives. The objective of Elshaer's paper is twofold: the first being to study the impact of the activities' sensitivity information on the forecasting accuracy of the ESM method [10]. The second is to test the claim that in normal conditions the project performance indicator provided by ESM at higher work breakdown structure is reliable. Vanhoucke presents two project tracking methods to address project problems [11]. In this paper, a bottom-up and a top-down project tracking approach within a corrective action framework is applied to a large and diverse set of fictitious projects which are subject to Monte-Carlo simulations to simulate fictitious project progress under uncertainty. Maravas and Pantouvakis proposed a new method to calculate project cash flows, in which the cost and duration of each activity are understood as fuzzy numbers [12]. Yao et al. present a fuzzy, stochastic, single-period model for cash management to provide financial decision makers with more insight into real cash management problems [13]. Mortaji et al. developed an efficient approach to calculate Estimate At Completion (EAC) [14]. Artificial Intelligence (AI) methods are used in different fields and have attracted much attention [15-17]. Cheng et al. employed AI approaches including fuzzy logic, K-means clustering, a genetic algorithm, and Neural Networks (NNs) to gain strategic control over project cash flows [18]. The evolutionary fuzzy support vector machine inference model for time series data as an alternative approach is presented by Cheng and Roy to predict cash flow [19]. Warburton proposes a formal method to include time dependence into EV management [20]. The model requires three parameters: the reject rate of activities, the cost overrun parameter, and the time to repair the rejected activities. The model is built on the well-established Putnam-Norden-Rayleigh (PNR) labor rate profile, which is a useful representation for large software projects.

Statistical approaches try to analyze series of observations that are obtained by the system and then they predict the future behavior. Many recent prediction theories have been developed based on statistical approaches due to the complexity of building an accurate model. However, intelligent systems such as NNs and fuzzy or a hybrid methods like Adaptive Network based on Fuzzy Inference System (ANFIS) have relatively large errors and cannot be applied to real projects [21,22]. The fuzzy systems can easily address uncertainties that exist in the complex physical systems behavior. Adaptive learning algorithms can generate fuzzy rules to predict chaotic time series when knowledge about the system and the input-output format of the numeric data is available [23]. Time series estimating based on Parallel Structure based on Fuzzy System (PSFS) is one of the methods used for this purpose [24]. PSFS consists of a number of parallel fuzzy systems. Each fuzzy system estimates the future value of the signal based on older signals with different time samples. PSFS considers the final estimation value based on the average output of each fuzzy system, except the minimum and maximum values [25].

A continuous increase in the complexity, efficiency, and reliability of modern industrial systems necessitates continuous development in the control and fault diagnosis theory and practice. This is associated with the increasing demands for higher system performance and product quality on one side and more cost efficiency on the other. The complexity and the automation degree of technical processes are also continuously growing. This development calls for more system safety and reliability. Today, one of the most critical issues surrounding the design of automatic systems is system reliability and dependability [26]. A traditional way to improve the system reliability and dependability is to enhance the quality, reliability and robustness of individual system components like sensors, actuators, controllers and computers. Even so, a fault-free system operation cannot be guaranteed. Process monitoring and fault diagnosis are therefore becoming an essential ingredient in a modern automatic control system [27]. The overall concept of fault diagnosis consists of the following three essential tasks: **Fault detection**: detection of the occurrence of faults in the functional units of the process. These lead to undesired or intolerable behavior of the whole system. **Fault isolation**: fault diagnosis system outputs are also alarm signals that indicate the occurrence of the faults or classified alarm signals that show which fault has occurred. The additionally provide information about the type or magnitude of the fault that has occurred [26,27].

In this paper, the plan value diagram is used to predict the earn value diagram using three methods. These three methods are based on the PSFS and NNs, which implement projects in organizations. The results of these three methods decreased the predicted error of the EV diagram by up to 2%.

This paper is organized as follows: In section 2, the structure of the new PSFS is presented. Section 3 shows the simulation. The NNs models are illustrated in section 4. A comparison of these techniques will be expressed and presented in section 5. Finally, in section 6, conclusions are made.

2. Proposed parallel structure based on fuzzy systems

PSFS is made by some fuzzy systems that have parallel connection and it is used for predicting time series. As shown in Fig. 2, the fuzzy system used in PSFS is ANFIS network, which includes 6 inputs, 1 output and 64 rules.

PSFS wants to predict the $\hat{x}_P(k+r)$ this is the data value in time step (k+r). P-th fuzzy system has an output $\hat{x}_P(k+r)$ that is based on previous data $\hat{x}(k+r-p)$, …, $\hat{x}(k+r-mp)$. PSFS can calculate the value of $\hat{x}(k+r)$ based on different outputs of fuzzy systems $\hat{x}_1(k+r)$, $\hat{x}_2(k+r)$, …, $\hat{x}_N(k+r)$. PSFS calculated the final value of $\hat{x}(k+r)$ by averaging the measurements between all fuzzy system outputs except maximum and minimum values. This value is shown in eq. (1).

$$\hat{x}(k+r) = \frac{1}{N-2}[\sum_{i=1}^{N} \hat{x}_i(k+r) - max_i\hat{x}_i(k+r) - min_i\hat{x}_i(k+r)] \quad (1)$$

The proposed PSFS is composed of multiple parallel fuzzy subsystems, which are used for time series [28]. Fig. 3

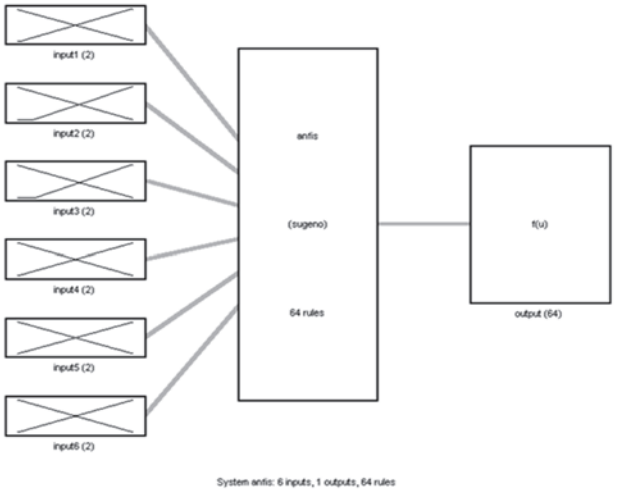


Figure 2. The proposed ANFIS network structure with 6 inputs and 64 rules. Source: the authors.

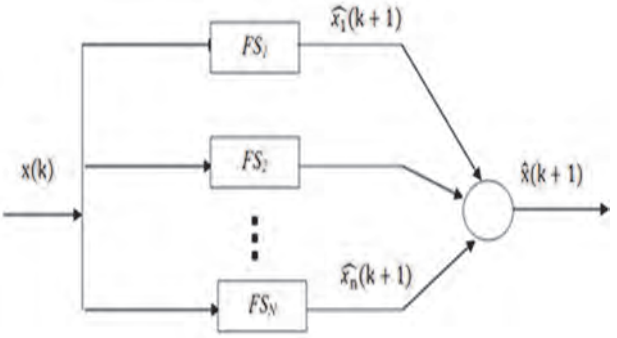


Figure 3. The proposed PSFS structure. Source: the authors.

shows the proposed PSFS. The PSFS consists of the N parallel fuzzy subsystems FS1, FS2... FSN. Each fuzzy system for a specific time with index k+1 predicts a value for the time series. The PSFS consequence predicts the end value $\hat{x}(k+\tau)$, in terms of the particular decision making.

In the previous PSFS structures, the average outputs of the fuzzy networks are considered as outputs that eliminate the maximum and minimum in each algorithm. Eliminating values will lead to missing information when the outputs of fuzzy networks are close to each other. By comparing the network output variance, the decision will be made to eliminate, or not, the extreme values in the proposed algorithm.

Firstly, the new algorithm is calculated based on output variance of a fuzzy network $\{\hat{x}_i(k+r)\}_{i=1}^{N}$, and also the standard deviation represented by σ_T. After that maximum output is calculated $\max\{\hat{x}_i(k+r)\}_{i=1}^{N}$, then it is removed from the list of data. The new variance is calculated by N-1 of the remaining samples and represented by σ_{max}. If $\sigma_T > K\sigma_{max}$, then data is sparse, so maximum a value must be removed by the previous algorithm; otherwise, the maximum value is set to zero. This process is also applied to minimum output. In the simulation step the K value is set to three.

3. Simulation

To facilitate network training and to prevent its divergence (discontinuity points), this paper undertakes an innovation. The input data includes the plan history and the project operation, as shown in Fig. 4. In this figure, the first half of the bell is considered as the plan, and the second half shows the activity operation.

1400 pieces of data are used to train the ANFIS network. Using these data, four ANFIS networks are trained with steps $P_1=2$, $P_2=3$, $P_3=4$,and $P_4=5$. Then, to test the trained networks with 200 further pieces of data, $(3*P+1)$ of the train data are used to generate the PSFS outputs. Finally, the generated data is compared to the original signal. The dataset (j303_10) that was generated by ProGen for RCPSP Kolisch and Hartmann is used in this paper [29]. A developed program was used in [30] by using VBA-MSP tools in Microsoft Project in order to generate periodical data.

It can be assumed that the operation plan is implemented until time 30 in order for a fault to be detected. To diagnose the fault, the absolute value of the error curve of the PSFS network should be measured with the curve output of the plan value at time E1. Also, the absolute value of the project output curve should be calculated by the plan value curve at time E2. By comparing these two values an error in the system can be addressed. Fig. 5 shows the progress plan and actual data for p1 = 2, p2 = 3, p3 = 4 and p4 = 5, respectively.

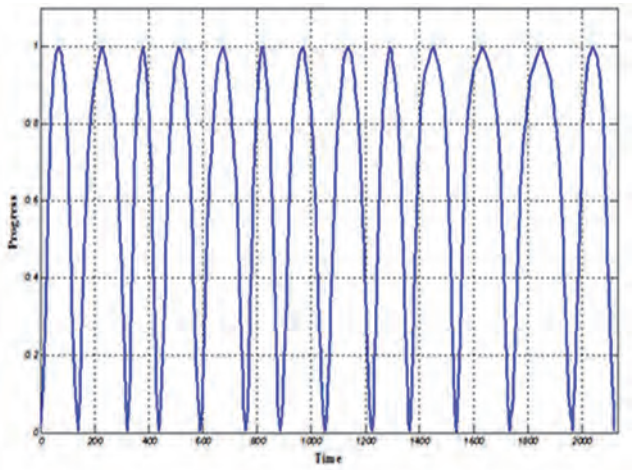


Figure 4. The input data includes the history of the program and project progress.
Source: the authors.

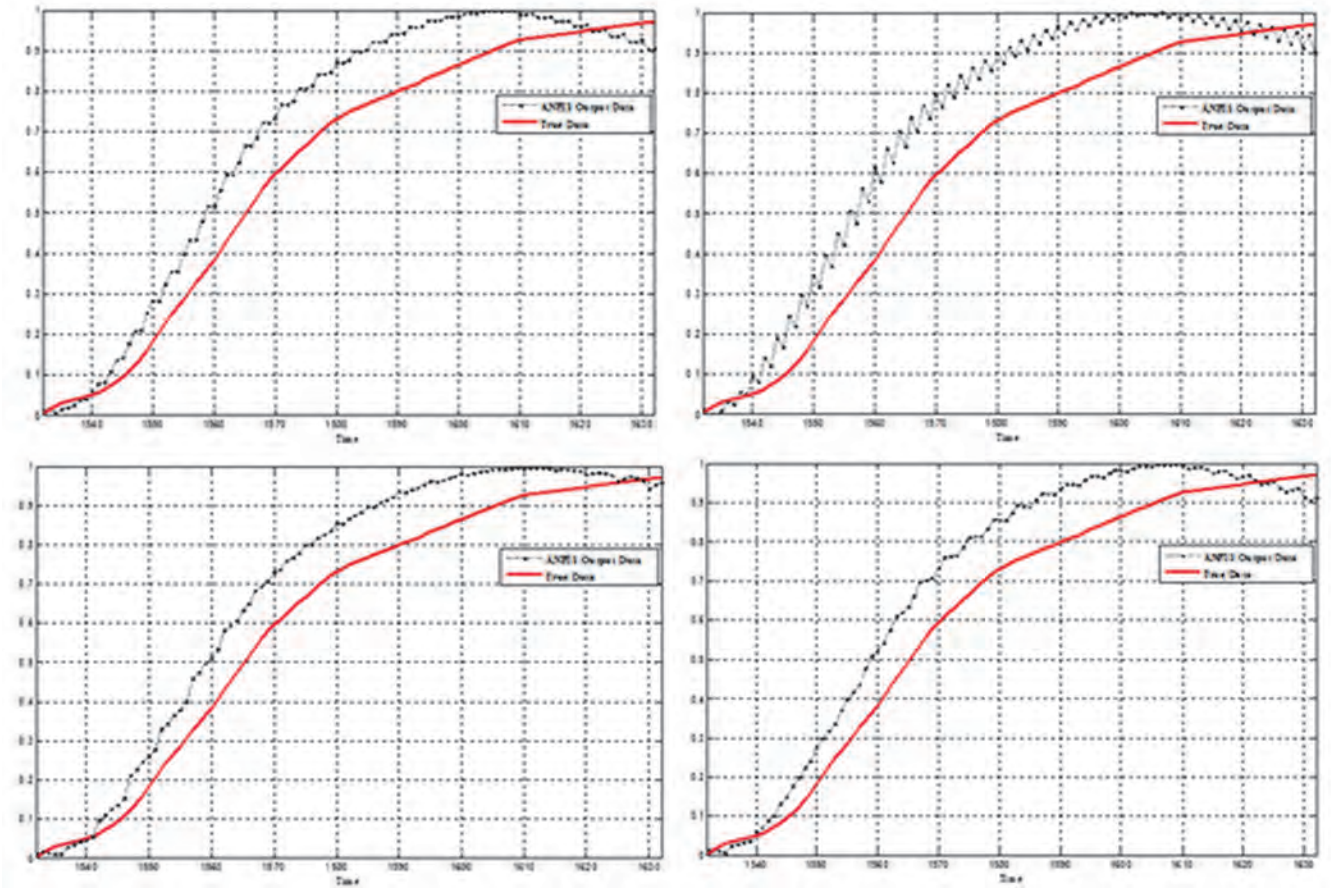


Figure 5. Comparing the progress plan and actual data for p1 = 2, p2 = 3, p3 = 4, and p4 = 5, respectively.
Source: the authors.

4. Neural network-based method

The NN has many applications, estimation being one of them. By using the NN in estimation, the subsequent data will be estimated, and a set of data with a good accuracy will be used [31]. Assume that *N* projects have been undertaken by a company, and the PV and EV diagrams of these projects are available. The purpose is now to estimate the EV diagram of the project $N+1^{th}$ using its PV diagram. There are two methods of doing this: the future state method, and by using previous data.

In the prediction problems with NNs, the future state prediction of a system is based on its previous state [31]. Since high precision and speed have a priority, different methods exist to implement networks that are similar to each other. Fig. 6 shows the proposed structure:

4.1. The first method simulation

The NN used in this project to predict the next state of the recurrent network is a two-layer NN structure that is shown in Fig. 7.

Also, the input data and the objective functions in the network can be selected below, if X indicates data:

$I_1=[X_1, X_2, X_3, X_4]$
$I_2=[X_2, X_3, X_4, X_5]$
…
$I_{n-3}=[X_{n-3}, X_{n-2}, X_{n-1}, X_n]$

$T_1=[X_5]$
$T_2=[X_6]$
…
$T_{n-3}=[X_{n+1}]$

When using previous data, the main issue is to train the NN with available historical data that it needs to use for new data. In this method, the PV graphs appear sequentially and are defined as NN inputs. Also, this is what happened for EV graphs, and they were defined as objective NN functions. There are two sine graphs in Fig. 4; one of is the input and the other (EV graph) is the objective function.

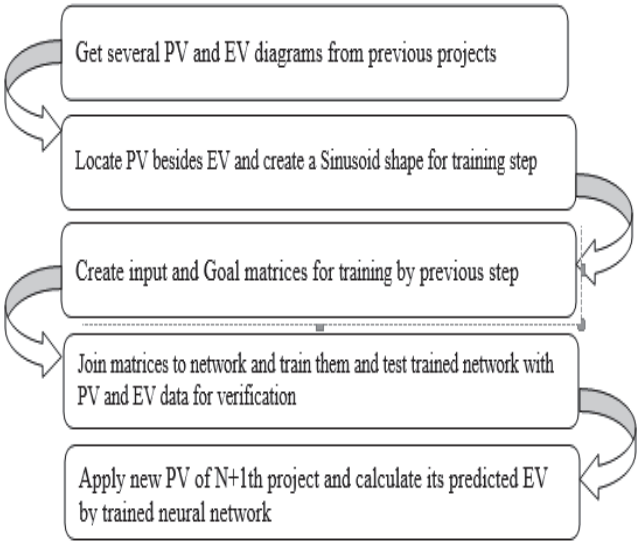


Figure 6. Different steps in the proposed method.
Source: the authors.

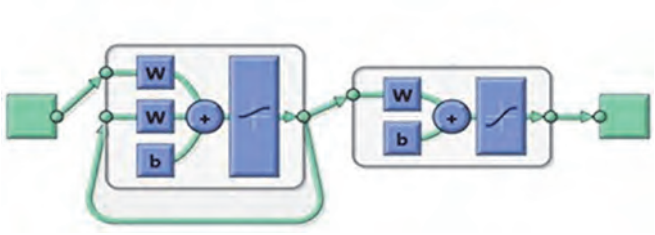


Figure 7. The NN used in the first method.
Source: the authors.

4.2. The second method simulation

In this method, the network is trained with 1400 pieces of data from PV and EV graphs and is then tested with another 800 pieces of data from them. The various NNs are used to train data; the best of them is chosen based on the error calculation measurements, (Root Mean Square (RMS) or Mean Squared Error (MSE)). It should be noted that the number of hidden layers and neurons in each of the networks is the same. According to the measurements taken and the relative uniformity of data, Radial Basis Function Neural Network (RBF NN) is the best network for aforesaid data.

5. Results

After collecting the required information from the organization and implementing the algorithm, the following results in three strategies' different sub-sections have been obtained.

5.1. PSFS results

Fig. 8 is drawn for the P = 5 step and ANFIS network. The actual and estimated EV is also shown in this figure.

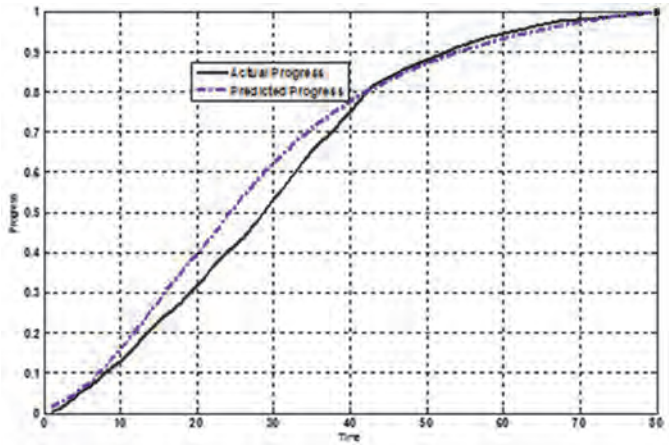


Figure 8. Comparing the progress plan data and PSFS output.
Source: the authors.

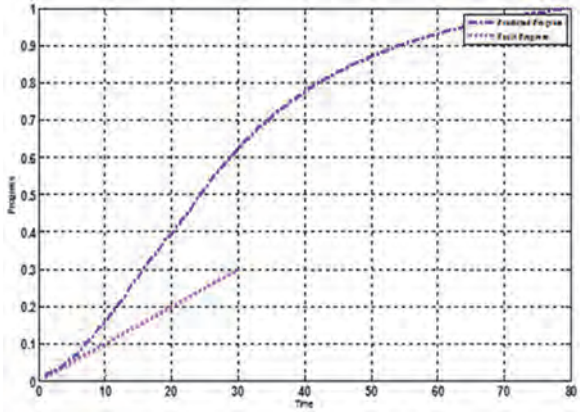


Figure 9. Comparing the predicted progress plan and a plan with fault.
Source: the authors.

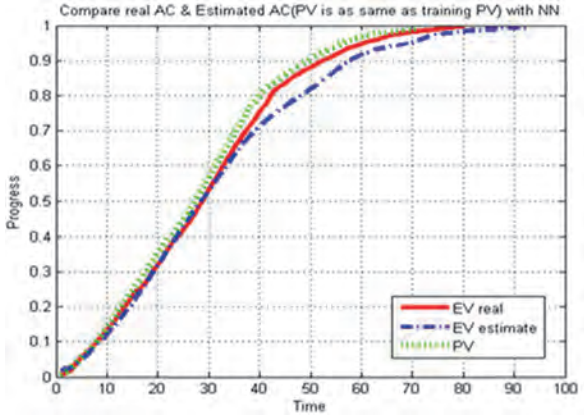


Figure 10. Comparing the predicted and actual EV graph, as well as the PV graph, with the first NN.
Source: The authors.

Using the network described in section 2, the PSFS system is built. The E1=1.6373 and E2 = 2.6978 are obtained in Fig. 9. It can be demonstrated that an error has occurred in the system due to the fact that E2> E1.

5.2. Results of the future state prediction based on neural network

The recurrent NNs were applied to the data listed in section 4.1 and the result is shown in Fig. 10.

5.3. Results of the previous data based on neural network

The RBF NN was applied to the data listed in section 4.2 and the result is shown in Fig. 11.

The fault detection in the NNs, in its best state (ANN Method II) is E1=0.02, whereas E2=2.69. This means that the second NN method is more accurate in detecting faults. The results are summarized in Table 1.

6. Conclusion

As is clear from the above results, the proposed algorithms are able to reduce the estimation error by up to 2

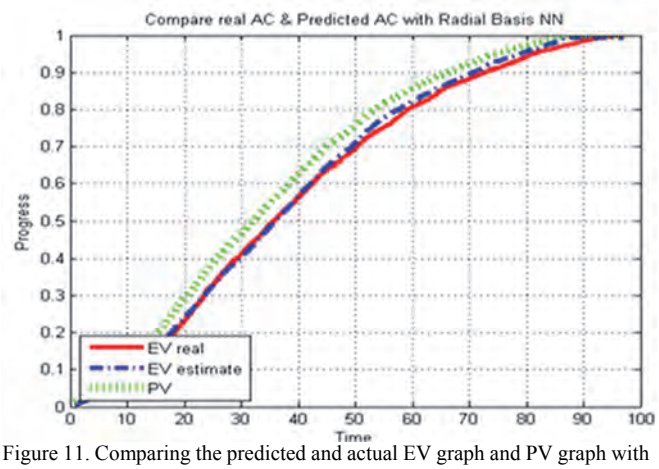


Figure 11. Comparing the predicted and actual EV graph and PV graph with the second NN.
Source: the authors.

Table 1.
Error rates of different methods for estimating the EV graph.

Error / Algorithm	RMS	MSE	MAPE (%)
PSFS	0.0521	0.00832	3.22
Future state prediction based on NN	0.0345	0.00132	1.35
Previous data based on NN	0.0116	0.00014	0.34

Source: the authors.

percent. However, in many cases, the maximum amount of the estimation error reduction is equivalent to 42 percent. In this paper, we developed new algorithms to estimate the EV graphs based on PSFS and NN. Also, the proposed algorithms effectively estimate the EV graphs and diagnose the fault with high accuracy. It must be mentioned that the first NN was introduced for the real data (non-uniform) because of its good performance.

References

[1] Kerzner, H., Project management case studies, Wiley, 2012.
[2] Fleming, Q.W. and Koppelman, J.M., Earned value project management, Project Management Institute, 2010.
[3] Naeni, L.M., Shadrokh, S. and Salehipour, A., A fuzzy approach for the earned value management. International Journal of Project Management, 29(6), pp. 764-772, 2011. DOI: 10.1016/j.ijproman.2010.07.012.
[4] Vandevoorde, S. and Vanhoucke, M., A comparison of different project duration forecasting methods using earned value metrics. International Journal of Project Management, 24(4), pp. 289-302, 2006. DOI: 10.1016/j.ijproman.2005.10.004.
[5] Lipke, W., Zwikael, O., Henderson, K. and Anbari, F., Prediction of project outcome: the application of statistical methods to earned value management and earned schedule performance indexes. International Journal of Project Management, 27(4), pp. 400-407, 2009. DOI: 10.1016/j.ijproman.2008.02.009.
[6] Acebes, F., Pajares, J., Galán, J.M. and López-Paredes, A., Beyond earned value management: A graphical framework for integrated cost, schedule and risk monitoring. Procedia-Social and Behavioral

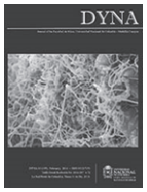
Sciences, 74, pp. 231-239, 2013. DOI: 10.1016/J.SBSPRO.2013.03.027.
[7] Cioffi, D.F., Designing project management: A scientific notation and an improved formalism for earned value calculations. International Journal of Project Management, 24(2), pp. 136-144, 2006. DOI: 10.1016/j.ijproman.2005.07.003.
[8] Plaza, M. and Turetken, O., A model-based DSS for integrating the impact of learning in project control. Decision Support Systems, 47(4), pp. 488-499, 2009. DOI: 10.1016/j.dss.2009.04.010.
[9] Von Wangenheim, C.G., Savi, R. and Borgatto, A.F., DELIVER! - An educational game for teaching earned value management in computing courses. Information and Software Technology, 54(3), pp. 286-298, 2012. DOI: 10.1016/j.infsof.2011.10.005.
[10] Elshaer, R., Impact of sensitivity information on the prediction of project's duration using earned schedule method. International Journal of Project Management, 31(4), pp. 579-588, 2013. DOI: 10.1016/j.ijproman.2012.10.006.
[11] Vanhoucke, M., On the dynamic use of project performance and schedule risk information during projecttracking. Omega, 39(4), pp. 416-426, 2011. DOI: 10.1016/j.omega.2010.09.006.
[12] Maravas, A. and Pantouvakis, J.-P., Project cash flow analysis in the presence of uncertainty in activity duration and cost. International Journal of Project Management, 30(3), pp. 374-384, 2012. DOI: 10.1016/j.ijproman.2011.08.005.
[13] Yao, J.-S., Chen, M.-S. and Lu, H.-F., A fuzzy stochastic single-period model for cash management. European Journal of Operational Research, 170(1), pp. 72-90, 2006. DOI: 10.1016/j.ejor.2004.06.017.
[14] Mortaji, S.T.H., Bagherpour, M. and Noori, S., Fuzzy earned value management using L-R fuzzy numbers. Journal of Intelligent and Fuzzy Systems, 24(2), pp. 323-332, 2013. DOI: 10.3233/IFS-2012-0556.
[15] Quiroga, J., Cartes, D. and Edrington, C., Neural network based system identification of a PMSM under load fluctuation. DYNA, 76(160), pp. 273-282, 2009.
[16] Pérez-Salazar, M.D.R., Mateo-Díaz, N.F., García-Rodríguez, R., Mar-Orozco, C.E. and Cruz-Rivero, L., A genetic algorithm to solve a three-echelon capacitated location problem for a distribution center within a solid waste management system in the northern region of Veracruz, Mexico. DYNA, 82(191), pp. 51-57, 2015. DOI: 10.15446/dyna.v82n191.51146.
[17] Agudelo, L., López-Lezama, J.M. and Galeano, N.M., Vulnerability assessment of power systems to intentional attacks using a specialized genetic algorithm. DYNA, 82(192), pp. 78-84, 2015. DOI: 10.15446/DYNA.V82N192.48578.
[18] Cheng, M.-Y., Tsai, H.-C. and Liu, C.-L., Artificial intelligence approaches to achieve strategic control over project cash flows. Automation in Construction, 18(4): pp. 386-393, 2009. DOI: 10.1016/j.autcon.2008.10.005.
[19] Cheng, M.-Y. and Roy, A.F., Evolutionary fuzzy decision model for cash flow prediction using time-dependent support vector machines. International Journal of Project Management, 29(1), pp. 56-65, 2011. DOI: 10.1016/j.ijproman.2010.01.004.
[20] Warburton, R.D., A time-dependent earned value model for software projects. International Journal of Project Management, 29(8), pp. 1082-1090, 2011. DOI: 10.1016/j.ijproman.2011.02.008.
[21] Jang, J.-S., ANFIS: Adaptive-network-based fuzzy inference system. IEEE Transactions on Systems, Man and Cybernetics, 23(3), pp. 665-685, 1993. DOI: 10.1109/21.256541.
[22] Iranmanesh, S.H., Mirseraji, G.H. and Shahmiri, S., An emotional learning based fuzzy inference system for improvement of the completion time of projects estimation Computers & Industrial Engineering, 2009. CIE 2009. International Conference on, pp. 470-475, 2009. DOI: 10.1109/ICCIE.2009.5223748
[23] Abarbanel, H.D., Brown, R., Sidorowich, J.J. and Tsimring, L.S., The analysis of observed chaotic data in physical systems. Reviews of Modern Physics, 65(4), pp. 1331, 1993. DOI: 10.1103/REVMODPHYS.65.1331.
[24] Jiang, J., Chen, J. and Qu, L., The application of correlation dimension in gearbox condition monitoring. Journal of Sound and Vibration, 223(4), pp. 529-541, 1999. DOI: 10.1006/jsvi.1998.2161.
[25] Kim, M.S. and Kong, S.G., Time series prediction using the parallel-structure fuzzy system. In Fuzzy Systems Conference Proceedings, 1999. FUZZ-IEEE'99, 1999. pp. 934-938.
[26] Ding, S., Model-based fault diagnosis techniques: Design schemes, algorithms and tools, Springer, 2013. DOI: 10.1007/978-1-4471-4799-2.
[27] Witczak, M., Modelling and estimation strategies for fault diagnosis of non-linear systems: from analytical to soft computing approaches, Springer, 2007.
[28] Kim, M., Lee, H., You, C. and Chung, C., Chaotic time series prediction using PSFS2. Proceedings of the 41st SICE Annual Conference, 2002. pp. 2358-2363.
[29] Kolisch, R. and Hartmann, S., Experimental investigation of heuristics for resource-constrained project scheduling: An update. European Journal of Operational Research, 174(1), pp. 23-37, 2006. DOI: 10.1016/j.ejor.2005.01.065.
[30] Iranmanesh, H., Mojir, N. and Kimiagari, S., A new formula to estimate at completion of a project's time to improve earned value management system. IEEE Conference on Industrial Engineering and Engineering Management, 2007. pp. 1014-1017.
[31] Mosavi, M.R., Comparing DGPS corrections prediction using neural network, fuzzy neural network and Kalman filter. Journal of GPS Solutions, 10(2), pp. 97-107, 2006. DOI: 10.1007/s10291-005-0011-7.

M.T. Hajali-Mohamad, received his BSc. and MSc. in Industrial Engineering at K. N. Toosi University of Technology, Tehran, Iran. He received his PhD. in Industrial Engineering in 2013, from the Iran University of Science and Technology (IUST), Tehran, Iran. His research interests include operations research, system dynamics, and EVM.
ORCID: org/0000-0003-1054-6982

M.R. Mosavi, received his BSc., MSc. and PhD. degrees in Electrical Engineering in 1997, 1998, and 2004, respectively from Iran University of Technology (IUST), Tehran, Iran. He is currently a faculty member (full professor) of the Department of Electrical Engineering of IUST. He is the author of 230 scientific publications in journals and international conferences. His research interests include circuits and systems design.
ORCID: org/0000-0002-2389-644x

K. Shahanaghi, is Assistant Professor (since 2004) at the school of Industrial Engineering, Iran University of Technology (IUST) where he lectures Simulation, Engineering Economy, and Uncertain Programming courses for Industrial Engineering postgraduate students. He received his BSc., MSc. and PhD. in Electrical Engineering at (IUST), Tehran, Iran in 2000. He has many papers in international journals and conferences. His research interests include Operations Research, Engineering Economy, Simulation, and Multicriteria Decision Making.
ORCID: org/0000-0003-0947-0773

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS



DYNA
http://dyna.medellin.unal.edu.co/

Fingerprint verification using computational geometry

Manuel Ramírez-Flores [a], Gualberto Aguilar-Torres [b] & Gina Gallegos-García [c]

[a] *Sección de Estudios de Posgrado e Investigación, Instituto Politécnico Nacional, Unidad Culhuacán, México. manuel300688@gmail.com*
[b] *Comisión Nacional de Seguridad, México D.F., México, autg79y@yahoo.com*
[c] *Sección de Estudios de Posgrado e Investigación, Instituto Politécnico Nacional, Unidad Culhuacán, México. ggallegosg@ipn.mx*

Received: October 17th, 2014. Received in revised form: August 12th, 2015. Accepted: January 10th, 2016.

Abstract
This paper presents a robust minutiae based method for fingerprint verification. The proposed method uses Delaunay Triangulation to represent minutiae as nodes of a connected graph composed of triangles. The minimum angle over all triangulations is maximized, which gives local stability to the constructed structures against rotation and translation variations. Geometric thresholds and minutiae data were used to characterize the triangulations created from input and template fingerprint images. The effectiveness of the proposed method is confirmed through calculations of false acceptance rate (FAR), false rejected rate (FRR) and equal error rate (EER) over FVC2002 databases compared to the results of other approaches.

Keywords: Angle of orientation, Delaunay Triangulation, Equal Error Rate, Fingerprint, Geometric Thresholds.

Verificación de huella dactilar utilizando geometría computacional

Resumen
Este trabajo presenta un método robusto con base en minucias para la verificación de huellas dactilares. El método propuesto utiliza Triangulaciones de Delaunay para representar a las minucias como nodos de un grafo compuesto por triángulos. El ángulo mínimo sobre todas las triangulaciones es maximizado, lo cual proporciona estabilidad local a las estructuras construidas contra variaciones de rotación y traslación. Umbrales geométricos y datos sobre minucias fueron utilizados para caracterizar las triangulaciones creadas con las imágenes de huellas dactilares de entrada y plantilla. La efectividad del método propuesto es confirmada con cálculos de la tasa de falsa aceptación (FAR), tasa de falso rechazo (FRR) y la tasa de igualdad de error (EER) sobre las bases de datos FVC2002, en comparación con los resultados de otras propuestas.

Palabras clave: Ángulo de orientación, Triangulaciones de Delaunay, Tasa de Igualdad de Error, Huella Dactilar, Umbrales Geométricos.

1. Introduction

Biometric technologies use some physical or psychological trait such as: fingerprint, face, iris, voice, etc. to verify or identify individuals and therefore restrict unauthorized access to computational systems. These traits are unique and inherent in individuals, making it difficult to falsify their I.D.; however, often they are noisy by nature [1].

Three types of degradations normally affect the quality of the fingerprint image: appearance of gaps between ridges, parallel ridges intercepts and natural effects such as cuts, wrinkles and injuries. Image enhancement processes are in charge of improving the contrast between ridges and valleys, and reducing the noise in the image [2]. Sometimes methods similar to medical image processing are also used, for example, via spatial domain filtering [3].

After enhancing the fingerprint image, comes the process of extracting and matching fingerprint features, which can be classified in three categories: based on minutiae [4-6], based on image [7-9] and hybrid [2,10,11]. Those based on minutiae, use a feature vector extracted from fingerprints as a set of points in a multi-dimensional plane. Some of the characteristics that the feature vector represents are: type of minutiae, position, and orientation, among others. After the extraction process, fingerprint matching becomes a non-rigid point-matching problem with unknown correspondence and differences in the number of points belonging to two sets (query and template). Moreover, skin elasticity changes the

relative position of the minutiae at each acquisition [6] and can cause genuine minutiae to get lost or pseudo minutiae to appear. [5]

The use of computational geometry can help to address image-processing problems that are difficult to understand and implement, such as distortion, rotation variations, and some others. One particular geometric structure that can resolve most of the shortcomings in the minutiae based methods for fingerprint verification processes is a Delaunay Triangulation.

For fingerprint verification processes, Delaunay Triangulation can be formed if the minutiae locations are taken as the set of points P. The advantages of this proposed method are that every minutia keeps the same neighboring structure even in the presence of distortion. Insertion of new points in the triangulations because of noise affects only locally. The same happens with missing or spurious minutiae. Thus, using geometrical thresholds, each triangle in the Delaunay Triangulation can be characterized uniquely. Because of this, each fingerprint can be classified and when a second sample of the same fingerprint requires to be verified, a quantitative comparison can be made. An Equal Error Rate of less than 1% can be obtained using these techniques, demonstrating the accuracy of the method under these thresholds.

1.2. Our contribution

The novelty of our contribution with respect to other approaches is in the full analysis of the Delaunay Triangulations to compare minutiae structures between fingerprints. Most minutia based approaches that use Delaunay Triangulations start their analysis by studying the edges that form Delaunay triangles instead of analyzing these triangles as a basic structure [4-6]. In some cases, even a threshold is established to decide whether to use Delaunay Triangulations or whether to process all possible combinations of minutiae forming an edge [4]. This kind of analysis involves a greater computational load because of the number of combinations to analyze and the operations required for each comparison.

On the other hand, if Delaunay Triangulations are used, most of the approaches presented until now perform a fingerprint alignment based on one reference triangle. To find this reference triangle an analysis of the triangle's edges is undertaken by using specific thresholds. Then, after the alignment, other thresholds have to be established to define a neighborhood where aligned minutiae should be. With this, several alignments and comparisons have to be made in order to count the matching aligned minutiae, calculate a matching score and find the best alignment [5,6].

In our proposal, from the beginning, every Delaunay triangle extracted from the query fingerprint is compared against every triangle extracted from the template fingerprint. In this comparison, the analysis of the three edges of the triangle and the three vertices of the triangle are performed. To avoid fingerprint alignment operations, a measurement and a comparison of the orientation of the triangles is performed. By establishing a specific threshold, the possible variation of fingerprint rotation is considered.

In the case of spatial displacement of the fingerprints, the structure of the Delaunay Triangulation is tolerant because if a general spatial displacement takes place, then all the minutiae move by the same proportion and the triangles extracted remain the same. If only some minutiae were displaced, then the length of the sides of some triangles changes, and the threshold for the comparison of the length of triangles covers that distortion. Finally, the matching score for the comparison of two fingerprints is pretty straightforward. The matching score is calculated as the rate of coincident triangles over the average of the total number of triangles between both fingerprints.

This paper is organized as following: Section 2 describes related work around the problem of fingerprint verification. Section 3 explains the mathematical and computational theory behind Delaunay Triangulations and the measurements calculated from this to characterize a fingerprint. Section 4 details our proposed solution and how our scheme was implemented, pointing out the tools used for the extraction of the fingerprint feature vector and the thresholds used in the matching process. Section 5 shows the results obtained after applying the proposed solution to 4 FVC2002 databases. Then a discussion of these results is presented. Finally, Section 6 contains the conclusions of this work in terms of the results obtained and the future work to improve the performance of the scheme proposed.

2. Related work

2.1. Image based approaches

Many studies have been undertaken to deal with the problem of fingerprint matching and many algorithms and schemes with different approaches have been proposed. Among image-based methods for fingerprint verification, there is a proposal based on features extracted from Wavelet and Fourier-Melling Transform (WFMT). Wavelet transform is used to preserve the local edges and reduce noise in the low frequency domain after image decomposition, which makes the fingerprint image less sensitive to shape distortion. Then Fourier-Melling transformation (FMT) served to produce a translation, rotation and scale invariant feature. The results obtained in [7], show that verification accuracy is 5.66 and EER is of 1.01%.

A second proposal in the category of image-based algorithms is the use of tessellated invariant moment features for fingerprint verification. In this scheme, a reference point is proposed to allow more accurate and quicker performance. In the scheme, all the intrinsic properties of the fingerprints are estimated, such as foreground region mask, local ridge orientation and local ridge frequency, to enhance fingerprint image. By using different complex filtering methods, a reference point is established. Then its orientation is calculated using the least mean square orientation algorithm. A Region of Interest (ROI) is centered on the reference point and tessellated in a predefined number of square cells. Seven invariant moments are extracted from the cells and these represent the fingerprint information of the local structure. The verification is based on measurements similar to the eigenvalue-weighted cosine (EWC) distance to match two corresponding feature vectors. The experiments carried out over the FVC2002 4 databases show an average Equal Error Rate (EER) of 3.57% using the EWC distance [8].

2.2. Hybrid approaches

One solution for hybrid fingerprint extraction and matching processes was proposed in [10] and uses both minutiae and ridge flow information. To capture the ridge strength at equal space orientations, a set of 8 Gabor filters in the spatial frequencies that correspond to the average inter-ridge spacing in fingerprints is used. Then an eight-dimensional ridge feature map is constructed with square tessellation of the filtered images. This map and the minutiae set of a fingerprint are used for matching purposes. For this scheme, the calculated EER is about 4%.

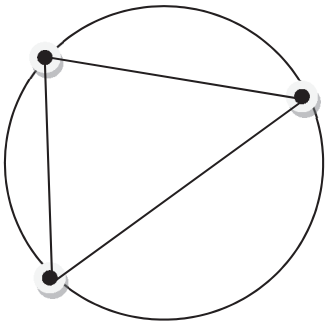


Figure 1. Triangle's circumcircle
Source: The authors.

2.3. Minutiae based approaches

Thus, regarding the proposals in the minutiae based category for fingerprint matching, there are few researches that utilize Delaunay Triangulations as a structure to compare a query and a template fingerprint image. In 2004, Parziale and Niel, proposed to establish the dependency among minutiae by applying Delaunay Triangulation over the point set representing them. In that structure, each minutia was used as a triangle's vertex. Then measures of distance between minutiae pairs, angular difference between orientations of minutiae pairs, and angles between the orientation of each minutia and the segment connecting them with another minutia, were calculated.

After applying three geometric filters, some triangulations in the query set were selected as candidates for matching with the triangles of the template set. For that, an alignment procedure was undertaken with the triangulations in the query set in terms of spatial coordinates and angles of orientation. If a minutia in the transformed query set is close enough to a minutia in the template set, it is counted and used later to calculate a matching score. [6]

In 2005, Liu, Yin, and Zhang, proposed a very similar fingerprint matching algorithm based on Delaunay Triangulations to find reference minutiae pairs known as RMPs. The analysis using Delaunay Triangulations begins with similar edge pairs formed from query and template sets of minutiae. Measurements of Euclidian distance, minutia orientation and edge orientation are then compared. If a pair of edges is very similar, then the triangles to which they belong become candidates for the next analysis phase. In the second phase, the distance of triangle's sides and internal angles are compared. If a coincidence between triangles exists, for each triangle, an alignment of the query set with the template set is carried out, using geometrical equations. Finally, for each alignment of points, these are counted and a matching score is calculated. [5]

The last proposal related to the use of Delaunay Triangulations, was made by Deng and Huo in 2005. Instead of finding the best-matching minutiae pairs, the objective was to find the best-matching edge pairs. Other important changes were: minutiae orientation was mapped to a range from 0 to 2Π instead of using the original from 0 to Π. The ridge count between minutiae was also used as data in the matching process.

The process of fingerprint matching starts by checking the number of minutiae in the fingerprint image, if it is below a threshold, Delaunay Triangulations are not calculated and instead all the possible edges connecting two minutiae are considered. Then, several geometric filters are applied to the edge pairs, like Euclidean distance, minutiae orientation difference, among others. Only the minutiae of the edges that satisfy those filters are used in the next phase. The remaining minutiae are sorted in ascending order and form all the possible triangles with the closest neighboring minutiae. Again, geometric comparisons between triangles are performed. If their characteristics satisfy the thresholds established, a matching score is calculated for each triangle. At the end, another matching score between the query and template image is calculated considering the triangle matching scores previously calculated. [4]

In later sections of this paper, some of these different schemes will be looked at again in a comparison against our proposed minutiae-based solution. Following that, we present a discussion about the obtained results in terms of the EER, FAR and FRR thresholds.

3. Background

A Triangulation can be defined, as the maximal planar subdivision whose vertex set is P, where P denotes a finite set of points in a plane. A maximal planar subdivision is a subdivision in which no edge connecting two vertices can be added to it without destroying its planarity (any edge that is not in the subdivision intersects one of the existing edges).

"Let P be a set of points in the plane, and let T be a triangulation of P. Then T is a Delaunay triangulation of P if and only if the circumcircle of any triangle of T does not contain a point of P in its interior."[12]

So any Delaunay Triangulation T maximizes the minimum angle over all triangles that compose it. What is stated in the above quote can be appreciated in Fig. 1.

To start building a Delaunay Triangulation with only the set of points P as initial data, an initial repository has to be created. From a geometrical point of view, the repository is the first triangle in the Delaunay Triangulation and it is large enough to contain the whole set of points P. The vertices of the first triangle are: three extra points, p0, p-1 and p-2. Fig. 2 represents this statement.

It is important to choose p_0, p_{-1} and p_{-2} far enough away, so they do not destroy any triangles in the Delaunay triangulation of P. Later p_0, p_{-1} and p_{-2} can be discarded together with all their incident edges.

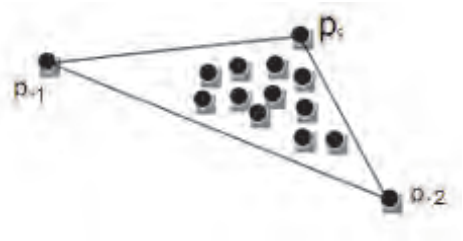


Figure 2. Delaunay Triangulation start
Source: The authors.

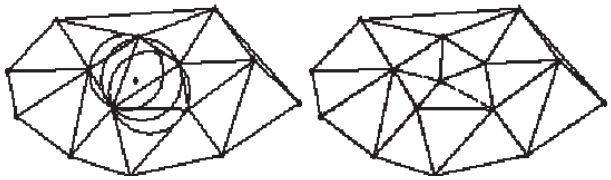


Figure 3. Addition of a point to A Bowyer-Watson triangulation.
Source: [13]

There are basically two types of existing algorithms, which can be used to implement a Delaunay Triangulation: The first type is a static algorithm where the triangulation is valid after every single point is processed. Some examples are: The recursive split algorithm, the divide and conquer algorithm, the step-by-step algorithm, the modified hierarchical algorithm, among others.

The second type of algorithms is the dynamic triangulation where the triangulation is valid during processing. This makes it possible to view the contribution of one point to the triangulated irregular network (TIN). Some algorithms to implement this are called incremental as the Bowyer and Watson algorithm.

Bowyer and Watson algorithm is known as an incremental delete and build algorithm because it adds points sequentially into an existing Delaunay triangulation [13]. The process follows the steps below:

For each point in set P:

1. Insert point p ϵ P into triangulation
2. Find all existing triangles whose circumscribing circle contains the point p.
3. All triangles found in step 2 are deleted and a convex cavity is created.
4. The point p is joined with all the vertices on the boundary of the cavity formed in 3 (re-triangulation).

Remove initial triangle.

An example of the beginning and the end steps of the process is shown in Fig. 3.

Finally the implementation of the Bowyer-Watson algorithm requires computational data structures such as: point based data structure for the vertices of the triangle, a triangle based data structure for the elements that compose the Delaunay Triangulation and finally a directed acyclic graph that represents the Triangulation with the edges of each triangle and the neighbors that share them.

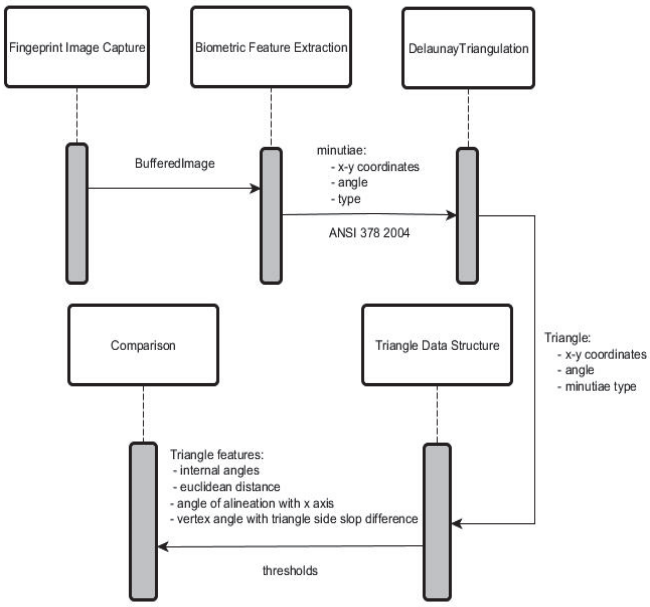


Figure 4. Sequence Diagram for our Fingerprint Verification method
Source: The authors.

For the implementation of Delaunay Triangulations in our scheme, the Bowyer and Watson algorithm was used because it provides the same theoretical optimum algorithmic complexity as other methods, which is $\Theta(N \log_2 N)$, but with an easier procedure.

Although Delaunay Triangulations allow the creation of a rotation and distortion tolerant structure, the characterization, identification and comparison between two different triangulations must be undertaken manually. Therefore, the second part of our proposal consists in a set of measures and geometric thresholds that allow to distinguish each triangle and its vertices in the Delaunay triangulations formed.

4. Proposed method

The proposed scheme is a minutia based fingerprint verification system that uses Delaunay Triangulations with geometrical measures and thresholds to validate the similarity between two different fingerprints. The complete proposed scheme is shown in Fig. 4., and explained right after.

Stage 1: Fingerprint image capture.

a. Use of Griaule Biometrics Fingerprint SDK 2009

Stage 2: Minutiae features extraction.

a. Creation of feature extraction vector in ANSI 278-2004 format.

Stage 3: Delaunay Triangulation creation.

a. Creation of reference triangle.
b. Insertion of a new point.
c. Search of the triangle containing the new point inserted.
   c.3 Check whether point is already inserted.
   c.4 Cross product calculation between vertices of triangle and point inserted.
d. Determine the Cavity produced by the new point in the Delaunay Triangulation.

d.1 Verify whether the new point is inside, on, or outside the Triangle's circumcircle and its neighbors' circumcircles.
d.2 If the new point is inside a circumcircle of some triangle, add it to cavity list.

e. Update Delaunay Triangulation
e.1 Remove triangles contained in the cavity calculated from Delaunay Triangulation.
e.2 Create of new triangles with remaining vertices around the old cavity and the new point.
e.3 Add new Triangles to Delaunay Triangulation.
e.4 Update each link between adjacent triangles.

Stage 4: Triangle Data Structure creation.

a. Triangle's internal angles calculation.
b. Euclidean distance between triangle's vertices calculation.
c. Triangle's angle of alienation with x-axis.
d. Difference between each pair of vertex's angle and triangle side slope.

Stage 5: Triangle features comparison

a. Alignment of triangles based on internal angles.
b. Arithmetic comparison of triangle's vertices based on minutiae type *
c. Arithmetic comparison between the length of the sides of the triangles *
d. Arithmetic comparison of differences between each pair of vertex angle and slope of a triangle side*
e. Arithmetic comparison between angles of alignment with x axis. *
f. Authentication or rejection of fingerprint.

*In each comparison, the corresponding threshold for error tolerance is added.

In Stage 2, the feature vector extracted using the Griaule Biometrics software for each minutia, contains the following biometric data:

$$m_i = \{x, y, \theta, \tau\} \quad (1)$$

where:

- m_i = ith minutia
- x = value of the spatial coordinate in the x-axis
- y = value of the spatial coordinate in the y-axis.
- θ = minutia orientation [0,180°]
- τ = type of minutia [end of ridge, ridge bifurcation, other]

For stage 4, different measures are calculated in order to uniquely characterize each triangle. An image interpretation of those calculations can be observed in Fig. 5.

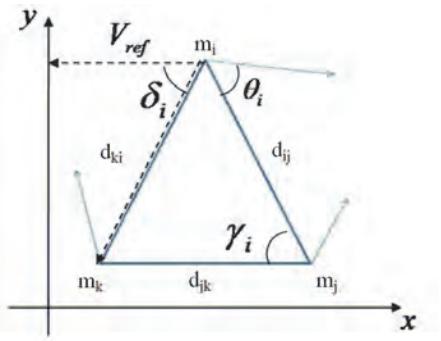


Figure 5. Angles used to characterize triangles
Source: The authors.

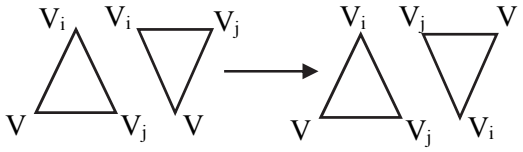


Figure 6. Alignment of triangles based on internal angles
Source: The authors.

The first measure, related to alignment of triangles can be resumed in Equation (2), which comes from the cosine law for triangles.

$$\gamma_i = \cos^{-1}\left(\frac{d_{ij}^2 + d_{jk}^2 - d_{ki}^2}{2d_{ij}d_{jk}}\right) \quad (2)$$

where:

- γ_i = triangle's internal angle
- d_{ij}, d_{jk}, d_{ki} = length of triangle's side

An alignment of triangles is needed to ensure that the order in which the vertices of two triangles in different Delaunay triangulations are described is correct. Otherwise, any further calculation will be meaningless because the vertices of the triangles could be swapped. To achieve this, the internal angles of a pair of triangles are paired so that the difference between them is a small as possible.

When a pair of angles has the smallest difference, the vertex of the triangle containing one of the angles is renamed so that it matches with the vertex's name in the other triangle. The process is repeated until the three vertices are matched with the best option in the second triangle. An example is shown in Fig. 6.

Once a pair of triangles is aligned, a process of comparison between the pair of triangles takes place. The first test checks the type of minutiae in the vertices with the same name in the different triangles. The main types of minutiae are classified as: termination and bifurcation.

If the first test is passed, the difference of one triangle's vertex angle (minutia orientation) and the angle of the segment connecting that vertex with another one is calculated, for each of the vertices of a triangle. Equation (3) describes how to calculate this measure. A threshold expressed in grades is established to allow a tolerance limit in the rotation that a triangle can have over another.

$$a_{ij} = \tan^{-1}\left(\frac{y_i - y_j}{x_i - x_j}\right) - \theta_i \quad (3)$$

where:

- a_{ij} = angle difference between triangle's side slope expressed as an angle and minutia orientation.
- y_i, y_j = y-spatial coordinate of minutiae *i* and *j* respectively.
- x_i, x_j = x-spatial coordinate of minutiae *i* and *j* respectively.
- θ_i = i's minutia orientation

The next comparison is undertaken using the Euclidean distance equation (4), between the vertices of a triangle. Again a threshold is established, this time in pixels, to tolerate a certain

degree of distortion in the shape of the triangle because of the different spatial allocation of the vertices.

$$d_{ij} = \sqrt{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2} \quad (4)$$

where:

- d_{ij} = Euclidean distance between minutiae *i* and *j*
- x_i, x_j = x-spatial coordinate of minutia *i* and *j*
- y_i,y_j = y-spatial coordinate of minutia *i* and *j*

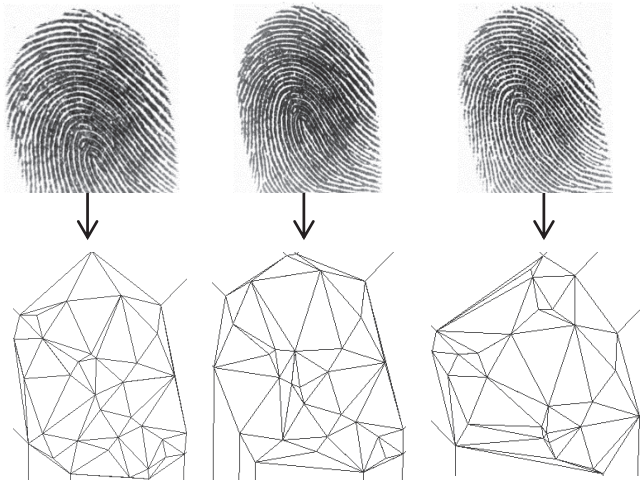
The final filter is related to Equation (5), which evaluates the angle between the first vertex of a triangle and the x-axis. A threshold expressed in pixels allows a tolerance degree of rotation for the triangle being evaluated in case the fingerprints captured are rotated. It is important to point out that this measure quantifies the degree of rotation that a local area of the fingerprint has in terms of triangles that belong to a Delaunay Triangulation.

With this, an alignment of all minutiae in the fingerprint is no longer needed, given that if there is a rotation of the fingerprint, every minutia and every triangle composed of them will present the variation and it will fit in the threshold established.

$$\delta_i = \cos^{-1}\left(\frac{V_{ref} \cdot d_{ki}}{|V_{ref}||d_{ki}|}\right) \quad (5)$$

where:

- δi = angle between one triangle's side ij and Vref which is a vector parallel to x-axis
- Vref = vector from triangle's minutia mi to y-axis, which is parallel to x-axis
- dki = euclidean distance between minutiae k and i



a) b) c)

Figure 7. Three samples of the same person and their Delaunay Triangles.
Source: The authors.

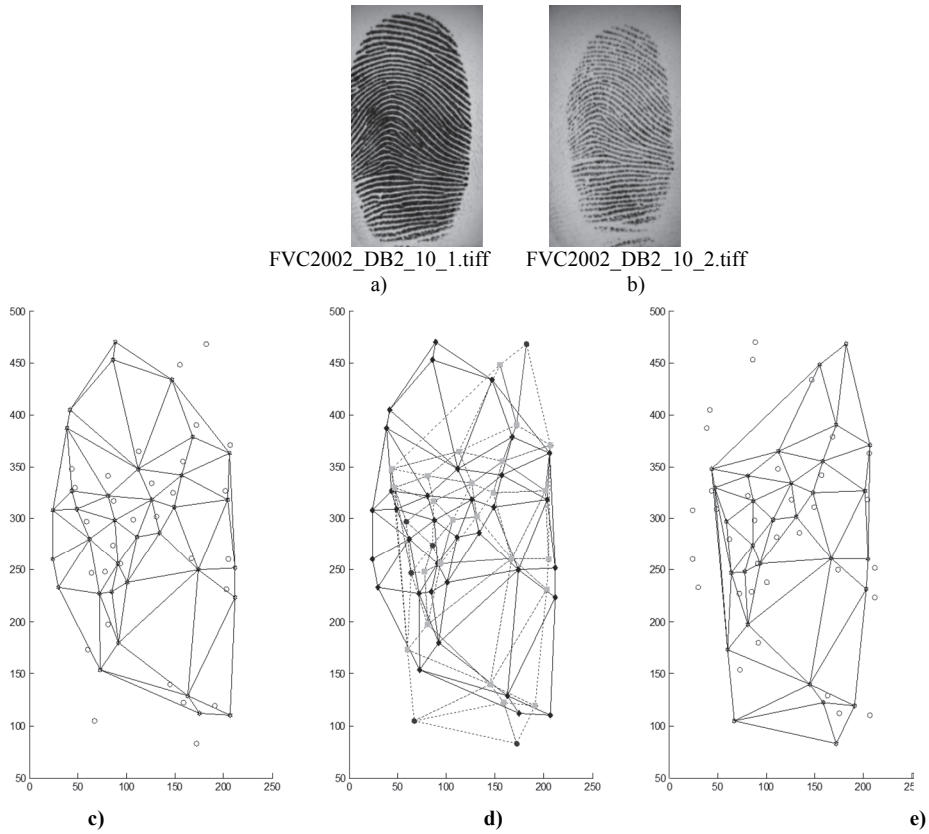


Figure 8. Delaunay Triangulation of fingerprints with small distortion
Source: The authors.

In the end, if each of the previously mentioned tests are passed successfully then the system recognizes both triangles as the same one and increases the count of equal elements between both of the fingerprint's Delaunay Triangulations.

Equal Error Rates below 1% can be obtained when using these techniques, which is a very good indicator for biometric verification systems.

5. Experiment

Before the experimental section, we must take into account that the fingerprint image obtained always changes due the pressure, angle and the tilt on capture device. For this reason, even if it is the same person, it is almost impossible for the same minutiae to appear in each one of the samples captured. This is one of the factors that mean that fingerprint recognition may present a high percentage of false rejection.

Because of the above, Delaunay Triangulations of the same fingerprint are also different each time a sample is captured, as shown in the Fig. 7.

During the experimental section, different tests were performed and it was found that it is almost impossible to have many similar Delaunay Triangulations in samples of two different fingerprints; however, in two samples of the same fingerprint, we found many similar Delaunay Triangulations. On average, when we have two samples of different fingerprints, we have found a maximum of two Delaunay Triangulations, but when we have two samples of the same fingerprint, we have found more than ten Delaunay Triangulations on average.

An image proof that describes how Delaunay Triangulations act as an effective filter for fingerprint images can be seen in the following 2 scenarios, represented with different graphs.

The first scenario shows two samples from the same fingerprint, whereby a change in pressure levels and a small rotation were registered. A group of images with the Delaunay Triangulations was processed. The results are presented in Fig. 8.

In Fig. 8, we can see that Delaunay Triangulations differ from one another but also that there are specific areas in the Triangulations that contain some triangles with the same characteristics in both image c) and e). In image d) some of the vertices that appear in both triangulations are marked in a softer color. This proves that Delaunay Triangulations can characterize fingerprints even if they differ because of displacement, distorsion, rotations, changes of pressure or other types of alterations.

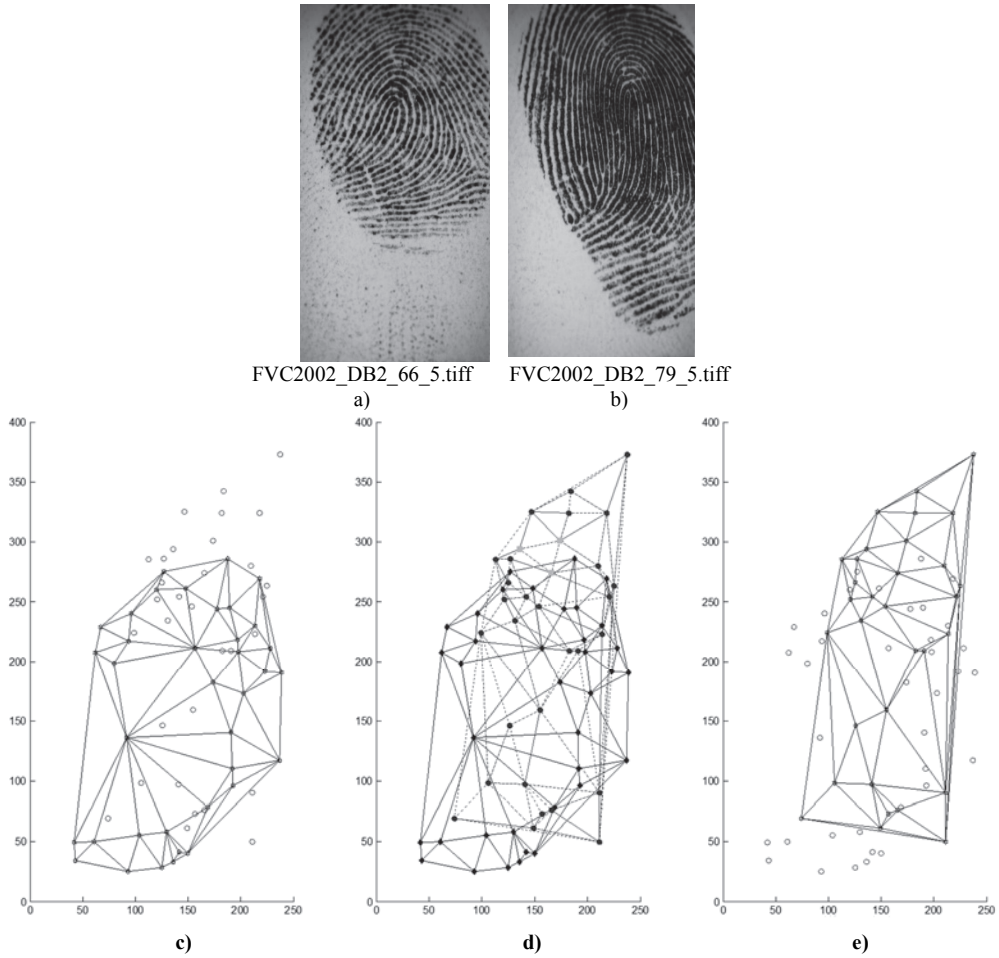


Figure 9. Delaunay Triangulation of two different fingerprints
Source: The authors.

Table 1. FVC2002 Databases characteristics

	Sensor Type	Model	Image Size	Set A (wide x deep)	Resolution
DB1	Low Cost Optical Sensor	Touch View II	300 x 300	100 x 8	500 dpi
DB2	Low Cost Capacitive Sensor	FX2000	256 x 364	100 x 8	569 dpi
DB3	Optical Sensor	100 SC	448 x 478	100 x 8	500 dpi
DB4	Synthetic Generator	SFinGe v2.51	240 x 320	100 x 8	about 500 dpi

Source: [14]

The second scenario shows two samples of two different fingerprints. Because of this, it is expected that no triangle in both Delaunay Triangulations calculated would be classified as equal. This is the most difficult task, because small triangles tend to be very similar in dimensions and angles with respect to others. Fig. 9 shows this.

From the images in Fig. 9 it can be seen that the calculated Delaunay Triangulations are totally different. Because of this, only one triangle is common between both triangulations. It is not easy to find the small triangle marked in a softer color in image d). in the other two images. But it is clear that the triangle marked as equal is very small and easy to classify wrongly.

It is very important to point out that for the experiments shown above, the different geometric thresholds to filter all the triangles in the Delaunay Triangulations had already been taken into account. Otherwise, if only a check up of the triangles present were carried out, almost none of the triangles would be equal to any other in the second triangulation. That can be seen easily in the images.

After testing the advantages of using our proposed scheme with some real fingerprint images and effectively characterize their minutiae, we amplified the number of tests in order to get the FAR and FRR thresholds to compare our scheme with other proposals. To obtain such thresholds, we considered 4 fingerprint databases from FVC2002 set A, which are summarized in Table 1.

The threshold depending fraction of the falsely accepted patterns divided by the number of all impostor patterns is called False Acceptance Rate (FAR). Its value is one, if all impostor patterns are falsely accepted and zero, if none of the impostor patterns are accepted. The fraction of the number of rejected client patterns divided by the total number of client patterns is called False Rejection Rate (FRR).

Note that if the score distributions overlap, the FAR and FRR intersect at a certain point. The value of the FAR and the FRR at this point, which is of course the same for both of them, is called the Equal Error Rate (EER).

The EER of a system can be used to give a threshold independent performance measure. The lower the EER, the better the system's performance, as the total error rate which is the sum of the FAR and the FRR at the point of the EER decreases [15].

For each fingerprint, 6 random images out of 8 were taken for training, and the remaining 2 were used for verification test purposes. This means that in each database, there were 600 training images and 200 test images. The selection was repeated 4 times with different test images each time. Then, FAR and FRR thresholds for each database were plotted in an image of "verification percentage vs. triangle matching percentage." Figs. 10 to 13 show each of these images.

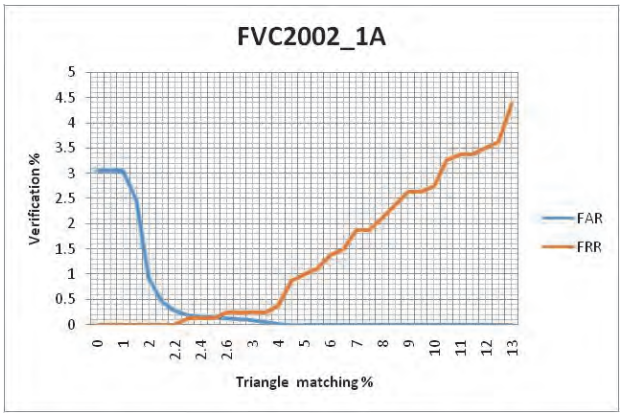


Figure. 10 FVC2A DB1 FAR and FRR thresholds
Source: The authors.

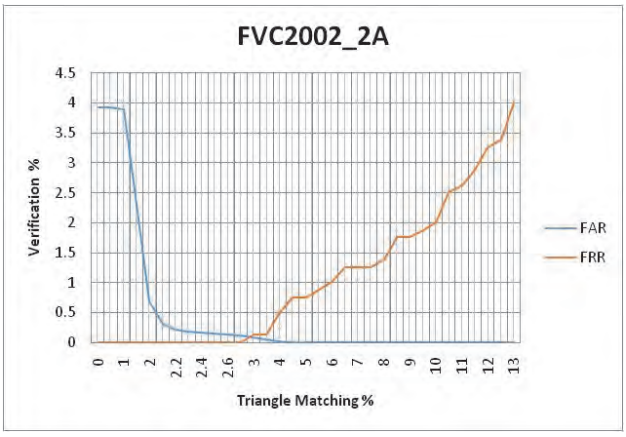


Figure. 11 FVC2A DB2 FAR and FRR thresholds
Source: The authors.

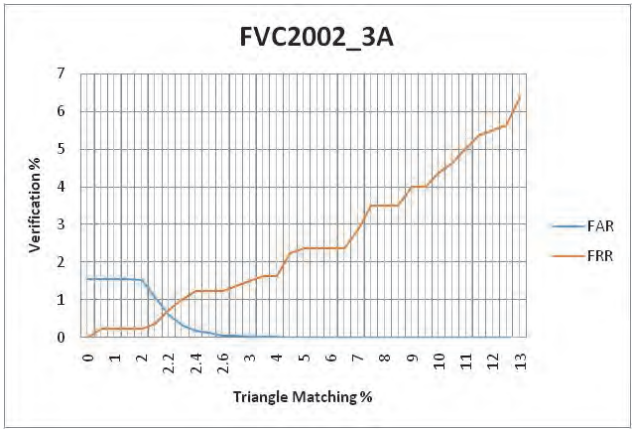


Figure. 12 FVC2A DB3 FAR and FRR thresholds
Source: The authors.

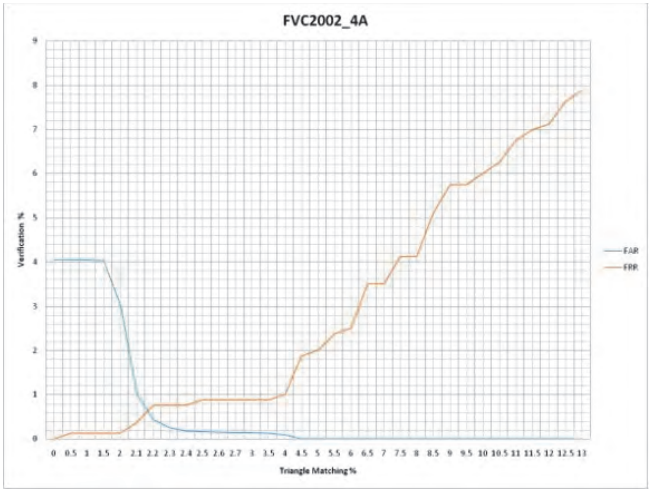


Figure. 13 FVC2A DB4 FAR and FRR thresholds
Source: The authors.

6. Results

After analyzing the images presented, the ERR was calculated and compared to another 3 proposed schemes. Table 2 shows the summary of the comparison.

As can be seen, the performance of the proposed method exceeds that of those presented in the related work of this article. The difference between the average percentages is about 7 times smaller in our proposal. The main reasons for this happening are related to the type of post processing stage in which our scheme takes place.

Instead of having a dispersion of points in an Cartesian plane, the analysis of minutiae takes place over a polygonal network that minimizes distortion and displacement of the vertices that compose them after image preprocessing and enhancement have been carried out. Also, the criteria established to decide whether a minutia in a fingerprint corresponds to another minutiae in a second fingerprint, strongly depends on a second set of measures from that polygonal network.

Other variables that play an important role in the identity verification results are the geometric thresholds, used to calibrate the training in the verification process. Depending on how strict the verification is required to be and the fingerprint images characteristics (size and resolution), the thresholds can be reduced to allow minimum or maximum variation among the polygonal networks. For the tests presented in this paper, a tolerance of 10 pixels in distance and 10 grades in angles were established.

If there were no alterations because of noise in the media and if the fingerprints do not suffer from erosion or aging, then they must remain equal. But because this is not possible, those thresholds were established based on the premise that different acquisitions from the same fingerprint even on a noisy media have the same degree of alteration in each of its minutiae such as the same degree of rotation, displacement, among others.

Table 2. EER from FVC2002 databases

Method	DB_1A	DB_2A	DB_3A	DB4_A	Average
Yang et al. [8]	1.63	3.78	4.20	4.68	3.57
Beng, et al. [7]	2.43	4.41	5.18	6.62	4.66
Ross et al. [10]	1.87	3.98	4.64	6.21	4.17
Deng. et al [4]	1.82	1.52	4.94	2.29	2.64
Our scheme	0.125	0.125	0.75	0.75	0.44

Source: Adapted from [2].

7. Conclusions

In this paper, a minutiae based fingerprint verification method was proposed. The innovation of this method relies on the use of Delaunay Triangulations and geometric thresholds to align fingerprint minutiae even in the presence of noise in the environment. Measures of distance, angles of orientation and angles of rotation make it possible to characterize geometric figures with its vertices composed of minutiae avoiding processes of minutiae alignment and matching score calculations. In this way, rotation and displacement tolerance is achieved within the verification process.

It is important to emphasize that a calibration of geometric thresholds must be carried out depending on the characteristics of the fingerprint images, such as size, resolution, etc. With all these conditions achieved, a better EER can be obtained in the tests realized.

Finally, the numerical values and graphs obtained in the tests, confirm that using Delaunay triangulation allows a strict discrimination of minutiae in a fingerprint. However, because the fingerprint images obtained always changes due to pressure, angle and the tilt on capture device, we need to store multiple biometric samples with different conditions to improve the scores. For example, for this paper, 6 images were necessary for each fingerprint..

Acknowledgements

This paper was supported by the CONACyT and the IPN.

References

[1] Wencheng, Y., Jiankun, H. and Wang, S., A delaunay triangle-based fuzzy extractor for fingerprint authentication, Trust, Security and Privacy in Computing and Communications (TrustCom), 2012 IEEE 11th International Conference on, 66(70), pp. 25-27, 2012. DOI: 10.1109/TrustCom.2012.23

[2] Khalil, M., Mohamad, D., Khurram, K. and Al-Nuzaili, Q., Fingerprint verification using statistical descriptors, Digital Signal Processing 20. ElSevier.Inc. pp.1264-1273, 2009. DOI:10.1016/j.dsp.2009.12.002

[3] Vianney, J., Rosales, A., Gallegos, F. and Arellano, A., Computer-aided diagnosis of brain tumors using image enhancement and fuzzy logic, DYNA, 81(183), pp 148-157, 2014. DOI: 10.15446/dyna.v81n183.36838

[4] Deng, H. and Huo, Q., Minutiae matching based fingerprint verification using delaunay triangulation and aligned-edge-guided triangle matching. Audio and Video Based Biometric Person Authentication. AVBPA, pp. 270-278, 2005. DOI: 10.1007/11527923_28

[5] Liu, N., Yin, Y. and Zhang, H., A fingerprint matching algorithm based on delaunay triangulation net, Proceedings, International Conference on Computer and Information Technology, CIT05. pp. 591, 595, 2005. DOI: 10.1109/CIT.2005.9

[6] Parziale, G. and Niel, A., A fingerprint matching using minutiae triangulation. ICBA2004, LNCS3072, Springer-Verlag Berlin Heidelberg. pp.241-248, 2004. DOI: 10.1007/978-3-540-25948-0_34

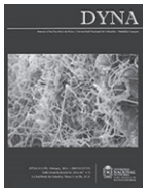
[7] Beng, J., Chek-Ling, D. and Song, O., An efficient fingerprint verification system using integrated wavelet and Fourier-Melling invariant transform, Image and Vision Computing. ElSevier, 22, pp. 503-513, 2003. DOI: 10.1016/j.imavis.2003.12.002
[8] Yang, J. and Park D., A fingerprint verification algorithm using tessellated invariant moment features, Neurocomputing. ElSevier, 7, pp. 1264-1273, 2008. DOI: 10.1016/j.neucom.2007.12.034
[9] Imamverdiyev, Y., Beng, A. and Kim, J., Biometric cryptosystem based on discretized fingerprint texture descriptors, Expert Systems with Applications. 40, pp. 1888-1901, 2012. DOI: 10.1016/j.eswa.2012.10.009
[10] Ross, A., Jain, A. and Reisman, J., A hybrid fingerprint matcher, Pattern Recognition. The Journal of the Pattern Recognition Society. 36, pp. 1661-1673, 2003. DOI: 10.1016/S0031-3203(02)00349-7
[11] Abraham, J., Kwan, P. and Gao, J.. Fingerprint matching using a hybrid shape and orientation descriptor, In: Yang J. (Ed.), State of the art in Biometrics INTECH, ISBN: 978-953-307-489-4, 2011, pp. 25-56, DOI: 10.5772/19105
[12] Berg, M., Cheong, O., van Kreveld, M. and Overmars, M., Computational geometry. Algorithms and applications, Berlin: Springer, 2008. DOI: 10.1007/978-3-540-77974-2
[13] Arens, C., The Bowyer-Watson algorithm. An efficient implementation in a database environment, Thesis, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, Holand, 2002, pp 5-9.
[14] Cappelli, R., Maio, D., Maltoni, D., Wayman, J.L. and Jain, A.K., Performance evaluation of fingerprint verification systems, IEEE Trans. Pattern Analysis and Machine Intelligence. 28(1), pp. 3-18, 2006. DOI: 10.1109/TPAMI.2006.20
[15] Technical Document About FAR, FRR and EER, Version 1.0, SYRIS Rechnology Corp., 2004, pp 1-4.

M. Ramírez-Flores, received the BSc. in Engineer in Telecommunications and Electronic Systems in 2010, with honorific mention from the Instituto Tecnológico y de Estudios Superiores de Monterrey, Campus Ciudad de México, Mexico. His areas of knowledge are: electronic voting process, secure cryptoimage applications and biometry.
ORCID: 0000-0001-5085-9604

G. Aguilar-Torres, has a PhD. in Communications and Electronics. His areas of knowledge are: signal processing, pattern recognition, neural networks and biometry.
ORCID: 0000-0002-1808-3962

G.Gallegos-García, has a PhD. in Communications and Electronics. Her areas of knowledge are: e-voting, design of secure cryptoimage applications, information systems and cryptography.
ORCID: 0000-0002-5212-350X

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS



DYNA
http://dyna.medellin.unal.edu.co/

Managing relationships and communications in higher education efficiently through digital social networks: The importance of the relational coordination model

Alexander Lacayo-Mendoza [a] & Carmen de Pablos-Heredero [b]

[a] *Universidad Rey Juan Carlos (URJC), Madrid, España. alexanderlacayo@yahoo.com*
[b] *Universidad Rey Juan Carlos (URJC), Madrid, España, Carmen.depablos@urjc.es*

Received: Febrero 22th, 2015. Received in revised form: November 12th, 2015. Accepted: January 10th, 2016.

Abstract
Digital social networks have proven to be of great support for organizations that are increasingly using new forms of social communication every day, seeking to improve their productivity and competitiveness. The main objective of this study is to explain how organizations, particularly Institutions of Higher Education, can improve their competitive position through the use of digital social networks from the perspective of relationship management and management of communications, in order to achieve better results in terms of teamwork satisfaction, relationship marketing and educational excellence. For this objective, the application of a relational coordination model, based on the sharing of objectives, optimal knowledge management, mutual respect and efficient communication mechanisms, is suggested.

Keywords: relational coordination, organizational practices, organizational results, university education, digital social networks.

Cómo gestionar las relaciones y comunicaciones de manera eficiente a través de las redes sociales digitales en instituciones de educación superior: Una propuesta desde el modelo de coordinación relacional

Resumen
Las redes sociales digitales han demostrado ser un gran apoyo para las organizaciones, que cada día las utilizan más en las nuevas formas de comunicación social en busca de mejorar su productividad y competitividad. Como objetivo principal de este estudio, se busca explicar cómo a través de la utilización de redes sociales digitales se puede mejorar la posición competitiva de las organizaciones, y en particular de las Instituciones de Educación Superior, desde la perspectiva de gestión de las relaciones y gestión de las comunicaciones, para conseguir mejores resultados en términos de satisfacción de los equipos, de marketing relacional y excelencia educativa. Para esto, se ha sugerido la aplicación del modelo de coordinación relacional que tiene sus bases en la compartición de objetivos, la óptima gestión del conocimiento, y el respeto mutuo unido a mecanismos eficientes de comunicación.

Palabras clave: coordinación relacional, prácticas organizativas, resultados organizacionales, educación universitaria, redes sociales digitales.

1. Introduction

In today's globalized world, production, coding and dissemination of information and knowledge have become the main pillars for technological innovation and economic growth. Simultaneously, the worldwide dissemination of information and communications technology (ICT) has launched a process of connectivity, unprecedented in human history, which has increased the possibility of generating and socializing information and knowledge by progressively removing barriers of time and space. Perhaps the most outstanding phenomenon of these changes is the Internet, a network of networks that has

had a substantial impact on economic processes, culture and the generation of new forms of interaction, communication and exchange of experiences between different actors, institutions and social movements [1].

The way companies interact has changed due to the emergence of information and communication technologies (ICTs). Since the first implementation of ICTs as marketing strategies in companies, the evolution and development of these trends in different fields have been remarkable. Everyday people tend to be more connected in order to perform most of their online activities and spend more time surfing the Internet, which shows the exponential growth in the use of ICT [2].

Currently, most organizations provide strong support to the development of information and communication technology. ICTs are fundamental not only for their instrumental character in meeting information needs for decision making, but also because they have become a vehicle for generating distinctive capabilities, which, when combined with other valuable resources, can result in a constitutive source of competitive advantage [3]. Therefore, the application of ICT must be aligned with organizational strategic planning, and as such becomes a strategic factor behind the growth of the organization [4].

With the emergence of the Internet, society has radically changed the way people communicate, their leisure habits, and their ways of meeting and relating to each other. Thanks to the technologies derived from the second Internet generation, known as Web 2.0, the consumer has become the main figure of the network. The label 2.0 represents a more collaborative Web that allows users to access and participate in the creation of unlimited knowledge, and as a result of this interaction, new business opportunities are created for businesses [5].

Digital social networks are configured to be the best tools of Web 2.0 for business strategies focusing on customer orientation, and on customizing the message, interaction with the recipient and maintenance of two-way communication with the client in order to achieve their loyalty [6].

Through digital social networks, company employees may get to know each other, fostering a better relationship. Connecting people creates a greater degree of trust and responsibility between them [7]. Therefore, digital social networks are a means of communication, with the distinctive feature of enabling direct and quick relationships to be formed amongst a community of a few or millions of users and where each person can also make a contribution. Digital social networks are a powerful tool that, if properly used, can allow us to highlight our strengths, hide our faults and, in both cases, offer us a better understanding of our environment that allows us improve relationships at higher levels of exchange value [8].

The growth of digital social networks coupled with the ability to target consumers, making strategic use of demographic information such as gender, ethnicity, education and political or religious affiliations, which have been provided by participants through their profiles, make this tool really attractive for companies [9].

While users of digital social networks seek to connect with friends, organizations have taken the opportunity to communicate through specific demographics. Digital social networks offer organizations several key advantages over regular Web sites, because they facilitate communication with users based on location, age and marital status using the information provided in their profiles. Organizations can also communicate with potential customers through different social networks [10].

Digital social networks are often managed from the area of marketing, facilitating communication and the exchange of information with clients, customers, prospects, suppliers, investors, candidates for vacant posts, competitors and any other relevant actors in the life of the company. Therefore they become useful tools for business and professionals; by enabling practices such as the use of phone, email or a corporate website. However, social networks add an extra feature in comparison with the rest of these media: "*the ability to create community*" [8].

In order to improve their online presence, higher education institutions have increased the use of digital social networks as a marketing strategy that will help them achieve better positioning and organizational growth [11]. A topic of interest in the field of educational organizations, since 2011, has been the use of digital social networks in order to analyze performance and as a way of building relationships and ensuring student loyalty and participation [12].

For effective communication with students, professionals in higher education must adopt new technologies and more importantly, explore opportunities to implement a social media marketing plan, which should be reevaluated to adjust to shifting trends and be adapted to the changing needs of students. Facebook, Twitter and YouTube are the preferred media for socializing and networking [13].

The quality of communication is a key factor in achieving effective organizational relationships by building relationships that enable higher education organizations fulfill their corporate mission [14].

However, communication is not the only factor to measure teamwork quality. This can be measured by many different factors, such as those raised in the relational coordination model proposed by Gittell [15], which has been applied in this research as a framework for analyzing the use of digital social networks as tools for achieving better results in the field of relational marketing, in terms of teams' satisfaction and educational excellence.

This model has been applied to different types of organizations, with promising results, and it explains that relational coordination occurs through relationships of shared objectives; shared knowledge; mutual respect; and frequent, timely, accurate and solution focused communication. Thus it is achieved by managing relationships in an efficient manner and ensuring the achievement of organizational goals [15,16].

The main objective of this paper is to describe how the use of digital social networks from a model-based approach of relational coordination can be applied to better explain educational excellence.

2. Using digital social networks in the education sector

It is increasingly evident that digital social networks are being used in different sectors to achieve their marketing

goals and the education sector is no exception. Institutions of higher education make use of this tool to meet their marketing goals but also for recruitment, admissions, interacting with students and stakeholders, fundraising, student-teacher relationships, and primarily to improve their competitiveness in the sector [17].

Digital social networks can be useful tools in recruiting efforts and very beneficial to improve the competitiveness of higher education institutions [18], disseminating humanizing stories of students and graduates, which promote student retention and help achieve marketing goals [19].

The president of Southwestern College recognizes that digital social networks have become fundamental and strategic tools for institutions of higher education for the purpose of recruitment, development, and public relations, to build relationships with students and graduates, and finally to improve their competitiveness, moving these institutions into a leadership position [20].

According to the American company Digital ZOG [21], dedicated to research on social media and new technologies, one of the benefits of the evolution of digital social networks is the ability to reach a lot of people. Educational institutions can reach not only potential students but also current students, alumni, colleagues in the industry and the community in general; position achieving a good reputation of the institution through the integration of digital social networks within their strategic planning. As previously mentioned, the use of digital social networks contributes to improving the competitiveness of higher education institutions.

According to Kaplan and Haenlein [22], although many colleges still only make use of traditional media, a noticeable growth in the number of institutions that are integrating technological tools and strategies, including digital social networks, to communicate and build relationships with students and other stakeholders is evident.

Thanks to the popularity of digital social networks among students, institutions of higher education have begun to implement strategies to attract prospective students, communicate with current students and maintain their relationships with alumni [23-25].

94% of institutions of higher education in the United States use Facebook to connect with their stakeholders. Other social media such as Twitter, Linkedin and YouTube, are also used for relationship marketing strategies, although at a slower rate. In addition, almost 60% of institutions have placed their different social media accounts on their institutional websites, confirming the importance of these platforms and the institutionalization of digital social networks [25].

According to research conducted by Pearson [26] thatsurveyed approximately 8,000 schools from a variety of disciplines and institutions of higher education in the United States on the use of social media, 78.9% of respondents considered that the use of digital social networks increased the level of communication with an institution's students and stakeholders. Similarly, respondents felt that sharing information through these means helped in creating better learning environments.

Digital social networks are popular as an information search tool, where comments and content can be shared and communicated, and are also used for other activities. Therefore, they represent a great opportunity for higher education institutions to interact, discuss and listen to different stakeholders, in a context of globalized higher education but are also a challenge [27].

Currently, the fact that young people make frequent use of the Internet and, particularly digital social networks, should not be considered as a waste of their time, because they are acquiring technological and communication skills useful in the modern world. The potential to use digital social networks as communication channels, as well as a source of information and entertainment, is great for higher education institutions. This is confirmed by the favorable attitude of students in relation to the academic use of this type of media.

Digital social networks promote firstly, self-learning; secondly, teamwork through information sharing; thirdly, access to other information sources that facilitate collaborative learning, and finally, communication amongst students, between students and teachers and also between students and administrative staff. All these advantages enable higher education institutions practice an interactive approach and promote a more dynamic learning environment, ensuring a better quality education.

3. Importance of the relational coordination in institutions of higher education

Gittell [28] defines the relational coordination model as a "mutually reinforcing process of interaction between communication and relationships carried out for the purpose of task integration".

Gittell [16] explains how this model differs from other ones, due to the fact that the proposal is based on three specific dimensions of relationship that are required for effective coordination. While other theories stress the importance of shared knowledge, the relational coordination model explains that although this is required, it is not a sufficient condition. According to this model, to achieve effective coordination, the agents that take part in the process must also be connected via shared objectives and a relationship of mutual respect.

The relational coordination model focusses on the relationships that emerge from the various roles that shape a process, more than on the relationship that some individuals maintain with each other during work [29].

The model is structured upon two kinds of relationships: relationship ties and communication ties.

Some of the relational dimensions that can be mentioned are shared goals, shared knowledge and mutual respect.

3.1. Shared objectives

In order to ensure teamwork quality it is important that each member shares his/her knowledge and experience of relevant tasks with the team [30]. Whenever different members of a team share the same objectives and values, inter-personal relationships are reinforced [31].

At a team level, establishing a set of objectives is considered to be the main factor in guaranteeing the

completion of tasks [32]. The quality and the degree of acceptance of the ideas generated by the members of the team increases whenever the members cooperate [33]. Therefore, sharing objectives is important to achieve quality teamwork and it is necessary if team objectives are to be met. The greater the support provided to the members of a team, the more effective the final results will be, and team moral can be improved too [34].

Shared objectives become an essential element for teamwork quality. Competition between members of a team can have a positive influence on their motivation and on returns derived from individual tasks, however, cooperation or support amongst team members is more important [35].

In a team, having shared objectives, instead of promoting competition between members, improves productivity amongst them when undertaking different tasks. Besides, when members share the same mission or vision, it increases the probability of being able to prioritize team tasks and of sharing a similar interpretation of work rules [36].

According to Gittell [16], thanks to the existence of a set of shared objectives in their work, stakeholders are able to develop ties that allow them to reach conclusions that are compatible with the different ways of thinking and acting, as new information is made available.

3.2. Shared knowledge

The knowledge found in work teams is considered to be shared knowledge or team knowledge [37-39]. The concept of shared knowledge consists in the fact that team members have compatible or similar knowledge that leads them to share expectations for team tasks [40,41]; these expectations allow the team to coordinate and forecast the behavior and needs of team members [37].

The generation of knowledge can appear at individual or team level, and it is the result of a communication process at an individual and team level, it is often the result of a bi-directional communication process in which information is shared with other people [42-44,36]. This process of communication serves to develop the spirit of work too [43].

The interchange of knowledge in teamwork is important to generate trust and improve teamwork effectiveness. Storck [45] and Herbsleb and Moitra [46] reinforce this observation, by affirming that with a lack of effective information interchange, projects can suffer from coordination problems that drive them to fail. Team interaction is more productive when information shared is relevant, accurate and useful [36].

Knowledge exchange is required in order to integrate the different disciplines represented on a team along with the ideas, knowledge and information held by team members as well as to ensure frequent communication [37,47,48]. When knowledge is not integrated, the process will be slower and more complex [49].

Shared knowledge refers to the provision of information and *know-how* that is required to help others to better undertake their work and fulfill obligations [50,51]. It can be offered by means of written or verbal communication, through the creation of networks with other experts or by documenting, organizing and obtaining knowledge from others [50,51].

Shared knowledge becomes the main way in which employees in a company can contribute to the application of knowledge, innovation and competitive advantage [52].

Gittell [16] shows that when the different members of the work group know how their objectives are related to the rest of the objectives of the group taking part in the same process, a dynamic is created in which everyone knows the implications that each change will have over any other task or role.

3.3. Mutual respect

A key aspect of teamwork quality is that team members openly share information [53,54]. Lack of respect amongst the members of a team becomes a barrier to teamwork that is the integration of a team's knowledge and experience in common tasks.

The members of a team that work towards a common objective must show each other mutual respect and help each other to develop the ideas and contributions of other team members whenever it is required, instead of fighting each other [36].

Mutual respect is "key to maintain a healthy labor climate" [55]. Therefore, an essential factor for teamwork quality is that contributions are made based on respect [36].

A leader must not treat any other member badly. In teamwork there are different levels and positions, but the contribution of skills and values to the team by all members is equally important [55].

Work teams based on mutual respect amongst members are more productive and maintain a positive work climate [55]. Productive interactions require mutual respect, a shared language and a willingness to participate in the teamwork necessary to resolve conflicts [56].

Mutual respect between people ensures that all the members of the team can express their ideas. In this way, the team will benefit from each other's talent [57]. The respect for another member's ability, implied in the process, establishes a powerful link that is applied holistically to the whole process, resulting in an effective coordination between team members [16].

Gittell's model includes several communication dimensions, some of these are: frequent communication, timely communication, accurate communication and problem solving communication.

3.4. Frequent communication

Teamwork members need to be connected continually, by making frequent and solid communication channels possible so as to increase feedback, and allow for the correction of errors, the generation of ideas and problem solving [58,59]. This frequent interaction would produce more and new ideas and improves the decision making process [31,35,60].

According to the theory of information processing, frequent communication is required to decrease the levels of uncertainty often found in complex processes [61].

Additionally, frequent communication lays the foundation for other factors that then determine the outcomes

of teamwork. For example, communication is needed to coordinate the efforts of team members and knowledge [62]; so as to understand the collective mission of the team [63], to be sure that the team shares the same mental model over time [44], and to facilitate trust in the team [64] o. Team performance can be defined as the extent to which the team is able to deliver on quality, costs and the established objectives on time [65].

Frequent communication is needed (a process where participants create and share information amongst themselves to reach a mutual understanding) and the team leader must coordinate and facilitate this communication [66], to help establish the relationships amongst different roles through the proximity generated as a consequence of repeated interaction [16].

3.5. Timely communication

Delays in communication can have negative implications on final organizational goals. Therefore it is important to have communication in the precise moment that it is required to have success in the organizational goals [67].

Communication is a key factor to achieve quality teamwork. It offers a means to exchange information, share ideas amongst the members of the team, coordinate efforts and offer feedback [54]. Today, information exchange is not only important but ensuring that the right information is offered to the right person at the right moment is very much desired [54,68,69] Communication refers to the capacity of team members to understand information, network exchange, and the use of these networks to improve the interchange of information [70]. Communication serves as a mechanism for the interaction of the members of a team. [71].

The communication style of an organization is related to its business growth, so a style based on open and accurate communication increases the interest and trust of employees [72].

3.6. Accurate communication

There is great evidence that the different aspects of an effective communication, such as frequency and precision over policies and organizational routines are positively related with an employee's productivity [73-76].

Accurate communication in terms of the relevant information plays a critical role in the effectiveness of group work performance [74].

It is important for team members to communicate directly to each other, since the interchange of information by means of intermediaries demands more time, and runs the risk of errors appearing in the message when it is sent [30,54]. To achieve this, it is extremely important that members of a team are open to the other members and avoid hiding important information, since it can damage the integration of knowledge and experience of the team members, this being the main function of teamwork [36].

The effective application of organizational strategies and human resources policies depend on the accuracy of information exchange amongst team members [77]. The lack of precise communication between managers can create strategic misalignment [78].

3.7. Problem solving communication

Effective communication requires commitment from the stakeholders that take part in the task to a communication policy oriented at solving problems that appear in groups characterized by high interdependence. Blaming other workers or avoiding one's own responsibilities can lead to negative consequences that have an impact on the final results [79].

Communication is a required component for coordination and teams perceive communication problems as a result of experiencing problems with coordination processes [71]. Communication is a required component to the collaborative solution of problems and teams will find it much more difficult to resolve an issue without it [80].

Often, some problems dealing with team work appear when the actions of one or more team members are non compatible with the rest of the team members [81], and they increase when the members feel that their different objectives cannot be reached at the same time as the rest of the members [82].

Problem solving communication allows team members to adapt in a more flexible way to unforeseen and negative circumstances that may appear. Therefore, this represents an important aspect of teamwork quality [83].

Quality relationships allow team members to coordinate an effective work process, based on frequent, accurate and timely communication oriented at the solution of problems [83].

A brief description of each of the dimensions is shown in Table 1.

The relational coordination model proposed by Gittell [28] can improve results for organizations or organizational processes where high levels of task interdependence [84], uncertainty [85], time restrictions [86] and tacit knowledge [87] exist.

The application of coordination mechanisms in upper educational institutions can improve their level of academic quality.

Effectiveness in teaching and learning processes depends on the ability to coordinate different agents properly in the exchange of ideas, knowledge, and objectives and by demonstrating mutual respect between all the stakeholders. [29]. This is how different relationships are managed efficiently,, by allowing organizational objectives to be reached and ensuring team satisfaction [15,16].

The relational coordination model has been applied to different industries such as, health [15,16], and it has been applied to different medical units in hospitals. Those medical units that presented high degrees in terms of relational coordination were the ones that achieved better final results; López et al. [88], and De Pablos y López [3], have applied the model to show the organizational excellence of the Spanish Organ Transplant System; Haider [89] and de Pablos and Haider [90] have measured teamwork quality by making use of the relational coordination model and their research has been applied to the process of accurate myocardial infarction. De Pablos et al. [90] have applied this model to show best results in cloud computing practices and in education. De Pablos et al. [3] [90] have demonstrated how

higher degrees of relational coordination in upper education institutions lead to higher degrees of satisfaction in lecturers and students.

The need for coordinating is a prerequisite to achieving better results in companies [90]. Thompson [84] describes the importance of effective coordination amongst high interdependent tasks. De Pablos and Haider [90] observe how mutual adjustment improves coordination mechanisms at organizations, such as for example, routines, timetables, previous planning and task normalization.

Coordination is the integration of work in conditions of task interdependencies [91]. Lawrence and Lorsch [92], and De Pablos et al. [90], have studied the relationship between work coordination and final results in organizations, they have found that coordination is positively related with final competitive outcomes in organizations.

The relational coordination model focuses on understanding the importance of coordinating relationships and the dynamics of communication in organizations to achieve best results. [16]. This model has been applied to different kinds of organizations with promising results so far [3,88].

From the model it can be confirmed that relational coordination is produced by means of frequent high quality communication with the support of shared objectives and knowledge and mutual respect.

According to Gittell [16], relational coordination allows firms to obtain better levels of quality and effectiveness. It has to do with the sharing of objectives, values and the existence of efficient communication mechanisms amongst workers in a firm. According to this model, the existence of high levels of relational coordination imply higher levels of excellence and competitiveness for upper educational institutions [90].

Table 1.
Dimensions and elements for the Relational Coordination Model

Dimensions	Elements	Description
Relational Dimensions	Shared objectives	Shared objectives offer teamwork members strong linkages and facilitate the conclusions on how answering to new information.
	Shared knowledge	It is important that the members of a team own the global knowledge that it is required to better know how changes impact in the process.
	Mutual respect	The respect for the rest of the members of the team creates strong linkages amongst them.
Comunication dimensions	Frequent communication	Frequent communication helps to build relationships.
	Timely communication	Communication can be frequent however of low quality. For certain work interdependencies, delays in the communication can be transformed into fatal errors.
	Accurate communication	Apart from frequent and on time, communication must be accurate. Otherwise, important mistakes can also take place.
	Problem solving communication	Interdependent tasks often generate problems that must be solved together.

Source: Adapted from [29]

Although communication mechanisms are not a serious problem in educational ecosystems, the lack of shared objectives and mutual respect can become barriers when the search for excellence is a must [93].

A quality upper education system is key to creating value and it has become a priority for policy makers worldwide [94,95]. Therefore, the improvement of quality in education must also be oriented towards the increase in the degree of coordination between lecturers, the internal organization and the learning objectives. In this respect, Marengo and Dosi [96] and Brunner [97] indicate that the organizational systems and the government structures of upper education systems determine the success and the achievement of objectives in terms of upper education.

4. Application of relational coordination model in higher education institutions to optimize the use of digital social networks

In previous sections the literature review on the suitability of higher education institutions (HEI) using social media (HEIs) was discussed. The importance of relational coordination in these institutions in order to achieve organizational goals and interpret higher levels of academic quality has also been described. This section will explain how the model of Relational Coordination can be applied to optimize the use of digital social networks in HEIs.

Digital social networks are now fully established in our society. They have completely and quickly changed the way people relate on a personal and professional level [98]. The implementation and use of social networks must be a part of the strategic planning of HEIs. However, how can HEIs apply the Relational Coordination model to optimize the use of online social networks?

The relational coordination model, as has been explained, provides an organizational framework that can help HEIs optimize the use of social media in their internal and external relationships. Digital social networks have radically changed the way in which internal relationships are managed in HEIs, substantially improving the quality of the institutions. Rather than understanding them as a substitute for personal relationships, social media should be seen as an extension of these, allowing for an enrichment of human experience and making it possible to maintain daily contact with a wider circle of agents than would be physically possible using traditional methods [99], as well as becoming a new channel through which institutions can share strategies with employees.

The option to record sound and share videos via Vine or Instagram generates a climate of trust in teams, allowing members to exchange information with the entire community and also transmit how culture is lived in the institution [100]. For HEIs that have multiple branches located in different geographical areas where communication is difficult, digital social networks will be useful, becoming increasingly the main source of information for people and providing solutions to their problems through direct communication

without geographical barriers [101]. Also, creating a corporate group on Linkedin or Facebook where employees share their knowledge, create debates, etc. will promote a sense of belonging and also helps position the HEI network.

Internal relations are central to achieving the goals set by the HEI, to building a positive corporate culture and to improving the morale and productivity of employees [102] and digital social networks can be a great ally for management, when managed correctly and applied to the relationships between the different roles that shape the processes, as posed by the Relational Coordination model [15].

Digital social networks can also be useful tools in recruitment efforts and can improve the competitiveness of a HEI, allowing it to establish relationships not only with potential students, but also with current students, alumni and the community in general [17]. To do this, it is necessary to have a marketing plan on social media that contains strategies and action plans that are aligned with the objectives of the institution, but also with the individual goals of each department, study area or faculty. To implement the plan, coordination mechanisms between team members are needed. The team would consist of different profiles, one important one being the Community Manager, who would be the person in charge of building, managing and moderating communities around a brand, product or service on the Internet, in this case the HEI.

The HEI should have several Community Managers that help optimize the institutional use of digital social networks to strengthen relations and communications with users creating an active and participatory community [103]. One Community Manager should be responsible for communicating the institution's identity and should focus on general objectives, and there should also be one for each department, faculty or study area who is responsible for managing relationships with stakeholders, but also focuses on the general objectives of the institution. The scope of the general objectives of the HEI can be reduced if there is not proper coordination between the members that are part of this process. It is here that the application of the Relational Coordination model is useful to optimize the use of digital social networks, because, as mentioned earlier, the HEI have general goals and these will be shared by the different Community Managers, who in turn will have to exchange information frequently. Therefore we can say that the Relational Coordination model allows team members to work together in achieving the individual and overall objectives of the HEI.

5. Conclusions

The use of social media by Higher Education Institutions (HEIs) as a tool to enhance communications and relationships with its stakeholders provides multiple opportunities to understand the quality of education.

The main goals that HEIs wish to achieve with their digital social networks strategies are: to establish trusting relationships with their students and target audience, to strengthen corporate image, to achieve greater recognition, to promote the brand and to improve competitiveness in the sector.

The literature review suggests that HEIs can achieve all of these goals, but that while the use of ICT and particularly digital social networks provides organizational benefits, it is not enough to just create an institutional account in all social media. HEIs should have a plan to optimize their use and develop mechanisms to address them, joining forces to share knowledge and objectives, promoting respect for the work of each team member, and also promoting effective communication mechanisms.

In conclusion, the usefulness of digital social networks in HEI cannot be questioned, it is a natural evolution in which institutions should be involved, as clear evidence exists of how it can improve key relationships and aspects of communication that directly enrich the competitiveness and reputation of the institution.

Additionally, the use of digital social networks in higher education institutions are very important not only in the formative stage, but also in their professional future, because companies now make use of these as a marketing and internal communication tool.

Finally, although the use of digital social networks provides a number of advantages to institutions of higher education, these are only tools, hence, success depends on the way in which these tools are used, which presents a big challenge to HEIs, who may use the relational coordination model to manage relationships and communications efficiently.

References

[1] Yáñez, M. y Villatoro, P., Las nuevas tecnologías de la información y de la comunicación (TIC) y la institucionalidad social. Hacia una gestión basada en el conocimiento. Publicación de las naciones unidas, Chile: Publicación de las Naciones Unidas, 2005.

[2] Grandi, C., Gonzáles, F. y Segarra, M., ¿Cómo potenciar el uso de las redes sociales como estrategia de marketing en empresas del sector hostelero?, Tec. Empresarial, 2014.

[3] De Pablos, C., López, J., Romo, S. y Medina, S, Organización y transformación de los sistemas de información en la empresa, España: Editorial ESIC, 2012.

[4] Arango, M., Branch, J. y Londoño, J, Arquitectura empresarial como instrumento para gestionar la complejidad operativa en las organizaciones, Revista DYNA, 81(185), pp. 219-226, 2014.

[5] Celaya, J., La empresa en la web 2.0., España: Gestión 2000, 2011.

[6] Castelló, A., Estrategias empresariales en la web 2.0. Las redes sociales online, España: Editorial Club Universitario, 2010.

[7] Keyes, J., Enterprise 2.0. Social networking tools to transform your organization, USA: Taylor & Francis group, 2013.

[8] Carballar, J., Social media, marketing personal y profesional. Informática y comunicaciones en la empresa, España: RC Libros, 2012.

[9] Steel, E. and Fowler, G.A., Big brands like Facebook, but they don't like to pay, The Wall Street Journal, 2011.

[10] Warren, C., Time spent on social networks has tripled, 2009.

[11] Griffith, S. and Liyanage, L., An introduction to the potential of social networking sites in education. Emerging technologies conference, 2008. [Online]. Available at: http://ro.uow.edu.au/cgi/viewcontent.cgi?article=1008&context=etc08.

[12] Stoller, E., Using social media to enhance engagement, yield, and retention. Inside higher education, [Online]. 2011. Available at: http://www.insidehighered.com/blogs/student_affairs_and_technology/using_social_media_to_enhance_engagement_yield_and_retention.

[13] Ratliff, A., The journal of technology in students affairs. University of Alabama, 2011.

[14] Scott, C., Allen, C. and Glen, B., Manual de relaciones públicas eficaces, Barcelona, España: Ediciones Gestión 2000, 2006.

[15] Gittell, J., High performance healthcare: Using the power of relationships to achieve quality, efficiency and resilience, Nueva York, USA: Mc Graw-Hill, 2009.

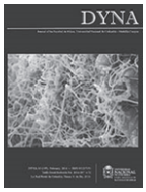
[16] Gittell, J., Relational coordination: Guidelines for theory, measurement and analysis, [Online]. 2010. Available at: www.jodyhoffergittell.info/content/rc.html.
[17] Wilson, C., Making connections: Higher educations meets social media, Change: The magazine of higher learning, 2013.
[18] Noel-Levitz, I., e-Expectations class of 2007 report: Building an e-recruitment network: 7, 2007.
[19] Solis, B., Customer service: The art of listening and engagement through social media, 32, 2008.
[20] Nolan, J., Social media in the business of higher education, 2012. [Online]. Available at: http://www.huffingtonpost.com/dr-james-michael-nolan/social-media-in-the-busin_b_1607245.html.
[21] Zog Digital., The benefits of social media for higher education, [Online]. 2013. Available at: http://www.business2community.com/social-media/the-benefits-of-social-media-for-higher-education-0445218#!byaijd.
[22] Kaplan, A. and Haenlein, M., Users of the world, unite! The challenges and opportunities of social media, business horizons, 53(1), pp. 59-68, 2010.
[23] Education Insider., Admissions offices turn to social media to connect with prospective students, 2010.
[24] Pearson., Social media in higher education, [Online]. 2011. Available at: http://www.pearsonlearningsolutions.com/higher-education/social-media-survey.php.
[25] Slover-Linett, C. and Stoner, M., Succeeding with social media: Lessons from the first survey of social media in advancement, 2011.
[26] Pearson., Social media for teaching and learning. Annual survey of social media use by higher education faculty, 2013.
[27] Smith, A., Rainie, L. and Zickuhr., College students and technology. Pew internet & american life project, 2011. [Online]. Available at: http://www.pewinternet.org/2011/07/19/college-students-and-technology/
[28] Gittell, J., Coordinating mechanisms in care provider groups: Relational coordination as a mediator and input uncertainty as a moderator of performance effects. Manag. Sci. 48, pp. 1408-1426, 2002. DOI: 10.1287/mnsc.48.11.1408.268
[29] Smith, A., Rainie, L. and Zickuhr., College students and technology. Pew internet & american life project, 2011. [Online]. Available at: http://www.pewinternet.org/2011/07/19/college-students-and-technology/
[30] Hackman, J., The design of work teams. In: Lorsch, J.W. ed. Handbook of organizational behavior, Prentice-Hall, Englewood Cliffs, NY, 1987. pp. 67-102.
[31] Hackman, J., Groups that work (and those that don't), San Francisco, 1990.
[32] Vinagre, J., Del Llano, J., De Pablos, C. and Nuitjen, M., Relational coordination and healthcare management in lung cancer. World Journal of Clinical Cases, 2(12), pp. 757-768, 2014. DOI: 10.12998/wjcc.v2.i12.757
[33] Cooke, R. and Szumal, J., The impact of group interaction styles on problem-solving effectiveness, Journal of applied behavioral science, 30(4), pp. 415-437, 1994. DOI: 10.1177/0021886394304005
[34] Heaney, C., Price, R. y Rafferty, J., Increasing coping resources at work: a field experiment to increase social support, improve work team functioning, and enhance employee mental health, Journal of organizational behavior, 16(4), pp. 335-352, 1995. DOI: 10.1002/job.4030160405
[35] Tjosvold, D., Cooperation theory, constructive controversy, and effectiveness: Learning from crisis. In: Guzzo, R.A., Salas, E. and Associates Eds., Team effectiveness and decision making in organizations, San Fransisco, 1995, pp. 79-112.
[36] Hoegl, M. and Gemuenden, H., Teamwork quality and the success of innovative projects: Theoretical concept and empirical evidence organization science, 12(4), pp. 435-499, 2001.
[37] Cooke, N., Cannon-Bowers, J. and Stout, R., Measuring team knowledge, Human Factors, 42, pp. 151-173, 2000. DOI: 10.1518/001872000779656561
[38] Ensley, M. and Pearce, C., Shared cognition in top management teams: Implications for new venture performance, Journal of Organizational Behavior, 22(2), pp. 145-160, 2001. DOI: 10.1002/job.83
[39] Rentsch, J. and Klimoski, R., Why do 'great minds' think alike?: Antecedents of team member schema agreement, Journal of Organizational Behavior, 22, pp. 107-120, 2001. DOI: 10.1002/job.81
[40] Cannon-Bowers, J., Salas, E. and Converse, S., Shared mental models in expert team decision making. In: Castellan, N.J., Jr. (Ed.), Individual and group decision making: Current issues, England: Lawrence Erlbaum Associates, 1990, pp. 221-246.
[41] Kozlowski, S. and Ilgen, D., Enhancing the effectiveness of work groups and teams. Psychological Science in the Public Interes, 7, pp. 77-124, 2006. DOI: 10.1111/j.1529-1006.2006.00030.x
[42] Safón-Cano, V., Creación y desarrollo del conocimiento en la organización, Revista Europea de Dirección y Economía de la Empresa, 6(2), pp. 115-126, 1997.
[43] Mickan, S. and Rodger, S., Characteristics of effective teams, Australian Health Review, 23(3), pp. 201-207, 2000. DOI: 10.1071/AH000201
[44] Salas, E., Burke, C. and Cannon-Bowers, J., Teamwork: Emerging principles, International Journal of Management Reviews, 2(4), pp. 339-356, 2000. DOI: 10.1111/1468-2370.00046
[45] Storck, P. and Lettenmaier, D., Trees, snow and flooding: an investigation of forest canopy effects on snow accumulation and melt at the plot and watershed scales in the pacific northwest, water resources series, technical report 161, university of Washing, 2000.
[46] Herbsleb, J. and Moitra, D., Global software development. IEEE Software, 2001. DOI: 10.1109/52.914732
[47] Edmondson, A. and Nembhard, I., Product development and learning in project teams: The challenges are the benefits. Journal of product innovation management, 26(2), pp. 123-138, 2009. DOI: 10.1111/j.1540-5885.2009.00341.x
[48] Ratcheva, V., Integrating diverse knowledge through boundary spanning processes – The case of multidisciplinary project teams, International Journal of Project Management, 27(3), pp. 206-215, 2009. DOI: 0.1016/j.ijproman.2008.02.008
[49] Sapsed, J., Bessant, J., Partington, D., Tranfield, D. and Young, M., Teamworking and knowledge management: A review of converging themes, International Journal of Management Reviews, 4(1), pp. 71-85, 2002. DOI: 10.1111/1468-2370.00077
[50] Pulakos, E., Dorsey, D. and Borman, W., Hiring for knowledge-based competition. In: Jackson, S.E., Hitt, M.A. and Denisi, A.S. (Eds.), Managing knowledge for sustained competitive advantage: Designing strategies for effective human resource management, 2003, 155 P.
[51] Cummings, J., Work groups, structural diversity and knowledge sharing in a global organization, Management Science, 50(3), pp. 352-364, 2004. DOI: 10.1287/mnsc.1030.0134
[52] Jackson, S., Chuang, C., Harden, E., Jiang. and Joseph, J., Toward developing human resource management systems for knowledge-intensive teamwork. In: Joseph, J.M. (Ed.), Research in personnel and human resources management, 25, pp. 27-70, 2006.
[53] Gladstein, D., Groups in context: A model of task group effectiveness. Administrative Science Quarterly, 29(4), pp. 499-517, 1984. DOI: 10.2307/2392936
[54] Pinto, M. and Pinto, J., Project team communication and cross-functional cooperation in new program development, Journal of Product Innovation Management, 7(3), pp. 200-212, 1990. DOI: 10.1016/0737-6782(90)90004-X
[55] Weinbach, R., The social worker as manager: A practical guide to success, Boston: Pearson Education, inc, 2003.
[56] Palinkas, L., Aarons, G., Chorpita, B., Hoagwooh, K., Landsverk, J. and Weisz, J., Cultural exchange and the implementation of evidence-based practices: two case studies, Research on social work practice, 2009. DOOI: 10.1177/1049731509335529
[57] Ferguson., Teamwork skills. Third edition, USA, 2009.
[58] Tushman, M. and Nadler, D., Information processing as an integrating concept in organizational design, Academy of Management Review, 3, pp. 613-624, 1978. DOI: 10.2307/257550
[59] Katz, R. and Tushman, M., Communication patterns, project performance, and task characteristics: An empirical evaluation and integration in an R&D setting. Organizational Behavior and Human Performance, 23(2), pp. 139-162, 1979. DOI: 10.1016/0030-5073(79)90053-9
[60] Moye, N. and Langfred, C., Effects of task autonomy on performance: An extended model considering motivational, informational and structural mechanisms, Journal of Applied Psychology, 89(6), pp. 934-945, 2004. DOI: 10.1037/0021-9010.89.6.934
[61] Souder, W. and Moenaert, R., Integrating marketing and R&D project personnel within innovation projects: An information uncertainty model,

Journal of Management Studies, 29(4), pp. 485-512, 1992. DOI: 10.1111/j.1467-6486.1992.tb00675.x
[62] Hans, H., Lee, J. and Seo, Y., Analyzing the impact of a firm's capability on outsourcing success: A process perspective. Information & Management, 45(1), pp. 31-42, 2008. DOI: 10.1016/j.im.2007.09.004
[63] O'Connor, M., The human capital era: Reconceptualizing corporate law facilitate labor-management cooperation, Cornell Law Review, 78(5), pp. 899-965, 1993.
[64] Jarvenpaa, S.L. and Leidner, D.E., Communication and trust in global virtual teams, Journal of Computer-Mediated Communication, vol. 3(4), 2006. DOI: 10.1111/j.1083-6101.1998.tb00080.x
[65] Schrader, S. and Goepfert, J., Structuring manufacturer-supplier interaction in new product development teams: An empirical analysis, Institut für Innovationsforschung und Technologiemanagement, Universität München, 1996.
[66] Rogers, E. and Kincaid, D., Communication networks: Toward a new paradigm for research. New York: Free Press, 1981.
[67] Waller, M., The timing of adaptative group responses to non-routine events. Acad. Manag. J., 42, pp. 127-137, 1999. DOI: 10.2307/257088
[68] Brodbeck, F., Communication and performance in software development projects, European Journal of Work and Organizational Psychology, 10(1), pp. 73-94, 2001. DOI: 10.1080/13594320042000043
[69] He, J., Butler, B. and King, W., Team cognition: Development and evolution in software project teams, Journal of Management Information Systems, 24(2), pp. 261-292, 2007. DOI: 10.2753/MIS0742-1222240210
[70] Stevens, M. and Campion, M., The knowledge, skill, and ability requirements for teamwork: Implications for human resource management, Journal of Management , 20, pp. 503-530, 1994. DOI: 10.1177/014920639402000210
[71] Marks, M., Mathieu, J. and Zaccaro, S., A temporally based framework and taxonomy of team processes, The Academy of Management Review, 26, pp. 356-376, 2001. DOI: 10.2307/259182
[72] Nelson, C., Communication styles and business growth, Journal of Financial Planning, third quarter, pp. 8-11, 2010.
[73] O' Reilly, C., Superiors and peers as information sources, work group supportiveness, and individual decision-makmg performance, Journal of Applied Psychology, 62, pp. 632-635, 1977. DOI: 10.1037/0021-9010.62.5.632
[74] Roberts, K. and O' Reilly, C., Some correlates of communication roles in organizations, Academy of Management Journal, 22, pp. 42-57, 1979. DOI: 10.2307/255477
[75] Snyder, R. and Morris, J., Organizational communication and performance, Journal of Applied Psychology, 69, pp. 461-465, 1984. DOI: 10.1037/0021-9010.69.3.461
[76] Kacmar, K., Witt, L. and Gully, S., The interactive effect of leadermember exchange and communication frequency on performance ratings, Journal of Applied Psychology, 88, pp. 764-772, 2003. DOI: 10.1037/0021-9010.88.4.764
[77] Goggin, M., Bowman, A., Lester, J. and O'Toole, L., Implementation theory and practice: Toward a third generation. Glenview, II: Scott, Foresman/little, Brown Higher Education, 1990.
[78] Tourish, D., Critical upward communication: Ten commandments for improving strategy and decision making. Long Range Planning, 38(5), pp. 485-503, 2005. DOI: 10.1016/j.lrp.2005.05.001
[79] Deming, W., Out of the crisis. Mit press, 1986. 507 P.
[80] Laughlin, P., Collective induction: Group performance, social combination processes, and mutual majority and minority influence, Journal of Personality and Social Psychology, 54, pp. 254-267, 1988. DOI: 10.1037/0022-3514.54.2.254
[81] Forsyth D., Group Dynamics, 2nd ed. Pacific Grove, CA: Brooks/Cole, 1990.
[82] Levine, J. and Moreland, R., Progress in small group research. Annual review of psychology, 41, pp. 586-634, 1990. DOI: 10.1146/annurev.ps.41.020190.003101
[83] Kyriakidou, O. and Özbilgin, M., Relational perspectives in organizational studies: A research companion. Edward Elgar Publishing, 2006. DOI: 10.4337/9781781950548
[84] Thompson, J., Organizations in Action, New York: McGraw-Hill, 1967.
[85] Argote, L., Input uncertainty and organizational coordination in hospital emergency units, Administrative Science Quarterly, 27, pp. 183-197, 1982. DOI: 10.2307/2392320
[86] Adler, C., Malhotra, A., Elman, I., Goldberg, T., Egan, M., Pickar, D., and Breier, A., Comparison of ketamine-induced thought disorder in healthy volunteers and thought disorder in schizophrenia, Am J Psychiatry 156(10), pp. :1646-1649, 1999.
[87] Nokata, I. and Takeuchi, H., The knowledge-creating company: How japanese companies create the dynamics of innovation, Publisher: Oxford University Press, 1995, 304 P.
[88] López, D., De Pablos, C., De La Puerta, E. and Fernández, C., Productivity in service systems: Towards a managerial framework, Service Science, 3(3), pp. 223-238, 2011. DOI: 10.1287/serv.3.3.223
[89] Haider, S., An organizational framework for the implementation of evidence based practices to reduce door to balloon time in patients undergoing primary angioplasty, Tesis. Universidad Rey Juan Carlos. Spain, 2013, 183 P.
[90] De Pablos, C., Haider, S. and Garcia, A., Relational coordination as an indicator of teamwork quality: Potencial application to the sucess of e-learning at Universities. International Journal of Emerging Technologies in e-Learning, 10(2), pp. 4-8, 2015.
[91] Faraj, S. and Xiao, Y., Coordination in fast response organizations, Management Science ,52(8), pp.1155-1169, 2006. DOI: 10.1287/mnsc.1060.0526
[92] Lawrence, R. and Lorsch, P., Organization and environment: Managing differentiation and integration. Boston: Harvard Business School., 1967.
[93] Torres, D., Moreno, J., Robinson, N., Delgado, E. and Herrera, F., ISI rankings of Spanish universities according to fields and scientific disciplines, Revista Profesional de la Informacion, 20(6), pp. 701-709, 2011.
[94] Brown, L., Securing a sustainable future for higher education in England, Report, 2010.
[95] Horizon, The new media consortium, UE, Report, 2012.
[96] Marengo, L. and Dosi, G., Division of labor, organizational coordination and market mechanisms in collective problem-solving, Journal of Economic Behavior & Organization, 58, pp. 303-326, 2005. DOI: 10.1016/j.jebo.2004.03.020
[97] Brunner, J., University governance: Typology, dynamics and trend., Revista de Educación, 355, pp. 137-159, 2011.
[98] Safko, L., The social media bible. Tactics, tolos & strategies for business success, Innovative Thinking, 2010.
[99] Mainar, H., Tu empresa en la Web 2.0 aumenta tus beneficios gracias a Facebook, Twitter y otras redes sociales, Bubok Publishing S.L., 2011.
[100] Noor, H. and Allen, J., Social media and strategic communications, USA: Editorial Matter, 2013. DOI: 10.1057/9781137287052
[101] Schroeder, J., Snapshot aesthetics and the strategic imagination, in visible culture, 2013.
[102] Brown, D., Internal communications should be of vital importance to any business, The Business Review, 2002.
[103] Guzmán, A., Del Moral, M., Gonzáles, F. y Gil, H., Contribución de Twitter a la mejora de la comunicación estratégica de las universidades latinoamericanas, Revista de Universidad y Sociedad del Conocimiento (RUSC), 10(2), pp. 236-251, 2013.

Lacayo-Mendoza, received a BSc. in Systems Engineering in 2005 from Universidad Politécnico Gran Colombiano de Colombia, an MSc. degree in International Business Management in 2011, from the Universidad Rey Juan Carlos de España, Madrid, Spain and is enrolled on a PhD program in Business Organization at Universidad Rey Juan Carlos de España. He works as a digital marketing consultant, and as a lecturer.
ORCID: 0000-0002-6629-5601

C. De Pablos-Heredero, holds a PhD in Economics and Business Studies from Universidad Complutense de Madrid, Madird, Spain. She is Director of the Master in Business Organization and co - director of the Master of Entrepreneurship and the Master in SAP ERP Logistics Project Management at Universidad Rey Juan Carlos, Spain. She has published books and articles in indexed journals as: El Profesional de la Información, Revista Mejicana de Ciencias Agropecuarias, Archivos de Zootecnia, Italian Journal of Animal Science, CEDE, Universia, Interciencia, TIBE, Intangible Capital, Service Science, Journal of Entrepreneurship Management, International Journal of Marketing Research, Revista de Ciencias Sociales, Revista de Economía Mundial, Dyna, Pensée, Medical Economics, REIS, etc.
ORCID: 0000-0003-0457-3730

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS



DYNA
http://dyna.medellin.unal.edu.co/

Use of residual powder obtained from organic waste to partially replace cement in concrete

Indara Soto-Izquierdo [a] & Marcio Antonio-Ramalho [b]

[a] *Departamento de Ciências Exatas e Tecnológicas, Universidade Estadual de Santa Cruz, Ilhéus, Brasil. indaritasi@gmail.com*
[b] *Departamento de Engenharia de Estruturas, Universidade de São Paulo, São Paulo, Brasil. ramalho@sc.usp,br*

Received: August 1rd, 2014. Received in revised form: May 15th, 2015. Accepted: January 28th, 2016.

Abstract
The main objective of this research is to investigate the feasibility of using the residual powder obtained from organic waste to replace a part of Portland cement during concrete production. The chemical and physical properties of the organic residual powder were first investigated using scanning transmission electron microscopy (STEM) and energy dispersive X-ray spectroscopy (EDS). The replacement ratios of cement have also been studied for 0%, 5%, 10%, 15%, and 20% of the residual powder by weight. Moreover, we evaluated the compressive strength, specific gravity, water absorption, and voids content of the concrete specimens that were modified with different concentrations of the residual powder. Results indicate that 5% residual powder concentration resulted in better physical and mechanical properties of the modified concrete when compared with the reference concrete. At this concentration, the addition of residual powder act as fillers in concrete, reducing the amount of voids and causing a higher density in the material.

Keywords: residual powder; organic waste; cement; concrete; mechanical properties.

Utilización de polvo de residuos orgánicos como sustitución parcial del cemento en el hormigón

Resumen
El objetivo principal de este trabajo es estudiar la viabilidad de utilizar el polvo obtenido a partir de residuos orgánicos para substituir parte del cemento Portland en la producción del hormigón. Inicialmente fueron evaluadas las propiedades químicas y físicas del polvo residual usando un microscopio electrónico de transmisión (TEM) y aplicando la espectroscopía de dispersión de energía de rayos X, (EDS). Los porcentajes del residuo remplazando el cemento fueron 0%, 5%, 10%, 15%, y 20%, en peso. Para los ejemplares de hormigón modificado con las concentraciones del polvo fueron evaluadas la resistencia a compresión, peso específico, absorción de agua, e índice de vacíos. Los resultados mostraron que 5% del polvo residual mejoraron las propiedades físicas y mecánicas del hormigón con relación a los ejemplares de referencia. Por lo tanto, el polvo puede actuar como filer en la pasta de cemento, disminuyendo el número de vacíos e aumentando la densidad y resistencia del material.

Palabras clave: polvo residual; residuos orgánicos; cemento; concreto; propiedades mecánicas.

1. Introduction

The construction industry uses approximately 14% to 50% of the natural resources extracted and is considered to be the second most significant contributor of carbon dioxide emissions [1]. As such, given the increasing environmental concerns, alternative materials using waste materials are a workaround solution to achieve sustainability in the construction industry.

Municipal solid waste (MSW) is defined as materials that are generated as a result of numerous activities in areas with human settlements [2]. Currently, incineration is one of the most commonly used solutions for waste management in developed and industrialized countries. For example, in Europe, the percentage of MSW incinerated has grown from approximately 15% in 1995 to 20% in 2009 [3]. In Japan, about 80% of MSW is being incinerated and reused [4]. China is continuing to build incineration plants for MSW each year in cities like Shenzhen,

Beijing, and Shanghai in order to generate electricity [5]. Countries like Belgium, Holland, Germany, and France have established laws and regulations for using municipal waste incinerated in construction [6].

In principle, the combustion of solid waste during the incineration process generates ashes, a new form of residue. This has been investigated as a potential raw material for the manufacture of cement by Lin et al. [7], Singhal et al. [8], Chen et al. [9] and Rodriguez et al. [10]. In Al-Rawas et al.'s research [11], incinerator ash was used at 0%, 10%, 20% and 30% replacement by weight for cement and with a w/c ratio of 0.70. Specimens prepared using 20% incinerator ash replacement for cement yielded a higher compressive strength than the control mix after 14 and 28 days of curing. Wu et al. [12] analyzed the feasibility of adding incinerated ash in the production of Portland cement. According to their study, good quality clinker can be obtained by heating the crude mixture at 1200 °C for 2 h with a 30% replacement of ash.

Incineration is a common practice to tackle the solid waste disposal issue. However, a chemical decomposition process may be an alternative method for this problem. Some researchers have studied this process to treat urban waste such as Jeronimo [13], Carneiro [14], Goyal et al. [15] and Kejun et al. [16]. This process is currently only being applied only to organic wastes. At the end of the process a residual powder is formed. As such, one of the potential uses of this recycled material is in the construction industry. An economic recycling approach will favor the effective use of nonrenewable natural resources and energy and prevent the dumping of toxic materials in the environment.

This paper discusses the suitability of the destination for residual powder obtained from organic waste, and the feasibility of incorporating the residual powder to replace a cement part in concrete. First, the chemical composition and particle size of the residual powder has been analyzed using scanning transmission electron microscopy (STEM) and Energy dispersive X-ray spectroscopy (EDS). Based on the results obtained, we assessed whether the residual powder is inert, noncontaminating, and suitable for use in concrete. Then, the residual powder was incorporated in the concrete with 5%, 10%, 15%, and 20% ratios. During the analysis, the water to cement ratio (w/c) and aggregates to cement ratio (a/c) were 0.70 and 10, respectively. Moreover, we analyzed the physical and mechanical properties of the concrete that were obtained by replacing a part of cement with the residual powder. According to the results obtained, incorporating 5% residual powder in cement will favor an increase in the density of the mixture and also improve the mechanical properties of the concrete. This could be ascribed to the small size of the particles that tend to occupy the voids in the cement paste.

2. Materials and experimental methodology

2.1. Materials and their properties

The cement used in this study was CPV ARI RS (high early strength Portland cement sulfate resisting), produced by Holcim. The physical, chemical, and mechanical characteristics of the cement, as summarized in Table 1, comply with the requirements of the Brazilian standard for cement [17].

Table 1.
Physical, chemical, and mechanical properties of the cement CP V ARI RS

Properties	Chemical composition (%)	Physical and Mechanical properties of cement and limits of Brazilian standards specifications for cement		
		Property	Test results	Limits of NBR.
SiO_2	19.20			
Al_2O_3	5.97	Fineness-Blaine measurement (cm^2/gm)	4 743	≥3000
Fe_2O_3	3.03	Fineness-Retained # 200 (%)	0.07	≤6.0
CaO	63.47	Initial setting time (minutes)	110	≥60
MgO	0.59	Final setting time (minutes)	163	≤600
SO_3	2.76	Le Chatelier Expansion (mm)	2.21	≤5.0
K_2O	0.83	Specific gravity	3.12	-
Na_2O	0.17	3 days Compressive strength (Mpa)	44.40	≥24
Free CaO	2.19	7 days Compressive strength (Mpa)	47.90	>34
Alkali equivalent	0.71	28 days Compressive strength (Mpa)	52.20	-

Source: The authors

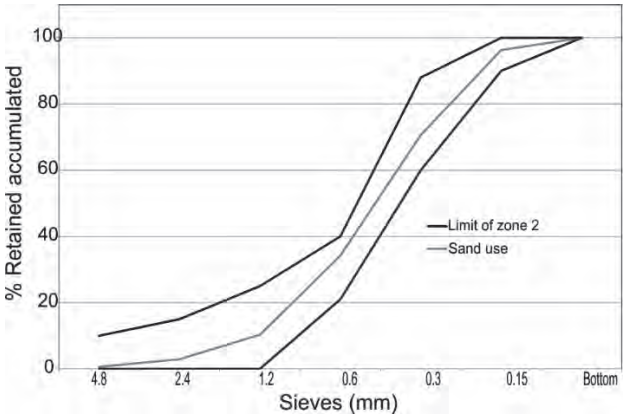


Figure 1. Comparison between the sand used in this study and that recommended by the Brazilian standard NBR 7211, 1987.
Source: Adapted from [18]

The fine aggregates used were sand and stone powders of basaltic origin, while the coarse aggregate is the zero graduation crushed rock basalt . These materials are readily available on the market and were used under the same conditions in which they are usually employed during construction.

According to the results of the granulometric analysis, the sand used in this study meets the requirements of the Brazilian standard [18] and can be classified as fine sand in zone 2, as shown in Fig. 1.

Fig. 2 shows the granulometric distribution of the stone powder, which conforms with the specifications recommended by the Brazilian standard [18] for the granulometric range of zone 4. Hence, it can be regarded as coarse sand.

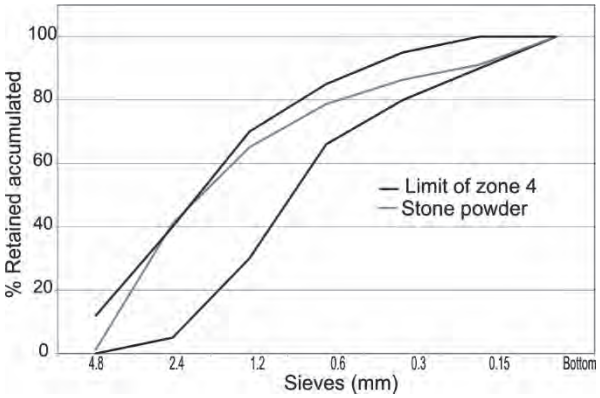


Figure 2. Comparison between the stone powder used in this study and that specified by the Brazilian standard NBR 7211, 1987.
Source: Adapted from [18]

Table 2.
Physical characteristics of sand and stone powder

Materials	Specific gravity (g/cm^3)	Bulk density (kg/m^3)	Fineness modulus	Maximum diameter (mm)	Powdery Material (%)
Sand	2.62	2.59	2.15	2.40	0.31
Stone powder	2.80	2.80	3.64	4.80	12.58

Source: The authors

Table 3.
Physical properties of the gravel

Properties	Sieves (mm)	Gravel
% Retained accumulated	12.5	0.0
	9.5	0.5
	6.3	48.3
	4.8	86.0
	2.4	98.8
	Bottom	100.0
Maximum diameter (mm)	9.50	
Fineness modulus	5.85	
Specific gravity (g/cm3)	2.86	
Bulk density (g/cm3)	2.70	

Source: The authors

The results of the characterization of the fine aggregates are presented in Table 2.

Table 3 shows the physical properties of the gravel.

2.2. Residual powder from organic waste

Furthermore, we analyzed the chemical and physical properties of the residual powder generated from the organic waste, including the chemical composition, particle size, analysis of heavy metals, bulk density, fineness modulus, and particle size. Table 4 summarizes the different tests and the methodology adopted for each of the trials.

2.3. Generation of residual powder from organic waste

In a typical process, all the received urban waste materials were placed on a magnetic mat for the extraction of metallic materials. Subsequently, all the inorganic materials, such as wood, glass, plastics, rubber, and batteries, were manually segregated, leaving behind the organic residues. The inorganics separated were to be recycled, while the organics were to be treated.

Table 4.
Tests and methodology adopted to characterize the organic residual powder

Material	Test	Experimental Methodology
Residual powder from organic waste	Analysis of the shape and size	Scanning Transmission Electron Microscope (STEM)
	Chemical analysis	Energy Dispersive X-ray Spectrometer (EDS)
	Analysis of heavy metals	Atomic emission spectroscopy (AES)
	Specific gravity (g/cm3)	NM 23. 2001
	Specific surface area	NBR 11579. 2012
	Granulometric analysis	NM 248. 2003

Source: The authors

Figure 3. Process adopted to obtain the powder: (a) mass withdrawal from the reactor and (b) conversion into a powder.
Source: The authors

The organic wastes were then passed through a grinding mill to reduce the particle size at the end of the belt. Following this, they were passed through a chamber for chemical reagents to be added. This enabled the stabilization of fatty waste, manure, and animal and human waste, as well as the chemical encapsulation of heavy metals; the corrosion process was thus accelerated.

The resulting mixture of chemical reagents and organic wastes was subjected to chemical reaction for approximately 40 min, until it turned into a mass that was free of chemical, biological, and pathogenic contaminants. This mass was removed from the reactor and dried outdoors. Subsequently, it was micro-ionized and transformed into a powder, as shown in Fig. 3.

2.4. Mixture proportion

To evaluate the feasibility of replacing the cement with the obtained residual powder, we prepared 5 different concrete mixtures by adding 0%, 5%, 10%, 15%, and 20% residual powder by weight as a cement replacement. For all the compositions, the humidity was set as 6% of the total amount of dry materials. The percentage of additive was maintained constant at 0.35% with a 0.70 w/c ratio. The proportion of aggregates with cement ratio (a/c) was set as 10, and the content of mortar was maintained constant at 54.5%. Details on the mixture design are provided in Table 5.

Table 5.
Mix proportion for 1 m3 of concrete (kg)

Materials	Refer.	5%	10%	15%	20%
Cement	205.10	194.85	184.60	174.34	164.08
Organic powder	-	10.25	20.50	30.76	41.02
Gravel			1025.6		
Stone dust			615.39		
Sand			410.21		
H_2O	143.60	136.40	129.20	122.00	114.90
w/c			0.70		
Proportion of aggregates (cement aggregate)			1:10		
Additive in relation to the mass of cement (%)			0.35		
Content of mortar (%)			54.50		
Moisture			6.00		

Source: The authors

Figure 4. Cylindrical specimens of dimension 5 cm x 10 cm.
Source: The authors

2.5. *Mixture and analysis of concrete*

For the analysis, cylindrical specimens with a 5 cm diameter and 10 cm height were produced, as shown in Fig. 4. A total of 8 specimens were prepared for each type of concrete.

In a typical process, the raw materials were mixed together at a temperature of 25±3 °C and relative humidity of 55±15% for 5 min using a mechanical mixer (G.Paniz) to obtain a homogeneous mixture. The photographs of the mixing process are shown in Fig. 5.

The order of collocation of the materials is as follows:

1st step–mixing and homogenization of the coarse and fine aggregates,

2nd step–addition of half the amount of water required for the process,

3rd step–addition of the cement and the homogenization of the mixture,

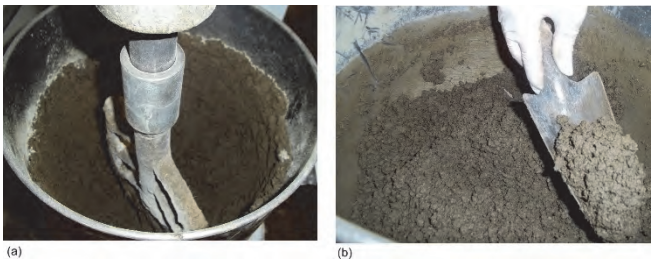


Figure 51. Mixing process: (a) mixing and homogenization of the materials and (b) final homogenization of the mixture.
Source: The authors

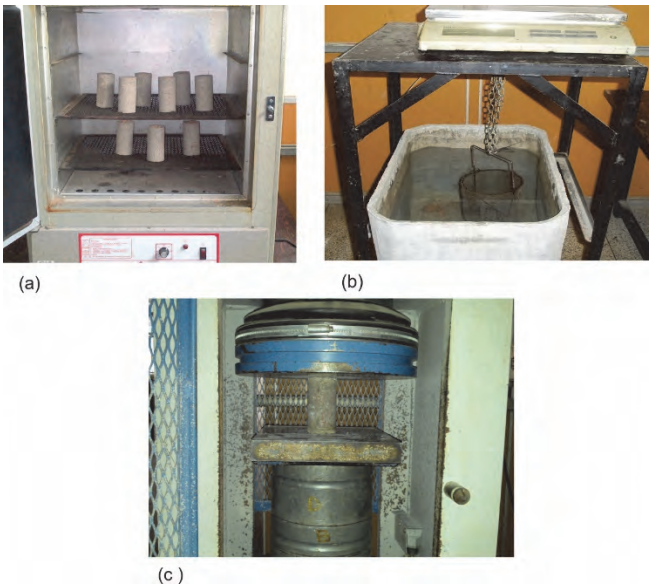


Figure 6. Equipment used for assessing the physical properties of the specimens: (a) drying of the specimen in the greenhouse, (b) determining the mass of the specimen using a hydrostatic weighing-machine and (c) compression test in the ELE machine.
Source: The authors

4th step–addition of the remaining water, along with the additive,

5th step–addition of the residual powder obtained from the organic waste,

6th step–final homogenization of the mixture.

We evaluated the physical characteristics of the concrete specimens, including density, water absorption, and void content, according to the Brazilian standard [19]. In addition, the surface texture and compressive strength of the concrete specimens were determined according to the Brazilian standard [20]. The specimens were dried in the greenhouse (Quimis). The density of the specimens was evaluated using a hydrostatic weighing-machine (Mettler, accurate to 0.1 g, maximum capacity of 24 000 g). The compressive strength of the specimen was determined using a computerized machine (ELE) with an Autotest model. The surface structure of the specimens was rectified using an appropriate processing machine to obtain regular and smooth surfaces. Fig.6 presents the details of these tests and the equipment used.

3. Results and discussion

3.1. Organic residual powder

The microstructure of the residual powder was observed using STEM. The corresponding images are shown in Fig. 7. The residual powder particles are composed of grains of different sizes and shapes. Some have a circular aspect, while some are lamellar with overlapping layers. Moreover, small particles are adhered on the surface.

The images obtained with the Image J program were used to measure the particle size. The corresponding results are summarized in Table 6. The particles in the residual powder have a broad size distribution is especially beneficial for using this material as concrete filler as it is expected to allow better filling of the voids in the cementitious matrix.

The chemical composition of the powder was determined using EDS. Fig. 8 illustrates the EDS spectra of three different organic residual powder samples. Table 7 summarizes the corresponding results obtained by averaging the values detected from three different organic residual powder samples. The proportions of sodium, magnesium, aluminum, silicon, phosphorus, sulfur, and chlorine are lower than 1%. The most abundant elements are carbon, oxygen, and calcium. Carbon is the fundamental element present in all animal and vegetable origin substances, while the abundance of oxygen indicates the nonmetallic and highly reactive nature of the material that easily forms compounds, especially oxides of the material.

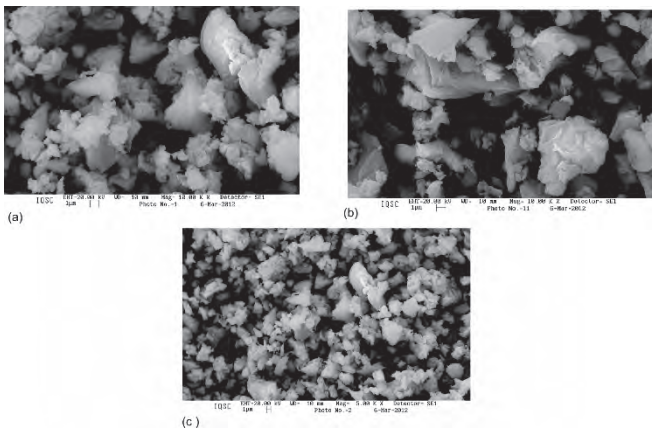


Figure 2. STEM images of the organic residual powder: (a) sample 1, (b) sample 2, and (c) sample 3.
Source: The authors

Table 6.
Particle size of the organic residual powder

	Sample 1 (µm)	Sample 2 (µm)	Sample 3 (µm)
	3.88	2.94	4.56
	3.50	4.36	6.37
	5.03	2.08	3.82
	5.31	2.75	2.79
	1.89	5.27	7.31
	2.18	2.26	3.69
	2.89	2.95	2.16
	1.75	1.59	1.34
	2.74	2.07	1.47
	1.70	2.89	1.84
	5.46	1.63	2.13
	5.17	3.06	1.69
	2.95	4.56	3.95
Average	3.28	3.16	3.26

Source: The authors

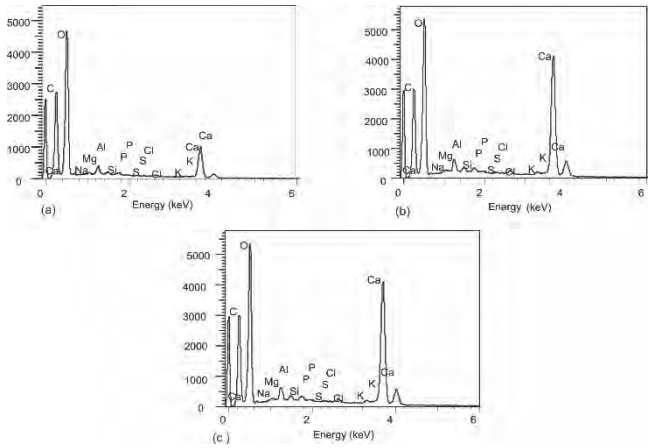


Figure 8. EDS spectra of three different samples of organic residual powder: (a) sample 1, (b) sample 2, and (c) sample 3.
Source: The authors

Table 7
Chemical composition of the organic residual powder

Element Type	% Element (by mass)	% Atoms
C	21.37	29.50
O	58.19	61.00
Na	0.28	0.21
Mg	1.19	0.83
Al	0.21	0.13
Si	0.32	0.19
P	0.07	0.04
S	0.06	0.03
Cl	0.34	0.17
Ca	18.00	7.90
Total	100	100

Source: The authors

Further analysis of heavy metals in the residual powder reveals the presence of different elements, listed in Table 8; < LOD (lower limit of detection) implies a very low concentration of the metal that is below the limit of precise quantification. The elements chromium, arsenic, cadmium, lead, and mercury, which are typically considered to be heavy elements, are found in very low concentrations. This

Table 8
Heavy metals present in the organic residual powder

Metals	Values (mg/l)	Maximum Limit (mg/l)
Copper	7.76	-
Chrome	< LOD	5.0
Arsenic	< LOD	1.0
Cadmium	< LOD	0.5
Lead	< LOD	1.0
Cobalt	< LOD	-
Phosphorus	72.59	-
Nickel	< LOD	-
Zinc	15.41	-
Mercury	< LOD	0.1

Source: The authors

Table 9.
Physical characterization of the organic residual powder

Physical Properties	Values
Fineness-Retained # 200 (%)	0.20
Fineness-Retained # 325 (%)	1.50
Specific surface (cm^2/g)	4.37
Specific gravity (g/cm^3)	0.72

Source: The authors

Table 10.
Physical characteristics of concrete with the organic residual powder

Concrete	Specific density – hardened state (g/cm3)			Water absorption (%)			Void index – hardened state (%)		
	Mean	Sd	CV (%)	Mean	Sd	CV (%)	Mean	Sd	CV (%)
Ref.	2.78	0.051	1.88	4.81	0.23	4.96	10.85	0.49	4.55
5% w. p	2.81	0.047	1.67	4.76	0.25	5.25	10.81	0.45	4.16
10% w.p	2.76	0.052	1.88	4.80	0.27	5.62	10.86	0.50	4.60
15% w.p	2.69	0.050	1.85	5.02	0.28	5.54	10.92	0.41	3.75
20% w.p	2.67	0.045	1.68	5.94	0.30	5.05	11.11	0.46	4.14

Source: The authors

complies with the limits highlighted in Appendix F of the Brazilian standard [21], which identifies hazardous waste. Based on these results, the residual powder obtained from organic waste can be rated "Not Dangerous Class II A Inert waste."

Table 9 shows the test values with respect to the physical characterization of the organic residual powder. The test for specific gravity, specific surface, and granulometric analyses were performed according to the Brazilian standards [22], [23] and [24], respectively.

3.2. Concrete

Table 10 shows the physical characteristics of the concrete with 0%, 5%, 10%, 15%, and 20% levels by weight of residual powder substituting the cement. The results include the mean, standard deviation (Sd), and coefficient of variation (CV) for each of the concrete.

The concrete with 5% organic residual powder exhibited a higher specific mass value. Hence, the concrete had a smaller value of water absorption and void content due to the refinement of the pores in the cement paste. The concrete with a higher ratio of organic residual powder (10%, 15%, and 20%) had lower performance in terms of its physical properties when compared with the reference concrete. This could be attributed to the less packaging that the mixture had.

Furthermore, we performed a detailed analysis of the compression strength of concrete using the Bootstrap technique. With this study, we obtained confidence intervals for the mean, standard deviation, and coefficient of variation of the study population with a 95% reliability, as shown in Table 11. Confidence intervals for populations were more informative than a result of only one value for a reduced number of observations because they provided a range of plausible values for an unknown parameter. Thus, one can better interpret the results and form the basis for decisions with a minimum level of risk. The Bootstrap method was applied with 5,000 resampling. The procedure is described in Appendix A.

Table 11.
Compressive strength of the concrete with organic residual powder

Concrete	Compressive strength (MPa)					
	Average		Sd		CV (%)	
	Lower	Upper	Lower	Upper	Lower	Upper
Reference	32.96	34.28	0.65	1.17	1.97	3.41
5% waste powder	33.69	34.98	0.51	1.33	1.51	3.80
10% waste powder	27.40	28.95	0.66	1.60	2.40	5.52
15% waste powder	19.31	20.63	0.63	1.17	3.26	5.67
20% waste powder	9.54	10.93	0.47	1.24	4.92	11.34

Source: The authors

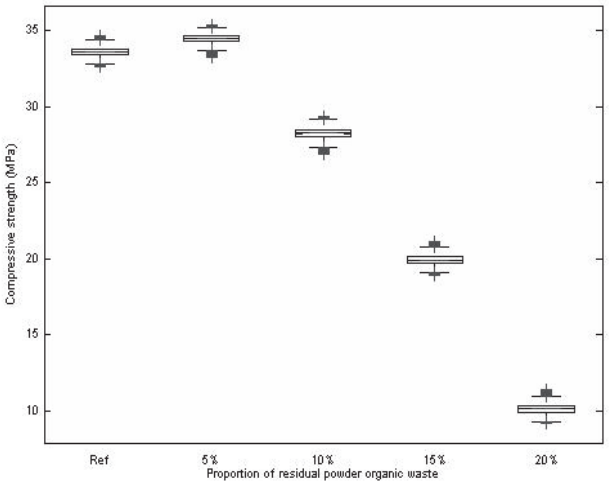


Figure 9. Box plot of the concrete's compression strength data with organic residual powder.
Source: The authors

The concrete with 5% residual powder showed superior results, and was better than that of the reference specimen. On the other hand, the use of higher concentrations of residual powder did not show satisfactory results, with the resistance being lower than in the concrete without organic residue.

For a better understanding of the results, we have made a box plot of the average resistance values of each concrete, which is shown in Fig. 9. The addition of 5% organic residual powder by weight that replaced the cement improved the cement binder; the size was decreased as was the number of pores. This resulted in a stronger concrete with lesser porosity.

The difference between the two means was analyzed using the Bootstrap technique. Therefore, the magnitude of these differences will be evaluated showing the number of times that the average compressive strength was statistically equal and different for a confidence level of 95%. This procedure is detailed in Appendix B.

Table 12 shows the percentage probability that the null hypotheses H01, H02, and H03 were accepted. In the matrix,

Table 12.
Percent probability that the null hypotheses were accepted

Conc/Conc	Ref.			5% w. p.			10% w. p.			15% w. p.			20% w.p		
	H01	H02	H03	H01	H02	H03	H01	H02	H03	H01	H02	H03	H01	H02	H03
Ref.	0	100	0	60.5	54.2	0.02	0	0	100	0	0	100	0	0	100
5% w.p.	0.02	54.2	60.5	0	100	0	0	0	100	0	0	100	0	0	100
10% w.p.	100	0	0	100	0	0	0	100	0	0	0	100	0	0	100
15% w.p.	100	0	0	100	0	0	100	0	0	0	100	0	0	0	100
20% w. p.	100	0	0	100	0	0	100	0	0	100	0	0	0	100	0

Source: The authors

each value corresponds to two populations, namely, u1 located in the row and u2 in the column. The percent value of 100% always lies on the diagonal to the null hypothesis H02 because it is compared with the same population. The probabilities of each of the null hypotheses do not necessarily add up a total of 100% given the fact that there may be cases in which the nulls hypotheses H01, H02 or H03 are accepted the same time.

The probability of compressive strength for the 5% concrete being statistically higher than the reference concrete is 60.58%, and being statistically equal is 54.22%, while the probability of the compressive strength being statistically lower is 0.02%. Therefore, based on the results of the hypotheses test, it can be concluded that the concrete modified with 5% residual powder has a better mechanical behavior.

On the other hand, it can also be concluded that the compressive strength in the reference concrete is statistically higher than the 10%, 15%, and 20% modified concretes.

The concrete modified with 5% residual powder is different from the concrete modified with higher concentrations of residual powder. This can be seen by observing the values in the 2nd row of the columns 3, 4, and 5 in Table 12. This shows that the null hypothesis H_{03} ($u_1 > u_2$) has 100% probability of strength to be statistically higher.

Fig. 10 shows the surface texture of the reference specimen and those modified with the organic residual powder. Compared with the reference specimen, the surface of concrete with 5% residual powder is relatively smooth and less porous. This implies that the fine residual powder at this specific concentration of 5% smoothens the texture of concrete and therefore, makes the elements less permeable, more resistant to aggressive environmental conditions, and esthetically appropriate.

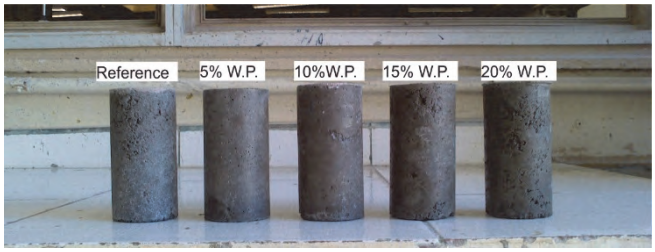


Figure10. Surface texture of the specimens
Source: The authors

4. Conclusions

- Residual powder obtained from organic waste had particles of varying sizes and shapes, which is expected to facilitate the better filling of void spaces in the cementitious matrix. Moreover, the concentration of heavy metals was found to be very low, making it a safe and nontoxic solid residue.
- The evaluation of the concrete's properties modified with the organic residual powder indicates that the powder does not present caking properties.
- Concrete modified with 5% organic residual powder exhibited superior compressive strength, better than that of the reference concrete. This implies that the correct packaging of the concrete with 5% residual powder leads to refined pores in the cement paste.
- The compressive strength of the concrete specimens modified with 10%, 15%, and 20% residual powder was lower than that of the reference concrete.
- The organic residual powder, when added at the proportion of 5%, acts as filler for concrete. This fills the voids, decreases the porosity, achieves a smooth texture, and provides an esthetically appropriate structure.
- Future studies to evaluate the same percentages of residual powder in different proportions of cement in concrete are underway.

Acknowledgements

The authors would like to thank the Department of Structural Engineering at the School of Engineering of São Carlos at the University of São Paulo for their permission to conduct the survey. The authors would also like to thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico-CNPq (Procs. # 142374/2011-3) and Fundação de Amparo à Pesquisa do Estado de São Paulo-FAPESP (Procs. # 2012/21541-7) for the financial support that they provided for this research.

Appendix A. Bootstrap Technique

Bootstrap technique, as proposed by Efron [25], is a computational method that simulates the experimentation by reusing the original data. In essence, the Bootstrap method randomly takes information from the original data and creates a new dataset. This procedure is called resampling. In this way, different statistics can be computed in each dataset, such as the sample mean and sample variance. The resampling and re-computation of the statistics is performed several times and, hence, can be considered to be repeated experiments. Therefore, by using the different statistics calculated in each resampling, the Bootstrap technique evaluates the distribution of the statistical parameter and estimates the confidence interval [26].

When considering the problem of finding the confidence interval of a statistical parameter θ, the Bootstrap assumes that the original data $X = \{X1, X2,\ldots, Xn\}$ itself constitutes the unknown distribution from the random variable X. Therefore, by resampling numerous times from X and computing $\hat{\theta}$ for each of these resamples, it is possible to

Table A.1.
Bootstrap procedure for calculating the confidence interval using the percentile method

Step 0: Conduct the experiment. Suppose that the original data is given by.

$$X=\{X_1,X_2,X_3...,X_n\}_{1xn}$$

Step 1: Resampling. Draw a random sample of *n* values, with replacement, from

X. the Bootstrap resample is.

$$X^*=\{?,?,?...,?\}_{1xn}$$

In this step is possible that some of the original data appear more than once and others not at all.

Step 2. Calculation of the Bootstrap estimate. Calculate the statistic $\hat{\theta}^*$, which can be the mean, the variance, etc.

Step 3. Repetition. Repeat Steps 1 and 2 a large number of times to obtain a total

of B Bootstrap estimates $\hat{\theta}_1^*,\hat{\theta}_2^*.....\hat{\theta}_B^*$.

Step 4. Confidence interval with percentile method. The $100(1-\alpha)\%$ Bootstrap confidence interval is $\left(\hat{\theta}_{q_1}^*,\hat{\theta}_{q_2}^*\right)$,where $q_1=B\alpha/2$, $q_2=B-q_1+1$ and α is of significance level.

Source: The authors

obtain a Bootstrap distribution for $\hat{\theta}$, which approximates the distribution of θ. Using this, a confidence interval can be derived for θ. This procedure is described in Table A.1.

Appendix B. Inference on the difference between the two means estimated using the Bootstrap method

Traditionally, the test of hypotheses is used to study the difference in the parameters of two populations. In essence, this method involves testing the equality of two parameters, according to a predetermined threshold probability. Thus, a significance level is defined to create a critical region. This critical region defines the set of all outcomes, which, if they occur, will lead us to decide that there is a difference between the parameters (negative result). In contrast, an outcome outside of the critical region will lead us to decide that there is no difference between the parameters (positive result) [26].

The Bootstrap method can estimate a range of likely values of the unknown population parameter, instead of only one value. Therefore, by calculating the hypothesis test of each value of the range, it is possible to evaluate the probability of either acceptance or rejection of the null hypotheses. This probability is more informative than a simple result of the hypothesis test, which either rejects or does not reject the null hypotheses.

The test of hypothesis considers the difference in means u1 and u2 of two normal distributions, where the variances σ1 and σ2 are unknown and not necessarily equal, i.e., σ1 ≠ σ2. Different null hypotheses, such as H01: u1<u2, H02: u1=u2 and H03: u1>u2, can be analyzed. A t-statistic is used to test these hypotheses. The solution obtained using the four-step hypothesis-testing procedure is as shown in Table B.1.

Table B.1.
Hypothesis test for the difference between two means

Step 0. Evaluate the parameters of interest $\hat{u}_1$, $\hat{u}_2$, $\hat{\sigma}_1$, $\hat{\sigma}_2$, n_1 and n_2 of original data,

Where:

n_1 and n_2 are sample size of populations 1 and 2, respectively.

Step 1. Null hypothesis $H_{01}:u_1<u_2$, $H_{02}:u_1=u_2$ and $H_{03}:u_1>u_2$

Step 2. Define significance level α .

Step 3. The test statistic.

$$t_0=\frac{\hat{u}_1-\hat{u}_2}{\sqrt{\frac{\hat{\sigma}_1^{\,2}}{n_1}-\frac{\hat{\sigma}_2^{\,2}}{n_2}}}$$

with degrees of freedom given by

$$v=\frac{\left(\frac{\hat{\sigma}_1^{\,2}}{n_1}+\frac{\hat{\sigma}_2^{\,2}}{n_2}\right)^2}{\frac{\left(\hat{\sigma}_1^{\,2}/n_1\right)^2}{n_1-1}+\frac{\left(\hat{\sigma}_2^{\,2}/n_2\right)^2}{n_2-1}}$$

Step 4. Accept H_{01} if $t_0<-t_{\alpha,v}$

Accept H_{02} if $t_0>-t_{\alpha/2,v}$ or if $t_0<t_{\alpha/2,v}$

Accept H_{03} if $t_0>t_{\alpha,v}$

Source: The authors

Table B.2.
Probability of the null hypotheses with the Bootstrap method

Step 0: The original data are given by.

$$X_1=\{X_{11},X_{12},X_{13}...,X_{1n}\}_{1xn_1}$$

$$X_2=\{X_{21},X_{22},X_{23}...,X_{2n}\}_{1xn_2}$$

Step 1: Resampling. Draw a random sample from X_1 and X_2, the Bootstrap resamples are:

$$X_1^*=\{?,?,?...,?\}_{1xn_1}$$

$$X_2^*=\{?,?,?...,?\}_{1xn_2}$$

Step 2. Calculate the statistics $\hat{u}_1$, $\hat{u}_2$, $\hat{\sigma}_1$ and $\hat{\sigma}_2$.

Step 3. Decide to either reject the null hypothesis or not reject it, according to Table B.1.

If the null hypotheses is true H_{01} then $Z_1=Z_1+1$.

If the null hypotheses is true H_{02} then $Z_2=Z_2+1$.

If the null hypotheses is true H_{03} then $Z_3=Z_3+1$.

where Z_1,Z_2 and Z_3 are counters of the number of times that the null hypotheses H_{01}, H_{02} and H_{03} are true. The initial value of Z_1, Z_2 and Z_3 are equal to zero.

Step 3. Repetition. Repeat Steps 1, 2 and 3 a large number of times to obtain a total

of B Bootstrap estimates.

Step 4. Calculate the probability of the null hypothesis H_{01}, H_{02} and H_{03}, as.

$$PH_{01}(\%)=100*\frac{Z_1}{B}\ ,\ PH_{02}(\%)=100*\frac{Z_2}{B}\ ,\ PH_{03}(\%)=100*\frac{Z_3}{B}$$

Source: The authors

Now, by resampling the populations X1 and X2 several times to compute $\hat{u}1$, $\hat{u}2$, $\hat{\sigma}1$, and $\hat{\sigma}2$ and by applying the hypothesis test for each of these resamples, it is possible to evaluate the probability of the different null hypotheses, as shown in Table B.2.

References

[1] Fioriti, C.F., Avaliação de compósitos de concreto com resíduos de borracha na produção de blocos para alvenaria. Tesis, Universidade Estadual Paulista, São Paulo, Brasil, 2002.

[2] Zanta, V.M. y Ferreira, C.F.A., Gerenciamento integrado de resíduos sólidos urbanos. Resíduos Sólidos Urbanos: Aterro Sustentável para Municípios de Pequeño Porte,1, pp. 1-16, 2003.

[3] Gori, M., Bergfeldt, B., Reichelt, J. and Sirini, P., Effect of natural ageing on volume stability of MSW and wood waste incineration residues. Waste Management, 33(4), pp. 850-857, 2013. DOI: 10.1016/J.WASMAN.2012.12.005.

[4] Jung, CH., Matsuto, T., Tanaka, N. and Okada, T., Metal distribution in incineration residues of municipal solid waste (MSW) in Japan. Waste Management, 24(4), pp. 381-391, 2004. DOI: 10.1016/S0956-053x(03)00137-5.

[5] Cheng, H. and Hu, Y., Municipal solid waste (MSW) as a renewable source of energy: Current and future practices in China. Bioresource Technology, 101(11), pp. 3816-3824, 2010. DOI: 10.1016/J.BIORTECH.2010.01.040.

[6] Van Gerven, T., Geysen, D., Stoffels, L., Jaspers, M., Wauters, G. and Vandecasteele, C., Management of incinerator residues in Flanders (Belgium) and in neighbouring countries. A comparison. Waste Management, 25(1), pp.75-87, 2005. DOI: 10.1016/J.WASMAN.2004.09.002.

[7] Lin, KL., Wang, KS., Lee, TY. and Tzeng, BY., The hydration characteristics of MSWI fly ash slag present in C3S. Cement and Concrete Research, 33(7), pp. 957-964, 2003. DOI: 10.1016/S0008-8846(02)01002-5.

[8] Singhal, A., Tewari, VK. and Prakash, S., Utilization of treated spent liquor sludge with fly ash in cement and concrete. Building and Environment, 43(6), pp. 991-998, 2008. DOI: 10.1016/J.BUILDENV.2007.02.002.

[9] Chen, H.X., Ma, X.W. and Dai, H.J., Reuse of water purification sludge as raw material in cement production. Cement & Concrete Composites, 32(6), pp. 436-439, 2010. DOI: 10.1016/J.CEMCONCOMP.2010.02.009.

[10] Rodriguez, N.H., Martinez-Ramirez, S., Blanco-Varela, M.T., Guillem, M., Puig, J. and Larrotcha, E., Evaluation of spray-dried sludge from drinking water treatment plants as a prime material for clinker manufacture. Cement & Concrete Composites. 33(2), pp. 267-275, 2011. DOI: 10.1016/J.CEMCONCOMP.2010.10.020.

[11] Al-Rawas, A.A., Hago, A.W., Taha, R. and Al-Kharousi, K., Use of incinerator ash as a replacement for cement and sand in cement mortars. Building and Environment. 40(9) pp. 1261-1266, 2005. DOI: 10.1016/J.BUILDENV.2004.10.009.

[12] Wu, K., Shi, H.S., De Schutter, G., Guo, X.L. and Ye, G., Preparation of alinite cement from municipal solid waste incineration fly ash. Cement & Concrete Composites. 34(3), pp. 322-327, 2012. DOI: 10.1016/J.CEMCONCOMP.2011.11.016.

[13] Jeronimo, F.dP., Stabilization of organic waste process and product thus obtained. Brasil Patent. 2013.

[14] Carneiro, M.S., Processo otimizado de beneficiamento de lixo urbano. Brasil Patent. 2013.

[15] Goyal, S., Dhull, S.K. and Kapoor, K.K., Chemical and biological changes during composting of different organic wastes and assessment of compost maturity. Bioresource Technology, 96(14), pp. 1584-1591, 2005. DOI: 10.1016/J.BIORTECH.2004.12.012.

[16] Kejun, S., Juntao, Z., Ying, C., Zongwen, L., Lin, R. and Cong, L., Accelerating the degradation of green plant waste with chemical decomposition agents. Journal of Environmental Management, 92(10), pp. 2708-2713, 2011. DOI: 10.1016/J.JENVMAN.2011.06.011.

[17] Associação Brasileira de Normas Técnicas (ABNT). NBR 5733: Cimento Portland de Alta Resistência Inicial. Rio de Janeiro, Brasil, 1991.

[18] Associação Brasileira de Normas Técnicas (ABNT). NBR 7211: Agregados para Concreto - Especificação. Rio de Janeiro, Brasil, 1987.

[19] Associação Brasileira de Normas Técnicas (ABNT). NBR 9778: Argamassas e Concretos Endurecidos: Determinação da Absorção de Água por Imersão – Índice de Vazios e Massa Específica. Rio de Janeiro, Brasil, 2005.

[20] Associação Brasileira de Normas Técnicas (ABNT). NBR 5739: Concreto - Ensaios de Compressão de Corpos-de-prova Cilíndricos. Rio de Janeiro, Brasil, 1991.

[21] Associação Brasileira de Normas Técnicas (ABNT). NBR 10004: Resíduos sólidos - Classificação. Rio de Janeiro, Brasil, 2004.

[22] Associação Brasileira de Normas Técnicas (ABNT). NBR NM 23: Cimento Portland e Outros Materiais em Pó - Determinação da Massa Específica. Rio de Janeiro, Brasil, 2001.

[23] Associação Brasileira de Normas Técnicas (ABNT). NBR 11579: Cimento Portland - Determinação do Índice de Finura por Meio da Peneira 75 μm (nº 200). Rio de Janeiro, Brasil, 2012.

[24] Associação Brasileira de Normas Técnicas (ABNT). NBR NM 248: Agregados - Determinação da Composição Granulométrica. Rio de Janeiro, Brasil, 2003.

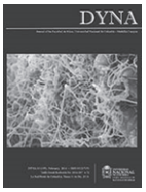
[25] Efron, B. and Tibshirani, R., Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Statistical Science. pp. 54-75, 1986.

[26] Montgomery, D.C. and Runger, G.C., Applied statistics and probability for engineers. John Wiley & Sons; 2010.

I.S. Izquierdo, received her BSc. Eng in Civil Engineering in 2008, from the Polytechnic Institute Jose Antonio Echeverria (ISPJAE), Havana-Cuba his MSc. in Structural Engineering in 2011 and her PhD. in Structural Engineering in 2015, both from the School of Engineering of São Carlos at the University of São Paulo, Brazil. She is currently a full professor in the Exact and Technological Science Department, Universidad Estadual de Santa Cruz- UESC, Brazil. She has experience in Civil Engineering with emphasis on Structural Concrete and Masonry, and conducts research into the following topics: waste in use of concrete, concrete technology and structural masonry.
ORCID: 0000-0001-7103-4353

M.A. Ramalho, received his BSc. Eng in Civil Engineering in 1980, his MSc. in Structural Engineering in 1983, and his PhD. in Civil Engineering in 1990, all from the University of Sao Paulo, São Carlos, Brazil. He is currently a full professor in the School of Engineering at São Carlos of the University of São Paulo, Brazil. He has experience in Civil Engineering with emphasis on Structural Concrete and Masonry, and conducts research into the following topics: structural masonry, finite element, reinforced concrete, structural analysis of buildings.
ORCID: 0000-0001-6411-331X

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS



DYNA
http://dyna.medellin.unal.edu.co/

Detection and localization of potholes in roadways using smartphones

Diego Andrés Casas-Avellaneda [a] & Javier Francisco López-Parra [b]

[a] *Facultad Ingeniería, Pontificia Universidad Javeriana, Bogotá, Colombia. casas.diego@javeriana.edu.co*
[b] *Facultad de Ingeniería, Pontificia Universidad Javeriana, Bogotá, Colombia. jlopez@javeriana.edu.co*

Received: August 8th, 2013. Received in revised form: March 12th, 2014. Accepted: January 25th, 2016.

Abstract
A system for the detection and localization of roadway potholes is proposed and is aimed at collecting data on the roadway potholes using the accelerometer, GPS and compass that are embedded into smartphones using the Android operating system. This system has a scalable mobile architecture that contributes to capturing environment information using multiple devices with a large geographic coverage.

Keywords: Geographical Information Systems; Location Based Services; Potholes; Mobile Computing; Mobile Architecture.

Detección y localización de imperfecciones viales utilizando smartphones

Resumen
Un sistema de detección y localización de imperfecciones viales ha sido propuesto con el fin de realizar la recolección de datos de las imperfecciones viales por medio del acelerómetro, GPS y brújula que se encuentran embebidos en los teléfonos inteligentes con sistema operativo Android. Este sistema cuenta con una arquitectura móvil escalable que contribuye a capturar información del entorno desde múltiples dispositivos con una gran cobertura geográfica.

Palabras clave: Sistemas de Información Geográfica; Sistemas Basados en Localización; Imperfecciones Viales; Computación Móvil; Arquitectura móvil.

1. Introduction

With the help of geographic information technologies, it is possible to establish new-strategy based solutions that use information about the condition of roadways. These solutions aim to help people and entities who are responsible for performing preventive as well as corrective maintenance to paved roads through continuous surveillance using their smartphones which are connected to a cartographic server that eases location and quality control of the work done.

Based on this context, we set out the need to build a cheaper system, such as the one we currently use, taking full advantage of the smartphone's sensors and their processing capabilities, which swiftly allow large distances to be reached in a short time. The roadways that have a flow of mass transit were given special priority, such as the ones on which articulated buses carrying more than 150 passengers in just one vehicle circulate.

The limitations we faced when carrying out the project were related to the roadways where the tests were to be performed and the type of mobile devices that were to be used. We did not want the tests to be too expensive as all that was needed was an accelerometer and a GPS. The main objective was to measure potholes on roads.

The following article summarizes the experience of specifying, designing and implementing an information system called RSMS (Roadway State Monitoring System) developed in the city of Bogota, Colombia.

The system collects information on the vehicle's acceleration and the roadway's potholes combined with their geographic and spatial location by using mobile devices that run on Android. The information collected is stored,

catalogued according to type of pothole and classified using the scale VIZIR category B [1], which deals with the surface damage of the roadway caused by local factors.

The system is formed by sensors that capture information by using tools such as the accelerometer, compass and GPS. The information is stored on a centralized and georeferenced database that is linked to an app server with web access. The communication between the mobile devices and the server takes place through SOAP (Simple Object Access Protocol) protocol [2].

The analysis processes are undertaken with a geographic information system and a series of information outputs, similar to maps, with the potholes' locations are created, along with a description, their features, closer station and buffer (area of influence) [3] and cluster (markers quantification) [4] analysis.

A roadway corridor approximately 10 kilometers long (6 miles) in Bogota through which there is mass transit was selected for the case study. The suggested architecture presented a number of advantages and made the integration of these devices easier with the Geographic Information System (GIS) and the sensor's added value.

2. State of art

A Geographic Information System (GIS) is "a specific software that allows users to create interactive queries, integrate, analyze and represent in an efficient way any type of geographical information referenced associated to one territory, connecting maps with databases" [5]. GIS systems are widely used as these systems support decisions with information provided by Localized-Based Systems (LBS), which "Is a service that adds value to target locations provided by one localization service. It uses knowledge of a mobile device's location to offer value to mobile subscriber or to a third-party" [6].

Smartphone sensors allow information about daily activities to be detected and recognized [7]. It is important to highlight the sensors' technological development in order to measure accuracy. The main sensors used in this project were the accelerometer, GPS and compass.

The accelerometer is a tri-axial sensor and measures movement, gravity variations, and human movements. As a result of the device's accuracy level, it can detect minimum movements, gravity variations and human movements that can be processed to detect what type of activity is being undertaken [7].

GPS (Global Positioning System) is a circuit that uses the Global Navigation Satellite System (GNSS) to obtain the current position of the sensor with an accuracy of 5 meters or more, depending on the nearest satellites' signal reception. Connection to a Wireless Access Point can improve the accuracy [8].

The compass is a sensor that detects the earth's magnetic field. This sensor locates the terrestrial magnetic north and reads the intensity of magnetic field in micro-teslas. Using this sensor is possible detect the device's direction. Combining the GPS location and compass reading on the roadway, the direction will be associated with the localization to obtain the correct track of the roadway.

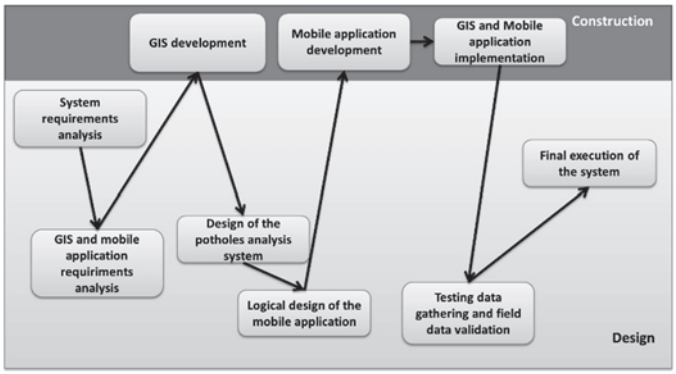


Figure. 1 Saw tooth lifecycle combined with FDD.
Source: Adapted from [5,6]

3. Methodology

The system was developed by applying the saw tooth lifecycle methodology [9], created by Robert Owen and Peter Coad as part of their Feature Driven Development (FDD) [10]. The methodological phases that are part of the development of this project are presented in Fig. 1.

The previously mentioned methodological phases are divided into the design and development of the system according to the combination of features defined in the system requirements analysis. From these features, prototypes are made to validate the conceptualization of the system that is designed from the previously described characteristics and the system is built in a recurring and incremental way [9].

The main reasons that have triggered the use of the Feature Driven Development [10] are the focus on the developer to produce results every two weeks and the ease of looking over the results in a simple way for each one of the functionalities developed without the need to perform big tests.

During the system's design, the web app's established requirements show the user all the detections that were made in a determined geographic area using a map supported by a Location Based System (LBS) [6, 11]. In terms of the information display about the potholes, the system draws influence areas defined by the mass transit system stations and locates them in critical zones according to the recurring detections that are performed via the mobile device. The user can modify the ratio of these areas to observe the affected areas.

In order to limit the responsibilities of the system's functionalities, the following actors were defined:

- Maintenance Manager: this is responsible for displaying the information processed by the system in order to make decisions. This actor can review the information through maps and standard reports to determine the critical roadways. Maps are provided with cluster and buffer analysis to give value information for decisions.
- Data collector: this is responsible for taking the data (location, direction and movement) along the roadway and making different rounds with the mobile device. The mobile device is configured with Android OS and

the mobile application is designed to collect data in the field.

- GIS analyst: is responsible for verifying the information processed by the system that is provided by the data collector. This actor can review the information through maps with buffer analysis, displayed on OpenStreetMaps [12] and Google Maps [13]. In the verification process, the ranking and location of potholes can be modified according to related data.
- Web client: this visualizes the relevant information obtained by the system. This actor can review the information through standard pothole location map and standard report.

The functionalities that are developed as part of the system are:

- Manual capture of data: the data is captured manually over a roadway pothole, and the type of pothole is specified using the mobile device. This functionality is performed by the data collector.
- Automatic data capture: this captures data while the mobile device is on a determined route. It does not specify the pothole type and captures multiple flaws. This functionality is performed by the data collector.
- Cataloguing the possible type of pothole: the flaw is catalogued according the scale given by the category B VIZIR methodology. This functionality is performed by the GIS analyst.
- Analyzing collected information: this is performed by the GIS analyst. The information must be analyzed following the Geographic Information System Methodology.
  - Analyzing information collected using the Buffer method: the information of the measurement is analyzed using Buffer's algorithm with GIS analyst supervision.
  - Analyzing information using clusters: a cluster's algorithm is used with supervision by the GIS analyst.
- Visualizing collected information: visualizing on a map all the information collected by the system during a date rank. This is performed by the web client and the maintenance manager.
- Showing the summary chart of roadway potholes: this is performed by the maintenance manager. It shows a summary of the detected potholes.
- Visualizing possible flaws per station: this shows a map with what has been detected in a determined station. It is performed by the maintenance manager.
- Visualization of flaws per date: this shows on a map the imperfections detected on a specific date. It is performed by the maintenance manager.

The information processing within the system's functionalities is defined in Fig. 2. This shows each one of the stages in which the data is processed; it begins with the capture of the information, analysis of the information, categorization of any type of flaw, storage of the categorization and display of the processed data through maps and reports.

To capture the information manually and automatically, the smartphone must be placed in the bus interior, as is shown in Fig. 3.

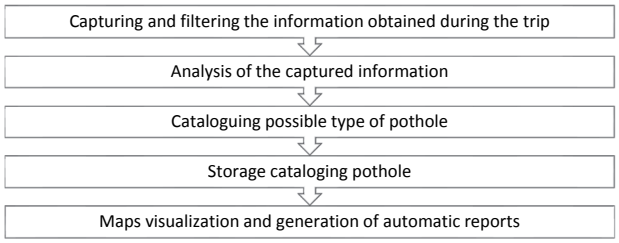


Figure 2. Processing information stages of the potholes' localization and detection system.
Source: The authors.

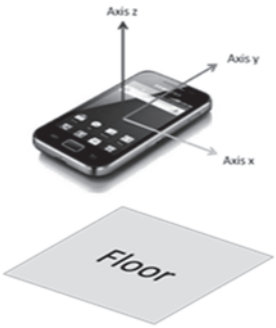


Figure 3. Spatial location of the smartphone to capture information on the roadway potholes.
Source: The authors.

The smartphone must be placed horizontal to the floor so that it is upright (positive axis z) and the phone's head (axis y) is in the direction in which the vehicle is moving.

The data stored in the system relates to the information on the detections of possible roadway potholes that are identified by mobile devices. The potholes are analyzed and classified through the system's functionalities. In terms of pothole detection, the system stores the latitude [degrees], longitude [degrees], GPS accuracy [meters], speed of the vehicle [km/h], accelerometer's acceleration on axis X, Y, Z [m/s2] and compass azimuth [degrees]. The pothole metadata stores the system's date and time, GIS analyst who classifies the pothole is identified, the rating according VIZIR scale (Chart 1), type of pothole (Chart 2) and detections analyzed.

The VIZIR ranking of the B category is used for the functional type potholes and is not related to the structural capacity of the road. [1] This classification is based on the Superficial Deterioration Index which ranks the pothole into three general cases with numerical values from 1 to 7, as is shown in Table 1.

Table 1.
Index of VIZIR Deterioration Surface Methodology [1]

Values	Ranking	Description
1 2	Good	Pothole in fine condition. Requires routine maintenance.
3 4	Fair	Pothole with some structural cracking and deformation. Requires maintenance and medium intensity treatment.
5 6 7	Deficient	Pothole with large amount of cracking and deformation. Requires major work.

Source: Adapted from [1]

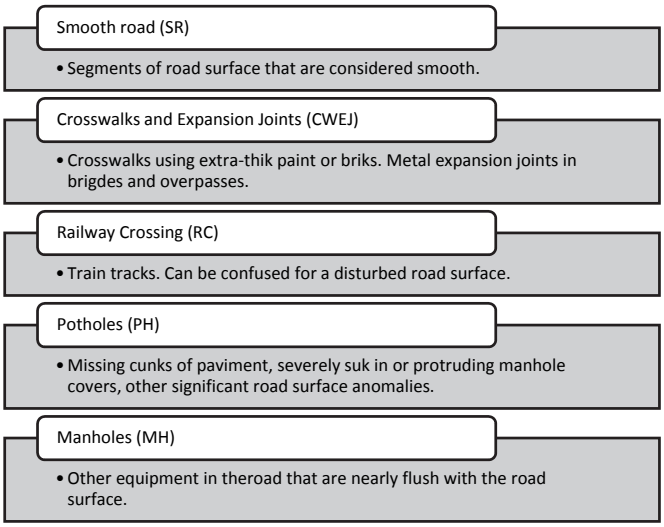


Figure 4. Roadway potholes rankings.
Source: Based on [14-16]

The superficial damage produced by roadway potholes may be caused by different structures located on the road network as well as by local transiting. According The Pothole Patrol [14] that was developed by the MIT, the main types of roadway potholes that should be taken into account during detection and location are ranked in Fig. 4.

From the data collected on pothole detections and ranking, the output displayed by the system to the final user is:

- Summary report of roadway potholes.
- Roadway potholes map (by date).
- Detections map (by date range).
- Buffer analysis map (by date range).
- Cluster analysis map (by date range).
- Roadway potholes map per station.

4. Proposed architecture

The architecture developed for the system is based on the service-oriented architecture for mobile applications [17] and on improvements made to the Information System of Flora and Fauna (ISFFA) [18]. The proposed architecture performs the initial data processing on the mobile device and the web server then performs the processing helped by the Google Location-Based Service using the principles of the Service Oriented Architecture (SOA) [19].

In order to improve the reliability of the data transmission through the mobile network the CarTel project [20] was taken into consideration; this was created by MIT and the Cafnet protocol (Carry and Forward Network) was designed. This protocol helps to maintain the data communication through a network that presenting intermittence and ranges of little coverage, as is shown in Fig. 5.

In order to reduce the impact of mobile device restrictions, there have been some algorithms implemented to the device's software that require low processing, low storage on the main memory and low usage of secondary storage in the disc on system's information. The dispatch of information

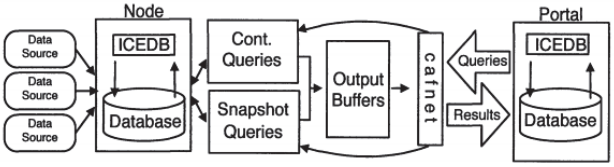


Figure 5. *CarTel* software architecture.
Source: [15]

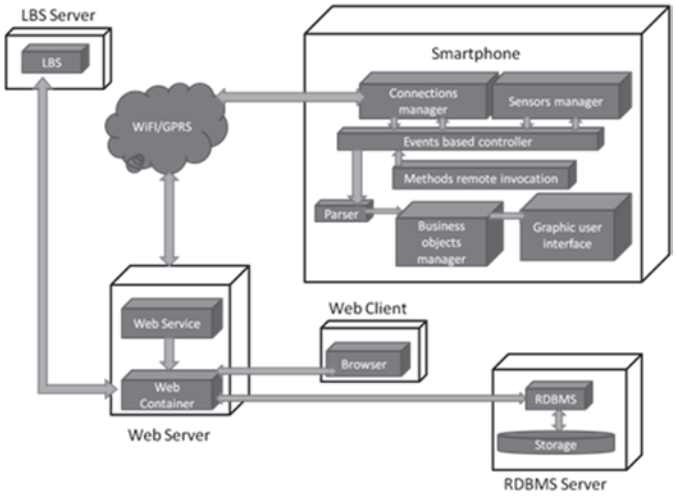


Figure 6. Mobile architecture proposed for the location and detection system of the roadway state.
Source: The authors.

in plain text through messages is created by using the SOAP protocol and XML [2].

The given architecture shown in Fig. 6 focuses mainly on an app for Geographic Information Systems and Location Based Systems. [2]:

In the proposed architecture great emphasis is made on the architectonical design of the mobile app due to the importance that the information capture has through the measuring and acceleration sensors (accelerometer), direction measuring (compass) and location (GPS).

The Sensors Administration component is responsible for collecting all data from the previously mentioned sensors simultaneously and with better accuracy taking into account external conditions. The High-Pass Filter (HPF) [21] has been implemented in this component in order to obtain more significant values and eliminate those that may cause noise and false detections.

5. Results

The system proposes a capture method that is simple and has great coverage compared to other traditional methods and is used in Civil Engineering in the Longitudinal Profile Analyzer (LPA) [22], for example. This requires the use of special equipment that has to be towed with a vehicle moving at a constant speed to make the captures and it also requires additional software and hardware to revise the information analysis. The RSMS (Roadway State Monitoring System)

Figure 7. Segment chosen for the study area and location of the imperfections in the section selected for testing.
Source: The authors.

Table 2.
Characterization of road imperfections located during testing.

#	Latitude	Longitude	VIZIR Score	Closest station	Direction	
1	4° 42' 33,329" N	74° 3' 12,272" W	6	Calle 127	South North	-
2	4° 41' 50,813" N	74° 3' 1,829" W	4	Pepe Sierra	South North	-
3	4° 41' 24,135" N	74° 3' 23,856" W	5	Calle 100	North South	-
4	4° 43' 35,867" N	74° 3' 1,829" W	3	Calle 142	South north	-

Source: The authors.

carries out the captures only with the sensors located within the smartphone and sends the captured information via mobile network to be processed on the GIS (Geographic Information System).

The chosen study zone was a 10 kilometer (6miles) Phase II roadway corridor belonging to the TransMilenio mass transit transportation system going North-South and South-North, as shown in Fig. 7.

This section was selected due to it having a significant number of potholes throughout the city [23] and also for the type of asphalt used in the construction of phases 1 and 2 (Caracas Avenue and North Motorway.) The asphalt used to construct the roadway chosen is rigid and segmented in tiles divided by transversal joints.

Four roadway potholes which are shown in Table 2 and visualized in Fig. 7 were located and typified in the test.

When the test is performed with data capture on automatic mode the purpose is to examine if the system is capable of detecting the pothole with the data capture on auto mode. Give the tests conditions, only the false negatives are taken into consideration to evaluate the system´s capacity for detection in the following circumstances: environmental conditions (cloud cover, rain, electromagnetic waves) and the phone location inside the bus. These variables affect the results obtained because the sensors are sensitive to the previously mentioned environmental conditions. [8].

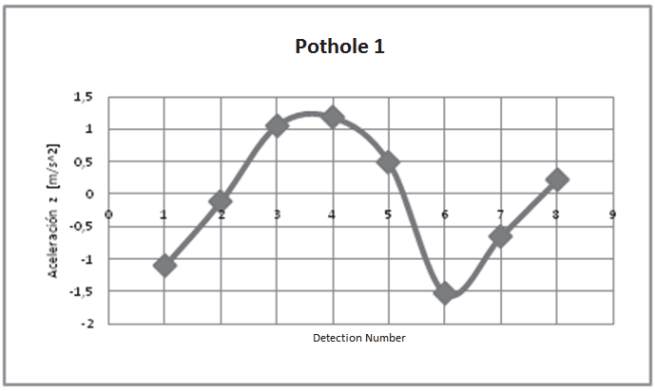


Figure 8. Data capture of pothole 2 during trip 3. Pothole located near to Calle 100 Station, direction North - South.
Source: The authors.

The tests were performed in the best environmental conditions possible: data intake during times of low passenger input, few clouds, sunny days, and the phone located on the window frame on the front part of the bus. Some variables such as additional weight and number of passengers were not taken into account since both of them are hard to control during normal operation of the system.

In order to analyze the detected data, we give an example from the pothole 1 in the round 3. As observed in Fig. 8 there are positive and negative accelerations because of the High-Pass filter effect used to improve the data quality captured by the accelerometer. Pothole 1 has maximum recorded values of 1,24 m/s^2 and minimum values of -1,55 m/s^2. It is possible to perform the detection because of the acceleration variation of 2,79 m/s^2 between the maximum and the minimum values.

The depth ranges of the recorded potholes in the study zones are between 1 and 6 centimeters. These potholes types are frequently found on the road network of TransMilenio phases 1 and 2 [23].

6. Architecture comparison

The most important CarTel contribution is that it uses elements such as the intermittent connectivity through Wi-Fi networks and its own communication protocol in order to obtain information in the environment. These elements allow CarTel to have a wide geographic range in order to obtain information and allow multiple sensors to connect to the system to obtain information from the environment.

In its architecture the system uses a connection to the mobile GPRS network to link the smartphones that are collecting data on roadway conditions from a wide geographic area, as was previously undertaken with the Flora and Fauna (ISFFA) Information System [18]. The use of the

mobile network benefits the coverage area and allows connectivity in a wide area of the city through the Connection Manager.

The architecture of the proposed system puts emphasis on the use of protocols and technologies that are nowadays standard in the development of information systems. The SOAP protocol is used for the interoperability between smartphones and web server as it is supported by the great majority of Android devices. This is an advantage as a great variety of smartphones from different manufacturers and OS can be added, providing they are compatible with the SOAP protocol.

Using the proposed architecture in RSMS with a larger number of smartphones and sensors available in the market, it is possible to monitor and obtain information beyond the roadway infrastructure that could lead to the new projects such as the model Internet of Things. This new model states: "Things having identities and virtual personalities using smart interfaces to connect and communicate with social, environmental and user contexts" [24]. Based on this concept it is possible to use an extended variety of sensors to monitoring roadways, vehicular traffic and Intelligent Transport Systems (ITS) that are focused on transportation management and supply chain execution. [25]

The following software and hardware tools were used:

Software: IDE NetBeans 6.9.1, Eclipse Indigo JEE, Android SDK for Windows, MySQL Server for Windows, Google Maps API.

Hardware: Smartphone: Samsung Galaxy Ace

Server: Intel Xeon 2,4 GHz QuadCore, Memory: 4 Gb. Operative System: Ubuntu 10.2, Network interface: 100/1000 Mbps.

7. Conclusions

It has been proved that our design is viable and can help as a reference model in future situations that require interoperability with acceleration, location and direction sensors that are available in mobile devices using a cell phone network.

The use of different devices all from same brand may affect in the accuracy of the sensor's data capture due to the heterogeneity of the embedded circuits that are located in each one of the models.

The developed system contributed to the test system's architecture and the data capture through the sensors. Architectural patterns Sensor-Computer-Control and the Android's operative system architecture for the mobile device, Model-View-Controller for the web client and the Java Enterprise Edition Framework architecture for the app server were chosen for the system design.

The system allows measurements to be made of the roadway's physical condition and infrastructure using standard commercial hardware, in this case smartphones available on the market. Mobile networks are then used to achieve a wide area range to gather low-cost information using smartphones in comparison to the traditional methods and devices used. The added value is that the information is displayed through the Location Based System [11] and the accelerometer allows access to the information gathered.

References

[1] Cerón-Bermúdez, V.G., Evaluación y comparación de metodologías VIZIR y PCI sobre el tramo de vía en pavimento flexible y rígido de la vía: Museo Quimbaya-CRQ Armenia Quindío (PR 00+000-PR 02+600), 2006.

[2] Smiatek, G., SOAP-based web services in GIS/RDBMS environment, Environmental Modelling & Software, 20(6), pp. 775-782, 2005. DOI: 10.1016/j.envsoft.2004.04.008

[3] Dong, P., Yang, C., Rui, X., Zhang, L. and Cheng, Q., An effective buffer generation method in GIS, in Geoscience and Remote Sensing Symposium, 2003. IGARSS'03. Proceedings. 2003 IEEE International, 2003, pp. 3706-3708. DOI: 10.1109/IGARSS.2003.1295244

[4] Grubesic, T.H. and Murray, A.T., Detecting hot spots using cluster analysis and GIS, in Proceedings from the Fifth Annual International Crime Mapping Research Conference, 2001.

[5] Confederación de Empresarios de Andalucía. ¿Qué es un SIG?. [Online] 2010. [date of reference August 9th 2015]. Available at: http://sig.cea.es/SIG

[6] Wang, S., Min, J. and Yi, B.K., Location based services for mobiles, Technology and Standards, [Online]. 1, 2008. [date of reference: August 10th 2015] Avaliable at: http://to.swang.googlepages.com/ICC2008LBSforMobilessimplifiedR2.pdf

[7] Khan, A.M., Lee, Y., Lee, S.Y. and Kim, T., Human activity recognition via an accelerometer-enabled-smartphone using kernel discriminant analysis, in Future Information Technology (FutureTech), 2010 5th International Conference on, 2010, pp. 1-6. DOI: 10.1109/FUTURETECH.2010.5482729

[8] Chun, S., Lee, S., Nah, J., Choi, J. and Park, J., Localization of wi-fi access point using smartphone's GPS information, in Mobile and Wireless Networking (iCOST), 2011 International Conference on Selected Topics in, 2011, 121 P. DOI: 10.1109/iCOST.2011.6085822

[9] Rowen, R., Software project management under incomplete and ambiguous specifications, IEEE Transactions on Engineering Management, 37, pp. 10-21, 1990. DOI: 10.1109/17.45260

[10] Coad, P., Luca, J. and Lefebvre, E., Java modeling color with UML: Enterprise components and process with Cdrom. Prentice Hall PTR, 1999.

[11] Kim, H., Song, J., Kamilov, M. and Han, S., Real-time u-LBS system for mobile vehicle in 3G networks, in Proceedings of the International Conference on Mobile Technology, Applications, and Systems, Yilan, Taiwan, 2008, pp. 1-9. DOI: 10.1145/1506270.1506282

[12] Coast, S., The Open Street Map Initiative, [Online]. 2008. [date of reference August 9th 2015] Available at: http://stevecoast.com/category/openstreetmap/

[13] Miller, C.C., A beast in the field: The Google Maps mashup as GIS/2, Cartographica: The International Journal for Geographic Information and Geovisualization, 41, pp. 187-199, 2006. DOI: 10.3138/j0l0-5301-2262-n779

[14] Eriksson, J., Girod, L., Hull, B., Newton, R., Madden, S. and Balakrishnan, H., The Pothole Patrol: Using a Mobile Sensor Network for Road Surface Monitoring. MobiSys '08 Proceedings of the 6th International Conference on Mobile Systems, Applications, and Services, 2008. DOI: 10.1145/1378600.1378605

[15] Monleón-Cremades, S., Ingeniería de puentes: Análisis estructural. Universidad Politécnica de Valencia, 1997.

[16] Hibbeler, R.C., Mecánica de materiales. Pearson Publications Company, 2006.

[17] Natchetoi, Y., Kaufman, V. and Shapiro, A., Service-oriented architecture for mobile applications, in Proceedings of the 1st International Workshop on Software Architectures and Mobility, Leipzig, Germany, 2008, pp. 27-32. DOI: 10.1145/1370888.1370896

[18] Casas-Avellaneda D.A. and López-Parra J.F., A mobile architecture for integration of smartphones with LBS for flora and fauna inventories, DYNA, 80(180), pp. 9-15, 2013.

[19] Erl, T., Soa: Principles of service design. Prentice Hall, 2008.

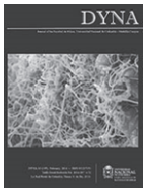
[20] Hull, B., Bychkovsky, V., Zhang, Y., Chen, K., Goraczko, M., Miu, A., Shih, E., Balakrishnan, H. and S Madden., CarTel: A distributed mobile sensor computing system, in Proceedings of the 4th

International Conference on Embedded Networked Sensor Systems, 2006, pp. 125-138. DOI: 10.1145/1182807.1182821
[21] Krause, G., High-pass filter, 1967.
[22] Caicedo, B., Murillo, C.A. and Tristancho, A., Medida perfil longitudinal de un pavimento mediante Navegación inercial, DOI: 10.16924%2Friua.v0i17.496
[23] Cámara de Comercio de Bogotá. Comportamiento de los indicadores de movilidad de la ciudad a Diciembre de 2012. [Online] 2012. [date of reference August 10th 2015] Available at: https://www.ccb.org.co/content/download/4646/57496/file/Observatorioo%20%202013.pdf
[24] Atzori, L., Iera, A. and Morabito, G., The internet of things: A survey, Computer Networks, 54, pp. 2787-2805, 2010. DOI: 10.1007/s10796-014-9489-2
[25] Arango, M., Gómez, R. and Zapata, J., Information systems applied to transport improvement, DYNA, 80(180), pp. 78-86, 2013.

D.A. Casas-Avellaneda, received his BSc. Eng in Systems Engineering in 2013 from the Pontificia Universidad Javeriana, Bogotá, Colombia. From 2012 to 2013 he worked for Oracle Colombia as an intern, and since 2013 he has worked as a Middleware Solution Specialist for Oracle Direct Latin America. His research interests include: mobile computing, integration systems and geographic information systems.
ORCID: 0000-0002-6567-9275

J.F. López-Parra, received his BSc. Eng in Systems Engineering in 1988 from the Universidad Nacional de Colombia, Bogotá, Colombia. He received his Specialization degree in Geoinformatics in 1992 and MSc in Environment and Development in 2000. He has worked in programs and on projects to do with geographic information systems, remote sensing and spatial data infrastructure in Colombia since 1995. He is currently part of the Pontificia Universidad Javeriana's GIS Research Team.
ORCID: 0000-0003-1298-1910

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS



DYNA
http://dyna.medellin.unal.edu.co/

Forecasting of short-term flow freight congestion: A study case of Algeciras Bay Port (Spain)

Juan Jesús Ruiz-Aguilar [a], Ignacio Turias [b], José Antonio Moscoso-López [c], María Jesús Jiménez-Come [d] & Mar Cerbán [e]

[a] *Intelligent Modelling of Systems Research Group, University of Cádiz, Algeciras, Spain. juanjesus.ruiz@uca.es*
[b] *Intelligent Modelling of Systems Research Group, University of Cádiz, Algeciras, Spain. ignacio.turias@uca.es*
[c] *Intelligent Modelling of Systems Research Group, University of Cádiz, Algeciras, Spain. joseantonio.moscoso@uca.es*
[d] *Intelligent Modelling of Systems Research Group, University of Cádiz, Algeciras, Spain. maríajesus.come@uca.es*
[e] *Research Group Transport and Innovation Economic, University of Cádiz, Algeciras, Spain. mariadelmar.cerban@uca.es*

Received: November 04rd, 2014. Received in revised form: June 12th, 2015. Accepted: December 10th, 2015.

Abstract
The prediction of freight congestion (cargo peaks) is an important tool for decision making and it is this paper's main object of study. Forecasting freight flows can be a useful tool for the whole logistics chain. In this work, a complete methodology is presented in order to obtain the best model to predict freight congestion situations at ports. The prediction is modeled as a classification problem and different approaches are tested (k-Nearest Neighbors, Bayes classifier and Artificial Neural Networks). A panel of different experts (post–hoc methods of Friedman test) has been developed in order to select the best model. The proposed methodology is applied in the Strait of Gibraltar's logistics hub with a study case being undertaken in Port of Algeciras Bay. The results obtained reveal the efficiency of the presented models that can be applied to improve daily operations planning.

Keywords: freight forecasting; classification; congestion; artificial neural networks; multiple comparison tests.

Predicción a corto plazo de la congestión del flujo de mercancías: El caso de estudio del Puerto Bahía de Algeciras (España)

Resumen
La predicción de la congestión en el tráfico de mercancías (picos de carga) es una importante herramienta para la toma de decisiones y es el principal objetivo de este trabajo. Predecir los flujos futuros de mercancías proporciona una potente herramienta en la cadena de suministro. En este trabajo, se presenta una metodología para conseguir el mejor modelo para predecir situaciones de congestión en flujos de mercancías. La predicción es modelada como un problema de clasificación, evaluando diferentes métodos (*K*–vecinos, clasificador Bayesiano y Redes Neuronales Artificiales). Para seleccionar el mejor modelo se desarrolla un panel de expertos (mediante métodos post–hoc del test de Friedman). La metodología propuesta se aplica a la cadena logística del Puerto Bahía de Algeciras. Los resultados obtenidos revelan la eficiencia de los modelos presentados, que pueden ser aplicados para mejorar la planificación diaria de operaciones.

Palabras clave: predicción de mercancías; clasificación; congestión; redes neuronales artificiales; test de comparación múltiple.

1. Introduction

Freight volume forecasting is a crucial component of the transportation system that can be used to improve the service quality and the operation planning in order to make correct decisions and support the system management. The flow of goods is one of the most important aspects in transport systems, and understanding the transport chain is central to its management, planning, improving facilities and operations [1]. Currently, companies that have a relationship with transportation require updates in terms of time to maximize their operating profit and also to improve the level of service [2].

The organization of multimodal transportation is complex because there are many actors involved in the supply chain. In this type of organization, in which there are many elements, the principal problem is information [3]. The availability of information about transport systems is vital to analyze trends [4]. Improving the quality of information using freight traffic predictions could be crucial in the short-term decision-making process.

Many studies have analyzed short-term forecasting problems pointing out the importance of prediction in different fields of research [5-9]. Short-term forecasting models can be categorized into two groups, depending on the kind of problem to be considered: regression or classification models [10]. Many of the studies relating to forecasting traffic flows can be found in maritime transport literature. The main focus of these papers was the avoidance of congestion problems. The large increase in trade has caused significant impacts in the supply chain operations, increasing traffic congestions and time delays in freight transport [11]. Many ports are congested and consequently costs are raised, which affects competitiveness. These effects can divert traffic to other more competitive ports [12]. Short-term predictions have been used to facilitate the implementation of daily port operation activities, such as the allocation and provision of personnel and necessary equipment for the proper planning of daily operations [13]. In this context, a prediction of certain traffic flows that lead to congestion could be a powerful tool to avoid those situations. Owing to the lack of literature relating to predicting freight congestion, a brief literature review on the proposed methods and other techniques for transportation problems was included in this paper.

There are, however, a wide range of works related to predicting traffic flows in transport using regression models [14-17]. Numerous types of methodologies have been used in the literature to predict traffic flows in shipping or maritime transport. Classic techniques such as regression analysis, classic decomposition or univariate forecast models (ARIMA) have been applied for years by several authors who have obtained satisfactory results. Seabrooke et al. [18] employed regression analysis to predict load growth in Hong Kong Port in Southern China. Similarly, Chou et al. [19] proposed a modified regression model to predict the volume of import containers in Taiwan. Fung [20] developed a forecasting model to predict the movement of containers in Hong Kong port and to provide a more accurate prediction than the one offered by the authorities. Klein [21] tried to find a model to predict the flow of maritime traffic in total tonnage in the Antwerp Port. Peng and Chu [13] performed a comparative analysis between different prediction models in order to find the most accurate model to predict container volumes. Ruiz-Aguilar et al. [22] developed a multi-step procedure to predict the number of freight inspections at ports. Finally, Yang et al. [23] modeled the road traffic in a Chinese container port environment, and the results indicated that the model had a high estimation quality at peak traffic times.

Many transport systems problems are represented as non-linear phenomena, which come from many sources and contain complex data. These nonlinear phenomena are difficult to plot in a dynamic context and are subject to constant changes. In this context, artificial neural networks (ANNs) have been used to model traffic flows due to their adaptability, nonlinearity and arbitrary function mapping capability. Therefore, several authors pointed out that the utilization of ANNs can provide a competitive advantage for planning and forecasting in transportation research [8,24-26]. Recent studies have introduced the use of intelligent and nonlinear techniques (specifically ANNs) for predicting and estimating maritime traffic flows, showing better results in comparison with classical statistical models. Lam et al. [27] applied neural network models to several types of goods movements in Hong Kong port, comparing the prediction performance with the one obtained with the classical regression methods. The results show that neural network prediction was more reliable. Recently, Gosasang et al. [28] suggested the use of an MLP model as a technique for predicting the volume of containers at Bangkok port. Furthermore, this author made a comparison with classic techniques that had, up to that point, been used to make calculations for the port.

Moreover, there are not many studies in the literature related to transport using classification methods to forecast congestion episodes of traffic flows, especially for maritime transport. Examples of these models are the well-known *k* nearest neighbor (*k*-NN) and the Bayesian Classifier (BC) models, which are widely, used classification techniques. As such, Robinson and Polak [29] used the *k* nearest neighbor method in order to model the urban link travel time in the city of London. The results provide a more accurate estimation than other techniques. More recently, *k*-NN was compared to other techniques in Bhave and Rao's work [30]. The authors studied the acoustic signatures of vehicle detection, providing traffic congestion estimations. Furthermore, [31] used a *k*-NN classification in a hybrid procedure, which involves other techniques such as multiple regression and principal component analysis in order to determine urban road categories using real traffic flow data.

Moreover, BC are proposed in several areas of transportation. Travel time forecasting has been emphasized by some authors. Thus, [32] a simple Bayesian estimator to forecast arterial link travel speed in Jeonju (China) was developed. Furthermore, a Naïve Bayesian Classification model was also used by Chowdhury et al. [33] for the same purpose. The Bayesian classifier has been applied with successful results in the area of transport security. Thus, a model based on BC was developed by Boyles et al. [34] in order to anticipate incident duration, whereas Oh et al. [35] proposed a nonparametric Bayesian model to estimate the likelihood of an accident in real-time from empirical data in California (US). More recently, other studies relating to transportation have identified the occupant postures in vehicles in order to reduce the risk of injuries [36], or vehicle classification using Bayesian networks [37].

Intelligence methods have also been successfully applied in classification problems related to transport. Particularly, ANNs have demonstrated good performances due to their inherent capabilities in classification tasks. In this sense, neural classifiers based on ANNs were employed in order to detect operational problems on intersections and signalized urban arterials [38] and to detect freeway incidents using probabilistic neural networks [39]. This kind of neural

network classifier was also proposed by Abdel-Aty and Pande [40] to improve the identification of patterns in a freeway that could forecast potential crashes. Moreover Wu et al. [41] adopted a multi-layer perceptron network, which was used to recognize and classify vehicles.

The main objective of this work is the prediction of peak loads in order to avoid congestion in port nodes. The authors have made a comparison between the above mentioned three different peak prediction techniques applied to Roll-on/Roll-off (Ro-Ro) traffic: *k*-NN, BC and ANNs. Multiple comparison tests have been used as an expert panel in order to find the best model to predict the peak volume of short-term (one-day ahead) traffic in the Strait of Gibraltar, which is the study case. To the best of our knowledge, classification techniques to predict workload peaks have not yet been proposed in the research literature.

The remainder of this paper is organized as follows: Section 2 presents the study area and the database; Section 3 describes the forecasting models used in this work and the methodology to investigate the performance of the different models; Section 4 analyzes the results obtained by the proposed methodology; and Section 5 presents the conclusions.

2. Materials and methods

The different proposed models were tested in the Strait of Gibraltar's logistics hub. The Strait of Gibraltar's geostrategic position means it is one of the main maritime routes for carriers, and consequently one of the most intense maritime traffic points.

Furthermore, the Strait of Gibraltar is the bridge between Africa and Europe. In 2012, 116,690 vessels crossed the Strait of Gibraltar, of which 40,214 ships connected African and European ports. Traffic on both sides consists essentially of passenger and Ro-Ro traffic. The distance between the ports on the Strait is only 12 miles and this makes transport from northern Africa to Europe more effective because of the close location. In 2012 there were 4,692,889 tons of good moved on this route. The Algeciras bay Port is in the top 10 Europe ports in terms of high Ro-Ro volume. As the main entry point of goods from Africa, the Algeciras Bay Port is in charge of managing their entrance to the EU, and also exercising several controls (health, industrial or tax) that are performed in the Border Inspection Posts (BIPs). Perishable goods are the most important traffic in this port. Due to this, freight transit must be performed in the most effective way possible. Hence, the proposed prediction model can add value to the supply chain management. It is important to adopt models to ease the process and to facilitate transshipment procedures.

The information used to develop this work is taken from a database provided by the Algeciras Bay Port Authority. The database contains all imports in Algeciras during the study period: between 2000 and 2007, and it was initially composed of more than three million records. After an extensive preprocessing stage was performed, the final database contained 2,970 daily records from January 2000 to December 2007. This work focuses on the first chapter (vegetables), which are the freight that is most frequently inspected in the BIP. The Border Inspection Post has a mean workload of 7000 ton/day. The main focus is to determine when the daily freight volume of vegetables increases over certain level (freight peak). Therefore, the forecasting tool can be used to avoid congestions, delays and cost increments. Stakeholders can use forecasts to help with resource planning.

After the preprocessing and review of the database, a daily increment of 400 tons was selected as a peak level, since this value diminished the performance of the system (the total time at BIP). More staff would be assigned during these peak levels of service in order to avoid congestion.

An autocorrelation analysis has been undertaken in order to determine what the most important lags (time delays) have been in the past (more informative). As time progresses the autocorrelation coefficient decreases. There is only an increment detected when the lag coincides with the day of the week. Therefore, lags = 1, 2, 7, 14, 21 and 28 have been chosen as experimental autoregressive window size (n) to be tested.

The database of samples used to teach the forecasting models was arranged in the form of autoregressive data (see Eq. (1), where n was the width of the observation window in the past (lags).

$$\hat{y}(t+nh) = f\big(y(t), y(t-1), \ldots, y(t-n)\big) \rightarrow \\ \rightarrow \hat{y}(t+nh) = 1(congestion)\ 0(normal\ situation) \quad (1)$$

This information was used to make the prediction, as it uses autoregressive models. In this case, nh was the time horizon when freight peak is predicted. That is, nh=1 means a one day-ahead prediction. The prediction for this work was 1-ahead (short-term), for which different prediction models and autoregressive inputs have been combined (see Table 1)

3. Methodology

The forecasting of freight peaks can be developed using a classification scheme between two categories (freight peaks and normal situations). Classification is the inferring of meaning (category, class) from observations. There are two basic stages when designing a classifier: the training (or design) phase and the test (generalization or application) phase. The result of the training phase is the set of classifier parameters that define the discriminant functions that form the class boundaries between disjoint class or category regions.

3.1. Bayes classifiers

Bayesian classifiers are based on probability theory and they give the theoretical basis for pattern classification [42]. The decision rules based on the Bayes theorem are optimal [43], but they are unattainable in practice because the complete information about the statistical distributions of each class is unknown. However, they do provide the logical basis for all statistical algorithms.

The statistical procedures attempt to provide the information about the distribution of two ways: parametric and non-parametric. The first way makes some assumptions about the nature of the distributions, and their parameters must be estimated. The latter is distribution-free.

We assume that the prior probabilities and the probabilities associated with each class are known.

Thereafter, in order to calculate a posteriori probabilities $P(A|X)$ and $P(B|X)$ the Bayes' theorem is used. Pattern is classified as A if $P(A|X) > P(B|X)$, otherwise the pattern will be classified as B. This is equivalent to classify a pattern within class A if $P(X|A)\cdot P(A) > P(X|B)\cdot P(B)$. If it is further assumed that $P(A) = P(B)$, then $P(X|A) > P(X|B)$. Assuming that the conditional probabilities are statistically independent and follow normal distributions with identical variances, $||X-\mu_A|| < ||X-\mu_B||$ can be obtained (μ_A and μ_B are the class means), which is a minimum distance-based classifier. The equation of the decision function is $||X-\mu A|| = ||X-\mu B||$. This method is known as naive Bayes, if variable independence is assumed [44].

More generally, it is possible to assume variable dependence and therefore there is a covariance matrix of the variables. In this case, two hypotheses can be assumed. One is that the model has the same covariance matrix for each class (homoscedasticity assumption) and the mean of each class vary (LDA: Linear Discriminant Analysis). The other is that the covariance and the mean of each class vary (QDA: Quadratic Discriminant Analysis) [45]. Both hypotheses have been used in this work.

3.2. k – Nearest neighbors

Nearest neighbor classifier is a well-known non-parametric method in pattern recognition [46]. These classifiers use all the available data as templates for classification. In the simplest form, for a given input vector, a nearest neighbor classifier searches the nearest template and classifies the input vector into the class to which the template belongs. In a more complex form, the classifier uses k nearest neighbors.

This algorithm does not have a training phase off line. The idea is to store the available data set, so that when a new instance is classified the algorithm searches for similar cases in the stored examples and assigns the most likely class. A common way to find the closest examples is through the Euclidean distance. In order to prevent ties, an odd number of neighboring observations are used in this work (k=1,3 or 5).

3.3. Neural networks for classification

A multilayer perceptron (MLP) feedforward network trained with backpropagation algorithm [47] are capable of approximating any non-linear mapping with arbitrary accuracy. In this sense, multilayer feedforward networks are universal approximators [48]. Multi-Layer Perceptron (MLP) based on backpropagation learning procedure is the most commonly used artificial neural network in numerous applications, including classification [49,50]. Gradient descent algorithms have some drawbacks that can be avoided with second-order Newton based methods [51]. The Levenberg-Marquardt algorithm was designed to approach second-order training speed without the need to compute the Hessian matrix, and it uses an approximation applying the Jacobian matrix of first derivatives. The application of Levenberg-Marquardt to ANN training is described in the work of Hagan and Menhaj [52]. This algorithm appears to be the fastest method for training moderate-sized feedforward neural networks, and it is used in this work.

In such a feedforward network, the output vector y is a function (non-linear) of the input vector X, and some weights w Training (or designing) the network involve searching in the weight space of the network for a value of w that produces a function which fits the provided training data.

The training process is a function minimization (i.e., adjusting w in such a way that the objective function E is minimized). For general feedforward neural networks, the backpropagation algorithm evaluates the gradient of the output y to update w (weights) by eq. (2) (where α is the learning rate).

$$w(t+1) = w(t) - \alpha \nabla E(w) \quad (2)$$

3.4. Evaluation

Generalization error is the estimation of the accuracy of learning algorithms that explains how well a learning machine generalizes with different data from the training process. The experimental procedure for each proposed model was repeated 20 times in order to compare the different models and to determine (by analyzing the mean and the variance of the indexes) the best one.

In order to compare the generalization error of learning algorithms, statistical tests have been proposed. To compare the different classification models in the experiments, different evaluation parameters can be used:

- *Accuracy* (ACC) is defined as the ratio between the sum of true results (number of peaks and non-peaks correctly predicted) against the total number of samples in the data series.
- *Precision* (PR) is defined as the proportion of the true positive against all the positive results.

The values of these indexes are in the interval [0-1]. An accuracy of 100% means that the measured values are exactly the same as the given values. Accuracy is how close to "true" measurements these values are. Precision is how consistent the results are over several measurements, or how repeatable the model is.

3.5. Multiple comparison of models

ANOVA test is a well-known statistical method for testing the differences between more than two related sample means. ANOVA divides the total variability into the variability between the models, variability between the data sets and the residual (error) variability. The null-hypothesis is that there are no differences between the models. If the between-models variability is significantly larger than the residual variability, the null-hypothesis can be rejected and it can be concluded that there are some differences between the models. ANOVA is based on assumptions that are most probably difficult to assure. First, ANOVA assumes that the samples are drawn from normal distributions. In general, there is no guarantee for normality in the results of the classifiers. The second and more important assumption of ANOVA is homoscedasticity, which requires the random variables to have equal variance. The Friedman test [53,54] is the nonparametric equivalent of ANOVA test, without the need to ensure assumptions of normality and homoscedasticity.

The issue of multiple hypothesis testing is a well-known statistical problem. The usual goal is to decrease the probability of making at least one Type 1 error in any of the

comparisons. If the null-hypothesis is rejected, a post-hoc test can be used. Salzberg [55] states a general solution for the problem of multiple testing, the Bonferroni method, and notes that it is usually very conservative. Pizarro et al. [56] use ANOVA and Friedman's test to compare multiple models (in particular, neural networks) on a single data set. The authors of this paper have also used the ANOVA and Bonferroni methods successfully in previous works [57,58].

The oldest and most popular technique to undertake these multiple comparisons procedure is the LSD (Least Significant Difference) detailed in Fisher [59]. This method determines the difference LSD (See eq. 3).

$$LSD = t_{\alpha/2;N-I}\sqrt{S_R^2\left(\frac{1}{n_i}+\frac{1}{n_j}\right)} \rightarrow \|m_i - m_j\| > LSD \qquad (3)$$

- Which follows a t-distribution with degrees of freedom N-I. I is the number of models, ni and nj are the number of observations for each mean mi and mj. The α parameter is the statistical significance or the probability of making a Type I error and SR is the estimation (mean square) of error variability.

The drawback of LSD is that it has the highest probability of making rejections, which increases with the number of comparisons performed. To avoid this problem, other procedures have been introduced to make multiple comparisons. These methods are based on the Bonferroni inequality. This procedure sets a significance level α that is shared between each of the comparisons taken into consideration. It would be better not to use the Bonferroni method when the number of pairwise comparisons is very large because the level of significance of each comparison may become too small to be considered useful. Another method is based on the Studentized range (q), which results in the significant difference method proposed by Tukey [60], also called HSD method (Tukey Honest Significant Difference), using the HSD difference shown in the following equation.

$$HSD = q_{\alpha;I;N-I}\sqrt{S_R^2/n} \qquad (4)$$

Fisher's LSD method is the one that provides the most significant difference, and is followed by Tukey's method. The method to be applied depends on the risk to be assumed; that is, accepting as significant differences those that are not (less conservative situation), or accepting less significant differences that do actually exist. Carmer and Swanson [61] conducted simulation studies by the Monte Carlo method, which concluded that the LSD procedure is a very efficient test to detect true differences. In this work, these methods have been used as an expert panel in order to consider their individual decisions and to ensemble a global rule.

In the case that there are no significant differences, Occam's razor's criterion should be used. Occam's razor is the principle that states a preference for simple theories: "Accept the simplest explanation that fits the data".

3.6. Validation and experimental design

A resampling strategy was developed in order to compare different models and to determine if difference among the models exists [56,62].

It is necessary to measure several test sets in which the examples have not been used in the training phase. There are a range of methods to achieve suitable validation of the results (cross-validation, bootstrap, etc.); specifically, two-fold cross-validation (2-CV), which divides the database into two disjointed sets (training and test). The two fold cross-validation procedure was applied to the database in this research.

The model parameters are determined (coefficients, weights, etc.) with the first set, and the error measures that have been established (precision and accuracy e.g.) are computed with the second one. Subsequently, the sets are inverted and the same operations are re-performed, and the average of the two experiments is obtained. This procedure was repeated 20 times in order to calculate the average quality measurements in this experiment. It is the most pessimistic validation technique as it leaves out the half the data to calculate the measures of generalization error. Therefore, any other validation method will provide better results. Thus, the results with 2-CV can be considered as a minimum value of the actual results. The authors have previously successfully applied this procedure in different applications [26,63].

In order to study the peak prediction of Ro-Ro traffic in Algeciras Bay Port of, three different classification methods have been tested: Bayesian classifiers (Linear and Quadratic Discriminant Analysis), *K*-Nearest Neighbors (*K =1,3,5*) and ANNs (with different numbers of hidden units and different epochs). The complete set of models can be shown in Table 1. In this work, 120 models have been tested in order to select the best forecasting performance. For each of the 20 models in Table 1, different inputs have been used. In each case, a different number (n) of lagged data has been used as inputs for the prediction model (different sizes of this autoregressive information window n=1, 2, 7, 14, 21 and 28). Actually, a greater number of inputs does not guarantee better results. This fact will be confirmed with the results that were obtained in the experimental design. The random resampling experiment explained above has been applied to calculate the average quality indexes in order to compare results.

4. Results and discussion

Comparing the different models tested, generalization error has been checked in order to select the model with the minimum error. This can be achieved by calculating the mean of the accuracy and precision values for test samples with different autoregressive window sizes.

The results are presented graphically in Fig. 1. It can be observed that the best network input configurations are obtained generally with *n=1* and *n=7* autoregressive window sizes, which exceed values of 0.75 for accuracy and 0.76 for precision. The worst results are obtained for *n=28* which indicates that introducing a higher size of autoregressive window does not improve the prediction.

Table 1.
Models Tested in the Resampling Procedure

Model number	Model	Parameters
1 - 3	K-NN	K = 1, 3, 5
4 - 5	Bayes	LDA, QDA
6 - 10	ANN-BP	Epochs = 100
		Hidden units = 1, 2, 5, 10, 20
11 - 15	ANN-BP	Epochs = 300
		Hidden units = 1, 2, 5, 10, 20
16 - 20	ANN-BP	Epochs = 500
		Hidden units = 1, 2, 5, 10, 20

Source: The authors

In those models based on *K*-NN, as parameter *K* increases, the results of the model improve. More stable results are obtained with the Bayesian models. In this case, the different functions studied (LDA and QDA) provide similar results. Moreover, for those models based on ANNs, it can be observed that the number of epochs used in the training stage is not very relevant. However, the selection of optimal number of neurons in hidden layer is crucial to find a balance between complexity and generalization error. Better results of precision and accuracy are obtained when the number of hidden neurons decrease, reducing the complexity of the classifier.

Models 6, 11 and 16 seem to be the best models after a preliminary visual analysis of Fig. 1. These models that are based on ANNs may be the optimal input configurations. Nevertheless, the best model must be chosen by using a statistical multiple comparison technique. Analysis of variance (ANOVA) techniques inform us about the existence of significant differences between the treatments. The Friedman Test is used for this goal instead of the ANOVA test due to its non-parametric properties. The Friedman test is an alternative to the ANOVA test, when the assumption of normality or equality of variance is not met. This, as is the case with many non-parametric tests, uses the ranks of the data rather than their raw values to calculate the statistic.

When the Friedman test reveals the existence of significant differences, it is necessary to analyze how the means differ. In order to achieve this, different techniques have been used (Fisher Least Significant Difference–LSD, Tukey's Honest Significant Difference–HSD and Bonferroni's method). The choice of the best model was made by contrasting the results obtained from each technique using a panel of experts.

The selection of the best global model was undertaken in two phases. This two-phase procedure is showed in Fig. 2. Phase I deals with the application of statistical techniques previously described (Friedman + post hoc test) in order to point out which autoregressive window size is the best one in each model. Phase II identifies the best global model (algorithm and parameters) using the same techniques. In Phase I, the quality indices (accuracy and precision) obtained for each model's different autoregressive windows have been compared by the Friedman test in order to verify the null hypothesis, with a significance level $\alpha=0.05$. The results of the Friedman test reveal the existence of significant differences between the average of all the samples, hence the null hypothesis has been rejected. In other words, there are certain autoregressive windows that are better than others.

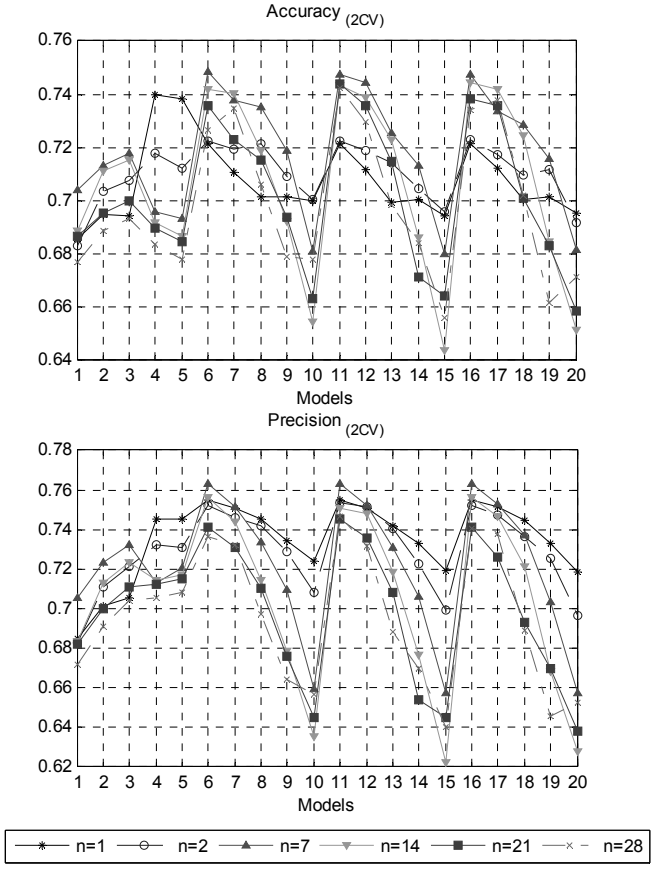


Figure 1. Accuracy and Precision results from the models collected in Table 1.
Source: The authors

Therefore, it is necessary to apply multiple comparison methods in order to find out which of the models or groups of models are significantly better in terms of accuracy and precision. Tukey's HSD, Bonferroni's and Fisher's LSD tests have been applied as a panel of experts to obtain the results presented in Table 2 (only Model 15 is presented as an example for simplicity).

The first step has been to point out those input configurations that are significantly different from the rest. The three experts agree that for the accuracy index, the input configurations 1 and 2 (*n=1* and *n=2*) are significantly different from the rest. In this case, the simplest computational model has been chosen (number 1). For precision values, the three experts agree that all the input configurations overlap one another. In this case, the input configuration with a better index should be chosen (*n=2*). Finally, due to the discrepancy between the selection of the better input configuration based on both quality indexes, an accuracy term was considered to be a critical criteria, which is due to the specific application of our research. In this work, the correct prediction of patterns of both classes it is the most relevant part (True Positives and True Negatives). Furthermore, in this study, false positive situations are preferable to false negatives because false negatives would produce overload and congestion situations decreasing the performance of the BIP system. Once the results obtained by Tukey's HSD, Fisher's LSD and Bonferroni's method for each of the 20 models have been analyzed, the best

input configurations have been chosen for each one.

In Phase II, in order to obtain the best global model, the methodology explained above has been applied again for the best input configurations. The results obtained in Phase II reveal the existence of significant differences between the input configurations. Therefore, the null hypothesis has been rejected with a level of significance $\alpha=0.05$. In order to find out which models are significantly different, the multiple comparison methods (Tukey's HSD, Fisher's LSD and Bonferroni's method) have once again been applied (as an expert panel). For Bonferroni's and Tukey's methods, all the models overlap; Fisher's LSD was the only one able to find differences between the groups. This is because it is a less conservative method than the others and offers more significant differences. The results obtained with Fisher's LSD are presented in Table 3. The HSD and Bonferroni methods have not been included due to the lack of space. The interpretation of the results collected in Table 3 shows that the best models are numbers 6, 11 and 16 (accuracy index). Therefore, the simplest model (number 6) can be selected as the best forecasting model based on Occam's razor rule. This model is a neural network with only one neuron in the hidden layer. In this sense, as the model is close to being linear and a more complex model (with a greater number of hidden neurons), it is not required. In Fig. 3, the results obtained for the best model (model number 6) can be observed. Results come from test data and one of the repetitions. The points on the graph correspond to the goods (measured in kg) that cross the Algeciras Bay Port on a daily basis. The circles correspond to right peaks predicted by the model and the triangles correspond to the misclassified ones.

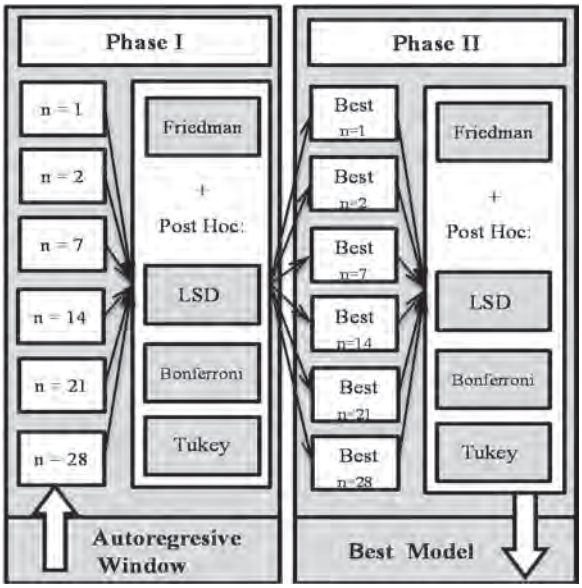


Figure 2. The two-phase statistical procedure for the selection of the best model. Source: The authors

5. Conclusions

A two-phase procedure has been carried out in order to find the best model to predict a significant increase in the amount of freight with the aid of a panel of statistical multiple comparison tests. In the first phase, the best autoregressive window can be selected, and in the second phase the best algorithm and its parameters are identified.

This panel of experts has been applied to a resampling procedure in order to measure the models' generalization error and then compare them using accuracy and precision indexes. In this way, the best model is an artificial neural network k with a three-layer structure with only one hidden unit using the Levenberg-Marquardt optimization method as a learning algorithm and an autoregressive window of $n=7$ as its input. Therefore, in this case, an ANN model achieves better performance than the Bayes classifier and KNN models. It is worth mentioning that the increase in complexity (adding hidden units) does not significantly improve the results. The Friedman non-parametric test rejected the null hypothesis and, therefore, it has been necessary to use post-hoc tests (Fisher's LSD, Tukey's HSD and Bonferroni's method). Fisher's LSD test has more power compared to other more conservative post-hoc comparison methods.

In this work, the risk consists of predicting a saturation situation when in reality does not happen. The only damage would be planning for a workload that was smaller than expected. The results confirm that it is possible to find a model that successfully predicts significant increases in the freight volume. This freight volume-forecasting tool (24 hours ahead) could be used by the different port organizations as a support decision aid for planning resources or facilities. Future work should be aimed at improving the performance of the models in order to provide more accurate predictions of the freight congestion at ports.

Acknowledgements

This work has been partially supported by a grant from the European project FEDER-FSE 2007-2013 and the Fundación Campus Tecnológico Bahía de Algeciras. In addition, the authors thank the Algeciras Bay Port Authority for kindly providing the Ro-Ro traffic database.

Table 2.
Window size comparison using Accuracy for Model number 15 (ANN-BP: 20 Hidden Units and 300 Epochs).

Model 15	Config.	Autoregressive window size (n)	No Significant				No Significant			
			ACCURACY	HSD	LSD	Bonferroni	PRECISION	HSD	LSD	Bonferroni
ANN-BP	1	n=1	0.718	1 2	1 2	1 2	0.694	1 2 3	1 2 3	1 2 3 5
	2	n=2	0.699	1 2	1 2	1 2	0.696	1 2 3	1 2 3	1 2 3
	3	n=7	0.658	3 4 5 6	3 5 6	3 4 5 6	0.681	1 2 3 5 6	1 2 3 5 6	1 2 3 5 6
	4	n=14	0.621	3 4 5 6	4 6	3 4 5 6	0.642	4 5 6	4 5 6	4 5 6
	5	n=21	0.647	3 4 5 6	3 5 6	3 4 5 6	0.666	3 4 5 6	3 4 5 6	1 3 4 5 6
	6	n=28	0.642	3 4 5 6	3 4 5 6	3 4 5 6	0.659	3 4 5 6	3 4 5 6	3 4 5 6

Source: The authors.

Table 3.
LSD Fisher multiple comparison of the best configurations.

LSD FISHER					
Model	**ACCURACY**	**No Significant**	**Model**	**PRECISION**	**No Significant**
1	0.706	1, 2, 10, 15, 20	1	0.704	1, 2, 3, 8, 9, 10, 13, 14, 15, 18, 19, 20
2	0.725	1, 2, 3, 9, 10, 14, 15, 19, 20	2	0.714	1, 2, 3, 8, 9, 10, 13, 14, 18, 19
3	0.734	2, 3, 4, 5, 9, 10, 13, 14, 18, 19	3	0.719	1, 2, 3, 8, 9, 13, 14, 18, 19
4	0.745	3, 4, 5, 7, 8, 9, 12, 13, 17, 18	4	0.739	**4, 5, 6, 7, 11, 12, 16, 17**
5	0.745	3, 4, 5, 7, 8, 9, 12, 13, 17, 18	5	0.738	**4, 5, 6, 7, 11, 12, 16, 17**
6	0.764	**6, 11, 16**	6	0.750	**4, 5, 6, 7, 11, 12, 16, 17**
7	0.753	4, 5, 7, 8, 12, 13, 17, 18	7	0.739	**4, 5, 6, 7, 11, 12, 16, 17**
8	0.745	4, 5, 7, 8, 12, 13, 17, 18	8	0.701	1, 2, 3, 8, 9, 10, 13, 14, 15, 18, 19, 20
9	0.733	2, 3, 4, 5, 9, 10, 13, 14, 18, 19	9	0.701	1, 2, 3, 8, 9, 10, 13, 14, 15, 18, 19, 20
10	0.723	1, 2, 3, 9, 10, 14, 15, 19, 20	10	0.698	1, 2, 8, 9, 10, 13, 14, 15, 18, 19, 20
11	0.764	**6, 11, 16**	11	0.749	**4, 5, 6, 7, 11, 12, 16, 17**
12	0.754	4, 5, 7, 8, 12, 13, 17, 18	12	0.747	**4, 5, 6, 7, 11, 12, 16, 17**
13	0.741	3, 4, 5, 7, 8, 9, 12, 13, 17, 18	13	0.698	1, 2, 3, 8, 9, 10, 13, 14, 15, 18, 19, 20
14	0.732	2, 3, 9, 10, 14, 15, 19, 20	14	0.699	1, 2, 3, 8, 9, 10, 13, 14, 15, 18, 19, 20
15	0.718	1, 2, 10, 14, 15, 19, 20	15	0.694	1, 8, 9, 10, 13, 14, 15, 18, 19, 20
16	0.764	**6, 11, 16**	16	0.749	**4, 5, 6, 7, 11, 12, 16, 17**
17	0.753	4, 5, 7, 8, 12, 13, 17, 18	17	0.734	**4, 5, 6, 7, 11, 12, 16, 17**
18	0.744	3, 4, 5, 7, 8, 9, 12, 13, 17, 18	18	0.699	1, 2, 3, 8, 9, 10, 13, 14, 15, 18, 19, 20
19	0.732	2, 3, 9, 10, 14, 15, 19, 20	19	0.701	1, 2, 3, 8, 9, 10, 13, 14, 15, 18, 19, 20
20	0.718	1, 2, 10, 14, 15, 19, 20	20	0.694	1, 8, 9, 10, 13, 14, 15, 18, 19, 20

Source: The authors

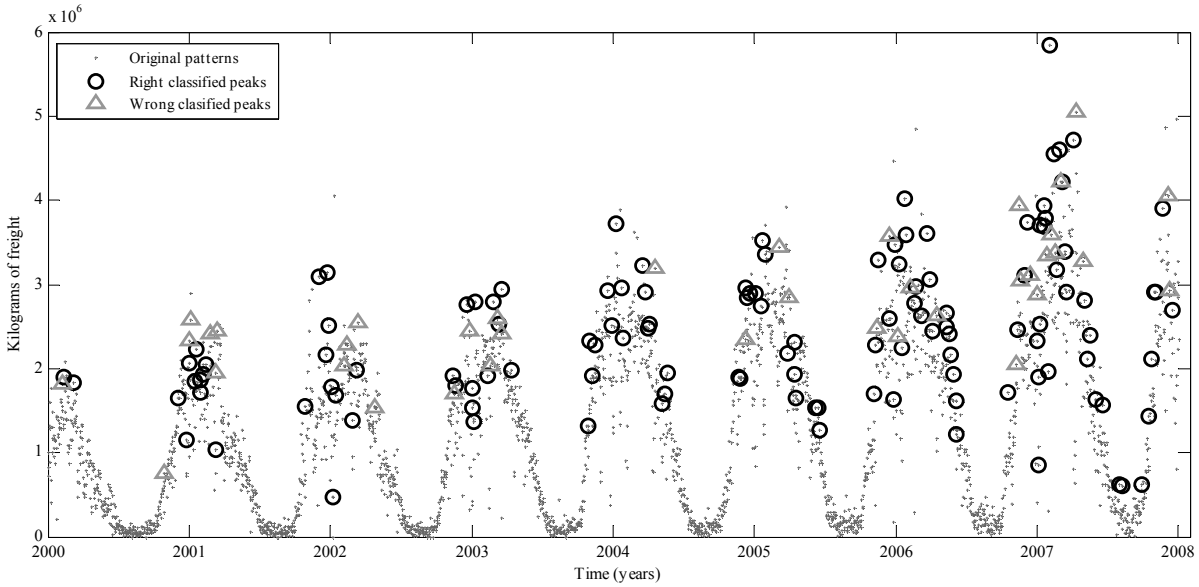


Figure 3. Freight peaks prediction results.
Source: The authors

References

[1] Chen, M.C. and Wei, Y., Exploring time variants for short-term passenger flow. Journal of Transport Geography, 19(4), pp. 488-498, 2011. DOI: 10.1016/j.jtrangeo.2010.04.003

[2] Yu, B., Lam, W.H.K. and Tam, M.L., Bus arrival time prediction at bus stop with multiple routes. Transportation Research Part C: Emerging Technologies, 19(6), pp. 1157-1170, 2011. DOI: 10.1016/j.trc.2011.01.003

[3] Bontekoning, Y.M., Macharis, C. and Trip, J.J., Is a new applied transportation research field emerging?–A review of intermodal rail–truck freight transport literature. Transportation Research Part A: Policy and Practice, 38(1), pp. 1-34, 2004. DOI: 10.1016/j.tra.2003.06.001.

[4] Bianco, L. and La Bella, A., Freight transport planning and logistics: Lucio Bianco and Agostino La Bella: Springer, New York. Computers & Operations Research, 20(5), pp. 562-564, 1993. DOI: 10.1016/0305-0548(93)90047-M.

[5] Zhang, G., Eddy-Patuwo, B. and Y-Hu, M., Forecasting with artificial neural networks: The state of the art. International Journal of Forecasting, 14(1), pp. 35-62, 1998. DOI: 10.1016/S0169-2070(97)00044-7

[6] Diaz, R., Talley, W. and Tulpule, M., Forecasting empty container volumes. The Asian Journal of Shipping and Logistics, 27(2), pp. 217-236, 2011. DOI: 10.1016/S2092-5212(11)80010-6.

[7] Clark, S.D., Dougherty, M.S. and Kirby, H.R., The use of neural networks and time series models for short term traffic forecasting: A comparative study, Proceedings of PTRC 21st Summer Anual Meeting1993.

[8] Dougherty, M., A review of neural networks applied to transport. Transportation Research Part C: Emerging Technologies, 3(4), pp. 247-260, 1995. DOI: 10.1016/0968-090X(95)00009-8

[9] 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
[10] Bishop, C.M., Neural networks for pattern recognition. Oxford university press, 1995.
[11] Ke, G.Y., Li, K.W. and Hipel, K.W., An integrated multiple criteria preference ranking approach to the Canadian west coast port congestion conflict. Expert Systems with Applications, 39(10), pp. 9181-9190, 2012. DOI: 10.1016/j.eswa.2012.02.086.
[12] Fan, L., Wilson, W.W. and Dahl, B., Congestion, port expansion and spatial competition for US container imports. Transportation Research Part E: Logistics and Transportation Review, 48(6), pp. 1121-1136, 2012. DOI: 10.1016/j.tre.2012.04.006.
[13] Peng, W. and Chu, C., A comparison of univariate methods for forecasting container throughput volumes. Mathematical and Computer Modelling, 50(7–8), pp. 1045-1057, 2009. DOI: 10.1016/j.mcm.2009.05.027.
[14] Smith, B.L. and Demetsky, M.J., Short-term traffic flow prediction models-a comparison of neural network and nonparametric regression approaches, Systems, Man, and Cybernetics, 'Humans, Information and Technology', IEEE International Conference On, pp. 1706-1709, 1994.
[15] Vlahogianni, E.I., Golias, J.C. and Karlaftis, M.G., Short-term traffic forecasting: Overview of objectives and methods. Transport Reviews, 24(5), pp. 533-557, 2004. DOI: 10.1080/0144164042000195072
[16] Moscoso-Lopez, J.A., Ruiz-Aguilar, J.J., Turias, I., Cerbán, M. and Jiménez-Come, M.J., A comparison of forecasting methods for ro-ro traffic: A case study in the strait of Gibraltar, in Zamojski, W., Mazurkiewicz, J., Sugier, J., Walkowiak, T. and Kacprzyk, J., Eds. Springer International Publishing, 2014, pp. 345-353. DOI: 10.1007/978-3-319-07013-1_33.
[17] Ruiz-Aguilar, J. J., Turias, I. J., and Jiménez-Come, M. J., A two-stage procedure for forecasting freight inspections at Border Inspection Posts using SOMs and support vector regression. International Journal of Production Research, 53(7), pp. 2119-2130, 2015.
[18] Seabrooke, W., Hui, E.C.M., Lam, W.H.K. and Wong, G.K.C., Forecasting cargo growth and regional role of the port of Hong Kong. Cities, 20(1), pp. 51-64, 2003. DOI: 10.1016/S0264-2751(02)00097-5.
[19] Chou, C., Chu, C. and Liang, G., A modified regression model for forecasting the volumes of Taiwan's import containers. Mathematical and Computer Modelling, 47(9-10), pp. 797-807, 2008. DOI: 10.1016/j.mcm.2007.05.005.
[20] Fung, M.K., Forecasting Hong Kong's container throughput: An error-correction model. Journal of Forecasting, 21(1), pp. 69-80, 2002. DOI: 10.1002/for.818
[21] Klein, A., Forecasting the Antwerp maritime traffic flows using transformations and intervention models. Journal of Forecasting, 15(5), pp. 395-412, 1998. DOI: 10.1002/(SICI)1099-131X(199609)15:5<395::AID-FOR628>3.3.CO;2-Z
[22] Ruiz-Aguilar, J.J., Turias, I.J. and Jiménez-Come, M.J., A novel three-step procedure to forecast the inspection volume. Transportation Research Part C: Emerging Technologies, 56, pp. 393-414, 2015. DOI: 10.1016/j.trc.2015.04.024.
[23] Yang, Z., Chen, G. and Moodie, D.R., Modeling road traffic demand of container consolidation in a Chinese port terminal. Journal of Transportation Engineering, 136(10), pp. 881-886, 2010. DOI: 10.1061/(ASCE)TE.1943-5436.0000152
[24] Park, D. and Rilett, L.R., Forecasting multiple-period freeway link travel times using modular neural networks. Transportation Research Record: Journal of the Transportation Research Board, 1617, pp. 163-170, 1998. DOI: 10.3141/1617-23
[25] Karlaftis, M. and Vlahogianni, E., Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. Transportation Research Part C: Emerging Technologies, 19(3), pp. 387-399, 2011. DOI: 10.1016/j.trc.2010.10.004
[26] Ruiz-Aguilar, J., Turias, I. and Jiménez-Come, M., Hybrid approaches based on SARIMA and artificial neural networks for inspection time series forecasting. Transportation Research Part E: Logistics and Transportation Review, 67, pp. 1-13, 2014. DOI: 10.1016/j.tre.2014.03.009
[27] Lam, W.H., Ng, P.L., Seabrooke, W. and Hui, E.C., Forecasts and reliability analysis of port cargo throughput in Hong Kong. Journal of Urban Planning and Development, 130(3), pp. 133-144, 2004. DOI: 10.1061/(ASCE)0733-9488(2004)130:3(133)
[28] Gosasang, V., Chandraprakaikul, W. and Kiattisin, S., A comparison of traditional and neural networks forecasting techniques for container throughput at Bangkok port. The Asian Journal of Shipping and Logistics, 27(3), pp. 463-482, 2011. DOI: 10.1016/S2092-5212(11)80022-2.
[29] Robinson, S. and Polak, J.W., Modeling urban link travel time with inductive loop detector data by using the k-NN method. Transportation Research Record: Journal of the Transportation Research Board, 1935(1), pp. 47-56, 2005. DOI: 10.3141/1935-06
[30] Bhave, N. and Rao, P., Vehicle engine sound analysis applied to traffic congestion Estimation, Proc. of International Symposium on CMMR and FRSM2011, 2011.
[31] Ćavar, I., Kavran, Z. and Petrović, M., Hybrid approach for urban roads classification based on GPS tracks and road subsegments data. PROMET-Traffic & Transportation, 23(4), pp. 289-296, 2011. DOI: 10.7307/ptt.v23i4.131
[32] Park, T. and Lee, S., A Bayesian approach for estimating link travel time on urban arterial road network, in Anonymous Computational Science and its Applications–ICCSA 2004, Springer, 2004, pp. 1017-1025. DOI: 10.1007/978-3-540-24707-4_114
[33] Lee, H., Chowdhury, N.K. and Chang, J., A new travel time prediction method for intelligent transportation systems, Knowledge-Based Intelligent Information and Engineering Systems, pp. 473-483, 2008. DOI: 10.1007/978-3-540-85563-7_61
[34] Boyles, S., Fajardo, D. and Waller, S.T., Naive bayesian classifier for incident duration prediction, Transportation Research Board 86th Annual Meeting, 2007.
[35] Oh, J., Oh, C., Ritchie, S.G. and Chang, M., Real-time estimation of accident likelihood for safety enhancement. Journal of Transportation Engineering, 131(5), pp. 358-363, 2005. DOI: 10.1061/(ASCE)0733-947X(2005)131:5(358)
[36] Adam, T. and Untaroiu, C.D., Identification of occupant posture using a Bayesian classification methodology to reduce the risk of injury in a collision. Transportation research part C: emerging technologies, 19(6), pp. 1078-1094, 2011. DOI: 10.1016/j.trc.2011.06.006
[37] Kafai, M. and Bhanu, B., Dynamic Bayesian networks for vehicle classification in video. Industrial Informatics, IEEE Transactions on, 8(1), pp. 100-109, 2012.
[38] Khan, S.I. and Ritchie, S.G., Statistical and neural classifiers to detect traffic operational problems on urban arterials. Transportation Research Part C: Emerging Technologies, 6(5), pp. 291-314, 1998. DOI: 10.1016/S0968-090X(99)00005-4
[39] Abdulhai, B. and Ritchie, S.G., Enhancing the universality and transferability of freeway incident detection using a Bayesian-based neural network. Transportation Research Part C: Emerging Technologies, 7(5), pp. 261-280, 1999. DOI: 0.1016/S0968-090X(99)00022-4
[40] Abdel-Aty, M. and Pande, A., Identifying crash propensity using specific traffic speed conditions. Journal of Safety Research, 36(1), pp. 97-108, 2005.
[41] Wu, W., Qi-Sen, Z. and Mingjun, W., A method of vehicle classification using models and neural networks, Vehicular Technology Conference, 2001. VTC 2001, Spring. IEEE VTS 53rd, pp. 3022-3026, 2001.
[42] Duda, R.O., Hart, P.E. and Stork, D.G., Pattern Classification. New York: John Wiley & Sons, 2001.
[43] Ripley, B.D., Statistical ideas for selecting network architectures, in Anonymous Neural Networks: Artificial Intelligence and Industrial ApplicationsSpringer, 1995, pp. 183-190.
[44] Keinosuke, F.. Introduction to Statistical Pattern Recognition Academic Press, 1990.
[45] Tinsley, H.E. and Brown, S.D., Handbook of Applied Multivariate Statistics and Mathematical Modeling, Academic Press, 2000.
[46] Bishop, C.M., Pattern Recognition and Machine, Learning Springer New York, 2006.
[47] Rumelhart, D.E., Hinton, G.E. and Williams, R.J., Learning internal representations by error propagation, in Rumelhart, D.E. and

McClelland, J.L., Eds. Parallel Distributed Processing Cambridge, MA: MIT Press, 1986, pp. 318-362.
[48] Hornik, K., Stinchcombe, M. and White, H., Multilayer feedforward networks are universal approximators. Neural Networks, 2(5), pp. 359-366, 1989. DOI: 10.1016/0893-6080(89)90020-8
[49] Zhang, G.P., Neural networks for classification: A survey. Systems, Man and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 30(4), pp. 451-462, 2000. DOI: 10.1109/5326.897072.
[50] Kotsiantis, S.B., Zaharakis, I.D. and Pintelas, P.E., Machine learning: A review of classification and combining techniques. Artificial Intelligence Review, 26(3), pp. 159-190, 2006. DOI: 10.1007/s10462-007-9052-3.
[51] Fletcher, R., Practical Methods of Optimization, 2nd Ed. New York: Wiley, 1987.
[52] Hagan, M.T. and Menhaj, M.B., Training feedforward networks with the Marquardt algorithm. Neural Networks, IEEE Transactions on, 5(6), pp. 989-993, 1994.
[53] Friedman, M., The use of ranks to avoid the assumption of normality implicit in the analysis of variance. Journal of the American Statistical Association, 32(200), pp. 675-701, 1937. DOI: 10.1080/01621459.1937.10503522
[54] Friedman, M., A comparison of alternative tests of significance for the problem of m rankings. The Annals of Mathematical Statistics, 11(1), pp. 86-92, 1940. DOI: 10.1214/aoms/1177731944
[55] Salzberg, S.L., On comparing classifiers: Pitfalls to avoid and a recommended approach. Data mining and knowledge discovery, 1(3), pp. 317-328, 1997. DOI: 10.1023/A:1009752403260
[56] Pizarro, J., Guerrero, E. and Galindo, P.L., Multiple comparison procedures applied to model selection. Neurocomputing, 48(1–4), pp. 155-173, 2002. DOI: 10.1016/S0925-2312(01)00653-1.
[57] Turias, I.J., Gutiérrez, J.M. and Galindo, P.L., Modelling the effective thermal conductivity of an unidirectional composite by the use of artificial neural networks. Composites Science and Technology, 65(3), pp. 609-619, 2005. DOI: 10.1016/j.compscitech.2004.09.018
[58] Martin, M., Turias, I., Gonzalez, F., Galindo, P., Trujillo, F., Puntonet, C. and Gorriz, J., Prediction of CO maximum ground level concentrations in the Bay of Algeciras, Spain using artificial neural networks. Chemosphere, 70(7), pp. 1190-1195, 2008. DOI: 10.1016/j.chemosphere.2007.08.039
[59] Fisher, R.A., Statistical methods and scientific inference (2nd Ed.). Hafner Publishing Co., New York, 1959.
[60] Tukey, J.W., Comparing individual means in the analysis of variance. Biometrics, 5(2), pp. 99-114, 1949. DOI: 10.2307/3001913
[61] Carmer, S.G. and Swanson, M.R., An evaluation of ten pairwise multiple comparison procedures by Monte Carlo methods. Journal of the American Statistical Association, 68(341), pp. 66-74, 1973. DOI: 10.1080/01621459.1973.10481335
[62] Dietterich, T.G., Approximate statistical tests for comparing supervised classification learning algorithms. Neural Computation, 10(7), pp. 1895-1923, 1998. DOI: 10.1162/089976698300017197
[63] Ruiz-Aguilar, J.J., Turias, I.J., Jiménez-Come, M.J. and Cerbán, M.M., Hybrid approaches of support vector regression and SARIMA models to forecast the inspections volume. Hybrid Artificial Intelligence Systems, pp. 502, 2014. DOI: 10.1007/978-3-319-07617-1_44

J.J. Ruiz-Aguilar, received his BSc. Eng. in Civil Engineering in 2006, his MSc. in Civil Engineering in 2008, MSc. in Computational Modeling in Engineering in 2010 and MSc. in Logistics and Port Management in 2012, and his PhD in Civil Engineering in 2014. From 2009 to 2010 he worked for consulting companies in the civil sector and since 2010 has worked for the University of Cádiz, Spain. He is currently a lecturer in the Department of Industrial Engineering and Civil Engineering at the Engineering School of Algeciras, and he is the coordinator of the Port Engineering section of the MSc in Logistics and Port Management, University of Cádiz, Spain. His present interests lie in the field of soft computing, simulation, modeling, forecasting and its applications in civil and logistics problems.
ORCID: 0000-0002-2170-0693

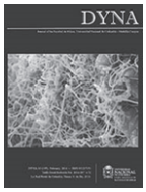
I.J. Turias, received his BSc. and MSc. in Computer Science from the University of Málaga, Spain, and his PhD. in Industrial Engineering in 2003 from the University of Cádiz, Spain. He is currently a professor (reader or associate professor) in the Department of Computer Engineering at the University of Cádiz, Spain. His present interests lie in the field of soft computing and its applications in industrial, environmental and logistics problems. He has coauthored numerous technical journals and conference papers, which are the result of his participation and leadership in research projects. He has also served as peer reviewer of several journals and conference proceedings. He has been contracted by a number of companies. He also was the Head of the Engineering School of Algeciras from 2003 to 2011. He is currently the principal researcher of the research group of Intelligent Modeling of Systems.
ORCID: 0000-0003-4627-0252

J.A. Moscoso-López, received his BSc. Eng. in Civil Engineering in 2001 and his PhD. in Engineering in 2013 from the University of Cádiz, Spain and his MSc. in Civil Engineering in 2003 from the Alfonso X el Sabio University, Spain. From 2003 to 2009, he worked for civil construction and consulting companies within the civil engineering (construction) sector, and from 2009 he has worked for the University of Cádiz, Spain. He is currently a lecturer in the Department of Industrial Engineering and Civil Engineering at the Engineering School of Algeciras. His research interests include simulation, modeling and forecasting nonlinear time-series in ports and logistics environments.
ORCID: 0000-0002-0080-0572

M.J. Jiménez-Come, after completing her BSc. in Chemical Engineering in 2007, from the University of Malaga, Spain, she obtained a PhD. in Engineering and Architecture Programing in 2013, from the University of Cadiz, Spain. Since 2010, she has been working at Cadiz University, Spain, where she is currently a lecturer in the department of Civil and Industrial Engineering. She is a Postdoctoral researcher in the Intelligent Modeling of Systems Research Group. Her research interests include modeling and simulation of industrial process using statistical and computational intelligence techniques.
ORCID: 0000-0003-0598-2544

M. Cerbán, obtained a BSc. of Economics in 1994 from the University of Seville, Spain and a PhD in Economics in 2006, from University of Cadiz, Spain. She has been a lecturer of applied economics at the University of Cadiz since 1998, since 1999 she has Director of the MSc. in Logistics and Port Management at the University of Cadiz, Spain and Co-Director of the Interuniversity Master in Port Management and Intermodality at the Universities of Cadiz, La Coruña, Oviedo, Politécnica de Madrid and Puertos del Estado since 2011. Her areas of research include Port Economics, Maritime Transports and Logistics.
ORCID: 0000-0003-4674-0453

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS



DYNA
http://dyna.medellin.unal.edu.co/

A mixed-integer linear programming model for harvesting, loading and transporting sugarcane. A case study in Peru

Marcela María Morales-Chávez [a], José A. Soto-Mejía [b] & William Sarache [c]

[a] *Programa de Ingeniería Comercial, Universidad Libre Seccional Pereira, Pereira, Colombia. mmorales@unilibrepereira.edu.co*
[b] *Facultad de Ingeniería Industrial, Universidad Tecnológica de Pereira, Pereira, Colombia. jomejia@utp.edu.co*
[c] *Facultad de Ingeniería y Arquitectura, Universidad Nacional de Colombia, Manizales, Colombia. wasarachec@unal.edu.co*

Received: March 4th, de 2015. Received in revised form: August 12th, 2015. Accepted: August 28th, 2015

Abstract
Due to opportunities for economic and social development in the biofuels market, improvement to the supply chain has become a relevant matter. In agro-industrial supply chains, procurement costs are highly relevant. Since sugarcane is a high performance raw material for ethanol production, this paper proposes a Mixed-Integer Linear Programming Model for cost optimization for harvesting, loading and transportation operations. The model determines the quantity of machines and workers to meet the biofuel plant requirements. Costs of resources for harvesting and loading as well as transportation costs from the land parcel to the production plant are minimized. Also, the model calculates the cost of penalties for shortages (unmet demand) and the cost of equipment idle time. The implementation of the model in a Peruvian biofuels company, showed a cost reduction of around 11 % when compared to the current costs.

Keywords: mixed-integer lineal programming; supply chain planning; sugarcane; biofuels.

Modelo de programación lineal entera mixta para el corte, carga y transporte de caña de azúcar. Un caso de estudio en Perú

Resumen
Debido a las oportunidades de desarrollo económico y social del mercado de los biocombustibles, el mejoramiento de su cadena de suministro se ha convertido en un tema altamente relevante. Dado que la caña de azúcar es una de las materias primas de mayor rendimiento para la producción de etanol, el presente artículo propone un modelo de Programación Lineal Entera Mixta para optimizar los costos en las operaciones de corte, cargue y transporte. El modelo determina la cantidad de máquinas y trabajadores para satisfacer los requerimientos de la planta de biocombustible. Se minimizan los costos de asignación de equipos, costos de transporte y adicionalmente se consideran los costos de penalización por demanda no satisfecha y por ociosidad de los equipos disponibles. La aplicación de este modelo en una empresa Peruana, presentó un porcentaje promedio de disminución de costos del 11 % al ser comparados con los costos actuales de la empresa.

Palabras clave: programación lineal entera mixta; planeación de cadenas de abastecimiento; caña de azúcar; biocombustibles.

1. Introduction

Biofuel supply chains have been identified as a strategic sector for Latin America [1]. Some countries are stimulating biofuel production to reduce fossil fuel dependency and, in this way, guarantee their energy security at lower prices [2-4]. Although several investigations on biofuel production have focused on identifying a more efficient feedstock, sugarcane is considered to be a high yield biomass in the production process [5].

In Peru, biofuel production is considered to be one of the seven most important sectors [6]; therefore, ethanol production from sugarcane is highly relevant for economic growth projections. The supply chain is made up of a group of companies involved in the flow of materials, information and capital, starting with the unprocessed raw materials and finishing with the end consumer [7-9]. Fig. 1 shows the general structure of the biofuel supply chain from sugarcane.

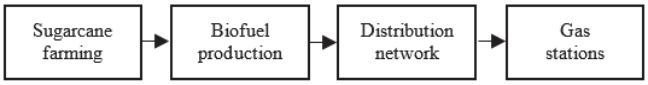


Figure 1. Supply chain structure
Source: Own elaboration

Bearing in mind that between 47% and 58% of ethanol production costs depend on the sugarcane crop [10], this stage becomes an important target to reduce costs in the entire supply chain. Some operations related to harvesting and loading cane are relevant. Also, given the complex nature of agro-industrial processes [9], it is important to consider the penalties derived from unmet demand and idle machines. In addition, harvesting, loading and transporting sugarcane are operations commonly affected by several constraints, such as land conditions, grinding requirements, availability and resource capacity (machinery and manpower) and production scheduling. Therefore, identifying the optimal resource allocation to reduce costs without affecting delivery goals is a decision of great complexity and economic relevance in supply chain planning.

In this sense, this paper shows a Mixed-Integer Linear Programming Model (MILP), for operations planning for the harvesting, loading and transportation of sugarcane to supply a biofuel production plant. The model takes into account optimization of four types of costs: allocation of machinery and workforce, transportation from the farm to the biofuel production plant, penalties for unfulfilled orders and idle machinery. Also, in the case of the mechanical harvesting method, land conditions for machinery selection are considered. The model was implemented in a case study in Peru, resulting in a cost reduction of around 11 % when compared to the current costs.

In order to explain the model, this paper has been structured as follows: First, a literature review is presented in Section 2. Several papers that discuss the improvement of a biofuel supply chain in the upstream stage are examined. Second, in Section 3, the structure of the proposed optimization model is explained in two phases: in the first, the general procedure is described by identifying its purpose, as well as the resources and costs that must be taken into account; in the second, the structure and some details of the mathematical model are explained. Third, in Section 4, a case study is solved using a well-known computational tool. Finally, in Section 5, the conclusions and some suggestions for future model applications are outlined.

2. Literature review

Biofuel supply chain improvement has been a matter of great interest in recent years. A systematic literature review was carried out using the ISI – Web of Science and SCOPUS databases. The search was confined to identifying papers that were concerned with the specific objective of the research presented in this paper. About 43 papers associated with the selected key words were found; however only 21 of these were considered to be relevant. According to the results, an important increase in contributions on this topic, especially over the last five years (73.9%), was detected. Most of these studies have been carried out in the United States (28%), Italy (12%), France and Cuba (8%). Papers on this subject in relation to Peru were not found. Table 1 summarizes a brief comparison of the 21 analyzed papers based on the scope of the objective function and the set of decision variables.

Table 1.
Literature Review of optimization models of sugarcane supply-chains

Objective function	**References**																				
	11	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**
Economic	x	x	x	x		x	x	x	x			x	x	x	x		x	x	x	x	
Government policies					x										x				x		
Environment	x								x		x		x	x	x				x		
Social aspects															x				x		
Computational response															x	x					
Type of biomass	x			x		x	x		x	x	x		x		x		x	x			x
Production capacity	x	x		x	x	x				x	x				x		x	x			
Number of production facilities	x						x	x		x	x			x	x						
Technology selection	x		x				x					x	x		x						x
Transportation fleet size						x		x									x	x		x	
Location of production facilities		x		x						x	x				x						
Machinery availability in the field								x									x	x		x	
Financial performance	x	x													x						
Number of land parcels to harvest																				x	
Land parcels assignment per type of crop		x																			

Source: Own elaboration based on the quoted authors

According to Table 1, the focus of a significant part of the studies is on supply chain design that included biomass type, technology selection, capacity allocation and the facilities required for the production stage. In other words, few contributions on supply chain planning were found. Regarding the first echelon of the supply chain (upstream), few studies examined the characteristics of the typical logistics operations for sugarcane, such as harvesting, loading and transportation from the land parcel to the production plant. It is important to emphasize that 91.3% of the articles examine operations where the crop is an input parameter, which reinforces the originality of this article where land parcel selection for sugarcane cultivation is a decision variable.

Additionally, the economic criterion is the most common objective function defined for sugarcane supply-chain optimization, although some of the articles take into account economic factors and environmental impact simultaneously. From the economic point of view, 21.41% of the studies aim to maximize expected net revenues and the rest are focused on cost minimization, emphasizing transportation, harvesting and warehousing operations. Likewise, only two researchers addressed opportunity cost analysis taking into account sugarcane maturity and crop investment.

Finally, models considering cost of equipment downtimes and cost of penalties for unmet demand were not identified in the literature review. Models focused on resource allocation (machinery and labor) to land parcels based on terrain constraints were not detected.

3. Structure of the model

The purpose of the mathematical model is to select the quantity of hectares to be collected per land parcel. It also allows for the selection of the type of harvesting method from three alternatives: mechanical, semi-mechanical and manual. In addition, the model selects the type of equipment for harvesting and loading operations, as well as the workforce required for a semi-mechanical or manual method. Table 2 shows the subscript indices used in the model; in turn, Tables 3 and 4 summarize the decision variables and the parameters respectively.

3.1. Objective function

The cost parameters in the objective function were grouped as follows:

- Machinery costs (harvesting and loading machines). This includes the operation costs and the cost of equipment idle time. The latter is obtained based on the fixed cost of unassigned machinery.
- Workforce costs for cut and loading operations.
- Costs related to harvesting method. This includes maintenance personnel costs, staff and other inputs.
- Penalty costs for unfulfilled demand. This cost is agreed in the supply contract.
- Transportation costs.

The main purpose of the model is the minimization of the total costs, which is determined by:

Table 2.
Subscript indices.

Subscripts	Variation
Loading machine *p*	p: 1,2,3,……v
Harvesting machine *r*	r: 1,2,3,……z
Land parcel *i*	i: 1,2,3,…….n
Harvesting method j	j = 1 (Mechanical); j = 2 (Semi-mechanical)*; j = 3 (Manual)

*Sugarcane is harvested manually; then, loading operations are carried out by loading machines.
Source: Own elaboration based on the quoted authors

Table 3.
Decision variables.

Symbol	Description
A_{pi}	Number of loading machines *p* assigned to the land parcel *i* during the week.
AO_p	Number of unassigned loading-machines *p* (idle machines) during the week.
EA_i	Number of employees assigned to the land parcel *i* for manual loading during the week.
K_{ri}	Number of harvesting machines *r* assigned to the land parcel *i* during the week (mechanical method)
KO_r	Number of unassigned harvesting-machines *r* (idle machines) during the week.
N	Quantity of sugarcane not sent to the production plant (t/week).
Q_i	Number of employees assigned to land parcel *i* for manual harvesting during the week.
QS_i	Number of employees assigned to land parcel *i* for semi-mechanical harvesting during the week.
X_{ij}	Size of land parcel *i* to be harvested by the method *j* during the week (ha)

Source: Own elaboration based on the quoted authors

Min (Costs) = Operation cost per harvesting method + Transportation cost for sugarcane delivered from each land parcel to the production plant + Penalty cost for unfulfilled demand + Cost per assigned harvesting machine + Cost per idle harvesting machine + Cost per assigned loading machine + Cost per idle loading machines + Cost of assigned workforce for manual harvesting + Cost of assigned workforce for manual loading.

$$\begin{aligned} MIN\_Z = & \sum_{j=1}^{3} CH_j \left(\sum_{i=i}^{n} X_{ij} \right) + \sum_{i=i}^{n} CT_i (TH_i) \left(\sum_{j=1}^{3} X_{ij} \right) + \\ & CN\ (N) + \sum_{r=1}^{z} \left(CK_r \sum_{i=1}^{n} K_{ri} \right) + \sum_{r=1}^{z} (CKO_r KO_r) + \\ & \sum_{p=1}^{v} \left(CA_p \sum_{i=1}^{n} A_{pi} \right) + \sum_{p=1}^{v} (CAO_p AO_p) + \\ & CQ \sum_{i=1}^{n} (QS_i + Q_i) + CEA \sum_{i=1}^{n} EA_i \end{aligned} \quad (1)$$

3.1. Model constraints

Model constraints are represented by equations 2 to 14. The set of constraints includes land availability, production plant demand, equipment capacity, workforce availability and land conditions.

The system of equations is as follows:

Table 4.
Parameters.

Symbol	Description	Unit
CA_p	Operating cost per assigned loading machine *p* during the week	US$/Week
CAO_p	Operating cost per unassigned loading machine during the week.	US$/Week
CEA	Operating cost per assigned employee during the week	US$/Week
CH_j	Fixed cost of the harvesting method *j*	US/ha
CK_r	Operating cost per assigned harvesting machine *r* during the week	US$/Week
CKO_r	Operating cost per unassigned harvesting machine during the week	US$/Week
CN	Penalty cost for unfulfilled demand	US$/t
CQ	Labor cost (for manual and semi-mechanical harvesting)	US$/person-week
CT_i	Transportation cost from land parcel i to production plant	US$/t
HD_i	Available area of the land parcel *i*	ha
MAS_p	Capacity of the loading machine *p*	t/week
MEAS	Loading capacity per employee	t/week
MKSr	Capacity of the harvesting machine *r*	t/week
MQS	Harvesting capacity per employee	t/week
NAA_p	Availability of loading machine *p*	Number of machines/week
NEA	Availability of employees for manual loading	Number of employees/week
NK_r	Availability of harvesting machine *r*	Number of machines/week
NQS	Availability of employees for manual harvesting	Number of employees/week
TH_i	Quantity of expected sugarcane per ha of land parcel *i*	t/ha
TMS	Maximum quantity of sugarcane delivered to the production plant	t/week
TRS	Minimum quantity of required sugarcane	t/week

Source: Own elaboration based on the quoted authors

Land:
Available area per land parcel (ha)

$$\sum_{j=1}^{3} X_{ij} \leq HD_i \qquad \forall i \tag{2}$$

Demand:

- Minimum amount of sugarcane required (t/week)

$$\sum_{i=1}^{n}\sum_{j=1}^{3} TH_i X_{ij} \geq TRS \tag{3}$$

- Maximum amount of sugarcane to be sent to the production plant (t/week)

$$\sum_{j=1}^{3}\sum_{i=1}^{n} TH_i X_{ij} + N = TMS \tag{4}$$

Equipment availability:

- Harvesting machines availability (quantity/week)

$$\left(\sum_{i=1}^{n} K_{ri}\right) + KO_r = NK_r \qquad \forall r \tag{5}$$

- Loading machines availability (quantity/week)

$$\left(\sum_{i=1}^{n} A_{pi}\right) + AO_p = NAA_p \qquad \forall p \tag{6}$$

Workforce availability:

- Workers' availability for harvesting (quantity/week): workers are divided into two groups; first one of them takes part in the semi-mechanical method; the second one is assigned to the manual method.

$$\sum_{i=1}^{n}\left(QS_i + Q_i\right) \leq NQS \tag{7}$$

- Workers' availability for manual loading (quantity/week)

$$\sum_{i=1}^{n} EA_i \leq NEA \tag{8}$$

Equipment capacity:

The equal symbol in the capacity constraints is required to achieve maximum equipment exploitation when assigned to every land parcel (uptime - maintenance time). The uptime for loading machines depends on the maximum time that sugarcane can wait to be processed after being cut without affecting its yield.

- Harvesting machines capacity per land parcel (t/week)

$$TH_i\left(X_{i1}\right) = \sum_{r=1}^{z} MKS_r\left(K_{ri}\right) \qquad \forall i \tag{9}$$

- Loading machines capacity per land parcel (t/week)

$$MQS\left(QS_i\right) = \sum_{p=1}^{v} MAS_p\left(A_{pi}\right) \qquad \forall i \tag{10}$$

Workforce capacity:

There are two groups of workers for harvesting operations: group 1 is assigned to semi-mechanical harvesting and group 2 is assigned to manual harvesting. The constraints are:

- Workers' capacity for semi-mechanical harvesting in every land parcel (t/week)

$$TH_i\left(X_{i2}\right) = MQS\left(QS_i\right) \quad \forall i \tag{11}$$

- Workers' capacity for manual harvesting in every land parcel (t/week)

$$TH_i\left(X_{i3}\right) \leq MQS\left(Q_i\right) \qquad \forall i \tag{12}$$

- Workers' capacity for manual loading in every land parcel (t/week)

$$MQS\left(Q_i\right) \leq MEAS\left(EA_i\right) \quad \forall i \tag{13}$$

Land conditions:

The proper operation of harvesting machines (powered by a caterpillar engine or wheel-mounted) depends on the land parcel characteristics; for example, wheel-mounted

machines encounter difficulties when assigned to wetlands. Therefore, depending on the foreman assessment of each type of land, the following constraint will come in to play:

- Harvesting machines t unassigned to the land parcel i due to land conditions.

$$K_{ri} = 0 \tag{14}$$

Non- negativity condition and integer variables:

All variables are constrained to values greater than or equal to zero. Furthermore, the following variables are constrained to integer values.

$$A_{pi}, AO_p, EA_i, K_{ri}, KO_r, Q_i, QS_i$$

4. Case study

This model was implemented in a Peruvian ethanol production company. The supply chain was comprised of sugarcane harvesting and loading operations, raw material transportation, the ethanol production process and finally, transportation to the international market. Although the company had several land parcels planted with sugarcane, the model was applied to support the operations programming of harvesting, loading and transportation for only two of these. This decision was made because, when the study was conducted, the generated raw material (41,300 t) exceeded the ethanol plant requirement (28,000 t). However, the model can be adjusted insofar as ethanol demand increases. Table 5 shows the input parameters given by the company; however, information about costs was omitted due to confidentiality agreements with the company.

By using GAMS (General Algebraic Modeling System) professional software, the model was solved. The model was run on a microcomputer provided with 1.66 GHz Intel processor and 4.6 GB of memory RAM. The problem was solved efficiently since the computational time was of a thousandth of second.

According to the results, 300 ha of land parcel 1 should be harvested using the mechanical method, assigning five machines for this operation (three Case brand, one John Deere brand powered by caterpillar engine and one John Deere brand powered by wheel-mounted). This outcome highlights another advantage of the proposed model, which allows for programming of different types of machines for this operation. Also, 40 ha of land parcel 1 must be processed using the manual method, which implies the assignment of 147 workers to harvesting and loading operations. Regarding land parcel 2, the model chose the mechanical method to collect 60 ha; however, due to terrain conditions, the harvesting machine powered by the caterpillar engine was assigned instead of the wheel-mounted machine. Based on this resource allocation, 28,000 tons of sugarcane (23,800 from land parcel 1 and 4,200 from land parcel 2) can be processed, satisfying 100% of the ethanol plant requirements.

Although the model penalizes equipment idle time, the semi-mechanical method was not included in the final solution, even though four loading machines and one harvesting machine are unnecessary. This decision shows the existing overcapacity in the agricultural echelon of the supply chain. As a consequence, the company must increase its marketing efforts. From another point of view, the company can take advantage of idle machines to implement preventive maintenance activities. The obtained total cost was approximately 11% less than the current costs, showing significant savings for the company. Table 6 summarizes the obtained solution.

Table 5.
Input parameters of the case study

TIME AVAILABILITY			
	Labor	Harvesting machine	Loading Machine
Hours/day	8	20	12
Days/week	7	7	7
Maintenance (Hours/week)		4	4
HARVESTING			
	Mechanical method		Manual method
	Case brand	Jhon Deere brand	Employee
Availability	4	3	200
Capacity (t/hour)	30	30	0.625
Capacity (t/day)	600	600	5
Capacity (t/week)	4200	4200	35
LOADING			
	Mechanical method		Manual method
	Cameco brand	John Deere brand	Employees
Availability	2	2	200
Capacity (t/hour)	24	24	0.75
Capacity (t/day)	288	288	6
Capacity (t/week)	2016	2016	42
DEMAND			
Minimum: 17500		Maximum: 28000	
LAND			
	Land parcel 1		Land parcel 2
Availability (ha)	500		90
Expected sugarcane (t/ha)	70		70
Land constraints	None		Humidity

Source: Own elaboration based on the quoted authors

Table 6.
Final results

NUMBER OF HARVESTED AND LOADED HECTARES		
	Land parcel 1	Land parcel 2
Mechanical	300	60
Semi-mechanical	0	0
Manual	40	0

RESOURCE ASSIGNMENT PER WEEK			
	Land parcel 1	Land parcel 2	Unassigned
Number of harvesting machines (Case brand)	3	0	1
Number of harvesting machines (John Deere brand; powered by caterpillar engine)	1	1	0
Number of harvesting machines (John Deere brand; powered by wheel-mounted)	1	0	0
Number of workers for semi-mechanical harvesting	0	0	
Number of workers for manual harvesting	80	0	
Number of loading machines (Cameco brand)	0	0	2
Number of loading machines (John Deere brand)	0	0	2
Number of workers for manual loading	67	0	

DEMAND SATISFACTION			
	Land parcel 1	Land parcel 2	Unfulfilled
Quantity of sugarcane (t)	23.800	4.200	0

Source: Own elaboration based on the quoted authors

5. Conclusions

Biofuel supply chain improvement has been a matter of great interest in recent years. Although the literature review shows an important number of papers on this topic, most of them are aimed at supporting decisions regarding supply chain design. The analysis showed that a significant part of the studies oriented to supply chain design, considered the biomass type, technology selection, capacity allocation and facilities as the most important decision variables; however few papers on optimization models for supply chain planning were found, specifically those oriented toward analyses of the particularities of the first echelon (upstream) of the sugarcane supply chain.

For instance, few studies examined the characteristics of the typical logistics operations for sugarcane, such as harvesting, loading and transportation from the land parcel to the production plant. Also, models that took into account the cost of equipment downtime and the cost of penalties for unmet demand were not identified in the literature review. Models oriented to resource allocation (machinery and labor) to the land parcels based on the terrain constraints were not detected.

In contrast, the proposed model shows several advantages that can be summarized as follows: 1) it analyzes operations programming related to harvesting, loading and transporting sugarcane, using land parcel selection as a decision variable; 2) the model considers the optimization of four types of costs: allocation of machinery and workforce, transportation from the farm to the biofuel production plant, penalties for unmet orders and idle machinery; 3) in the case of the mechanical harvesting method, terrain constraints for machinery selection were taken into account.

According to the results obtained in the case study, the current cost of the company was reduced by approximately 11%. By analyzing three alternatives for harvesting and loading operations (mechanical, semi-mechanical and manual), 300 ha with the mechanical method and 96 ha with the manual method were assigned. Because the model did not take into account the use of the semi-mechanical alternative, four loading machines and one harvesting machine were not used affecting the idle time cost. In general, the model established the required resources to support the operations programming in the first echelon of the supply chain; thus, the results show the number of machines, work force and staff necessary to meet the raw material requirements of the production plant.

Due to the level of complexity of the present model in relation to the number of land parcels analyzed, computer time was not a problem; nevertheless the effect on the set of variables and constraints must be checked for a greater amount of land parcels. The model shows some disadvantages that suggest some future research lines. For example, it is necessary to analyze other features related to operations such as internal transportation, yield for several sugarcane varieties and other soil constraints. Other variables such as demand uncertainty and multi-period and environmental conditions could be analyzed in order to provide a better tool to support decision-making.

Acknowledgments

The authors wish to thank the Universidad Tecnológica de Pereira for its academic and financial support through the research project assigned code number 1110-622-38514 (Colciencias).

References

[1] Pareja, P.C., Sevilla, S. y Coello, J., Estudio sobre la situación de los biocombustibles en el Perú, Lima, [Online] 2008, 68 P. Available at: http://www.cedecap.org.pe/uploads/biblioteca/48bib_arch.pdf

[2] Franco, C.J., Flórez, A.M. y Ochoa, M.C., Análisis de la cadena de suministro de biocombustibles en Colombia. Revista de Dinámica de Sistemas, 4(2), pp. 109-133, 2008.

[3] Duarte, A., Sarache, W.A.and Cardona, C., Cost analysis of the location of Colombian biofuels plants. DYNA, 79(176), pp. 71-80, 2012.

[4] Cortés-Marín, E., Suarez-Mahecha, H. and Pardo-Carrasco, S., Biocombustibles y autosuficiencia energética. DYNA, [Online] 76(158), pp. 101-110. 2009 Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-75249097474&partnerID=tZOtx3y1

[5] Moncada, J., El-Halwagi, M.M. and Cardona, C.A., Techno-economic analysis for a sugarcane biorefinery: Colombian case. Bioresource Technology, 135, pp. 533-543, 2013. DOI: 10.1016/j.biortech.2012.08.137.

[6] Ministerio de Educación del Perú. Plan nacional estratégico de ciencia, tecnología e innovación para la competitividad y el desarrollo humano PNCTI 2006-2021, Lima, [Online]. 2014. Available at: http://www.minedu.gob.pe/normatividad/reglamentos/PlanNacionalCTI-CDH2006-2021.php.

[7] Council of Supply Chain Management Professionals. Supply Chain Management, [Online]. 2013. Available at: https://cscmp.org/about-us/supply-chain-management-definitions.

[8] Sarache, W.A., Costa, Y. and Martinez, J., Environmental performance evaluation under a green supply chain approach. DYNA, 82(189), pp. 207-215, 2015. DOI: 10.15446/dyna.v82n189.48550.

[9] Duarte, A., Sarache, W.A. and Costa, Y., A facility-location model for biofuel plants: Applications in the Colombian context. Energy, 72, pp. 476-483, 2014. DOI: 10.1016/j.energy.2014.05.069.

[10] Consejo Nacional de Política Económica y Social. Conpes 3510. Lineamientos de política para promover la producción sostenible de biocombustibles en Colombia, Bogotá,[Online]. 2008, 44 P. Available at: http://www.minminas.gov.co/minminas/downloads/UserFiles/File/Conpes 3510.pdf

[11] Marvin, W.A., Schmidt, L.D., Benjaafar, S., Tiffany, D.G. and Daoutidis, P., Economic optimization of a lignocellulosic biomass-to-ethanol supply chain. Chemical Engineering Science, 67(1), pp. 68-79, 2012. DOI: 10.1016/j.ces.2011.05.055

[12] Giarola, S., Bezzo, F. and Shah, N., A risk management approach to the economic and environmental strategic design of ethanol supply chains. Biomass and Bioenergy, 58, pp. 31-51, 2013. DOI: 10.1016/j.biombioe.2013.08.005

[13] Osmani, A. and Zhang, J., Stochastic optimization of a multi-feedstock lignocellulosic-based bioethanol supply chain under multiple uncertainties. Energy, 59, pp. 157-172, 2013. DOI: 10.1016/j.energy.2013.07.043

[14] Ortiz-Gutierrez, R.A., Giarola, S. and Bezzo, F., Optimal design of ethanol supply chains considering carbon trading effects and multiple technologies for side-product exploitation. Environmental Technology, 34(13-14,SI), pp. 2189-2199, 2013. DOI: 10.1080/09593330.2013.829111

[15] Sparks, G.D., Ortmann, G.F. and Lagrange, L., An economic evaluation of soybean-based biodiesel production on commercial farms in Kwazulu-natal, South Africa. Agrekon, 50(3), pp. 68-89, 2011. DOI: 10.1080/03031853.2011.617862

[16] Shastri, Y., Hansen, A., Rodríguez, L. and Ting, K.C., Development and application of biofeed model for optimization of herbaceous biomass feedstock production. Biomass and Bioenergy, 35(7), pp. 2961-2974, 2011. DOI: 10.1016/j.biombioe.2011.03.035

[17] Giarola, S., Patel, M. and Shah, N., Biomass supply chain optimisation for Organosolv-based biorefineries. Bioresource Technology, 159, pp. 387-396. 2014. DOI: 10.1016/j.biortech.2014.02.109

[18] Lin, T., Rodríguez, L.F., Shastri, Y.N., Hansen, A.C. and Ting, K.C., Integrated strategic and tactical biomass-biofuel supply chain optimization. Bioresource Technology, 156, pp. 256-66, 2014. DOI: 10.1016/j.biortech.2013.12.121

[19] Balaman, Ş.Y. and Selim, H., Multiobjective optimization of biomass to energy supply chains in an uncertain environment. Computer Aided Chemical Engineering, 33, pp. 1267-1272, 2014. DOI: 10.1016/B978-0-444-63455-9.50046-5

[20] Yoda, K., Furubayashi, T. and Nakata, T., Design of automotive bioethanol supply chain using mixed integer programming. Nihon Enerugi Gakkaishi/Journal of the Japan Institute of Energy, 92(11), pp. 1173-1186, 2013. DOI: 10.3775/jie.92.1173

[21] Ivanov, B.B., Dimitrova, B. and Dobrudzhaliev, D., Optimal location of biodiesel refineries: The Bulgarian scale. Journal of Chemical Technology and Metallurgy, 48(5), pp. 513-523, 2013.

[22] Ortiz-Gutiérrez, R., Penazzi, S., Bernardi, A.L., Giarola, S. and Bezzo, F., A spatially-explicit approach to the design of ethanol supply chains considering multiple technologies and carbon trading effects. Computer Aided Chemical Engineering, 32, pp. 643-648, 2013. DOI: 10.1016/B978-0-444-63234-0.50108-1

[23] Čuček, L., Martín, M.J.P., Grossmann, I.E. and Kravanja, Z., Multi-objective optimization of a biorefinery's supply network. AIChE Annual Meeting, Pittsburgh, 2012.

[24] Song, H., Dotzauer, E., Thorin, E., Guziana, B., Huopana, T. and Yan, J., A dynamic model to optimize a regional energy system with waste and crops as energy resources for greenhouse gases mitigation. Energy, 46(1), pp. 522-532, 2012. DOI: 10.1016/j.energy.2012.07.060

[25] You, F. and Wang, B., Optimal design and operations of cellulosic biofuel supply chains under uncertainty. AIChE Annual Meeting, Minneapolis, 2011.

[26] Shastri, Y.N., Hansen, A.C., Rodríguez, L.F. and Ting, K.C., A novel computational approach to solve complex optimization problems involving multiple stakeholders in biomass feedstock production. American Society of Agricultural and Biological Engineers Annual International Meeting, Pittsburgh, 2010, pp. 542-555.

[27] Shastri, Y.N., Hansen, A.C., Rodríguez, L.F. and Ting, K.C., Biomass feedstock production and provision: A system level optimization approach. AIChE Annual Meeting, Nashville, 2009.

[28] Shastri, Y.N., Domdouzis, K., Hu, M., Hansen, A.C., Rodríguez, L.F. and Ting, K.C., System level analysis of biomass feedstock production for bioenergy sector. American Society of Agricultural and Biological Engineers, Annual International Meeting, Reno, 2009, pp. 2203-2222.

[29] Rozakis, S., Sourie, J. and Vanderpooten, D., Integrated micro-economic modelling and multi-criteria methodology to support public decision-making : The case of liquid bio-fuels in France. Biomass and Bioenergy, 20, pp. 385-398, 2001. DOI: 10.1016/S0961-9534(01)00004-6

[30] Malik, S.B., Satsangi, P.S., Tripathy, S.C. and Balasubramanian, R., Mathematical model for energy planning of rural India. International Journal of Energy Research, 18(4), 469–482, 1994.

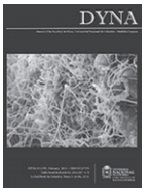
[31] Tripathy, S. C., Satsangi, P. S., Balasubramanian, R. and Malik, S. B. Artificial neural network application to energy system planning. International Journal of Engineering Intelligent Systems for Electrical Engineering and Communications, 7(3), pp. 121-126, 1999.

M.M. Morales-Chávez, received a BSc.an Industrial Engineering in 2006 and the MSc. in Operations Research in 2011 both from the Universidad Tecnologica de Pereira, Colombia; she is currently enrolled in the PhD program in Engineering at Universidad Nacional de Colombia, Sede Manizales, Colombia. From 2006 until today, she has been working as a consultant in several companies. Currently, she is professor in the Comercial Engineering Program, Universidad Libre-Seccional Pereira. Her research interests include: operational research, supply chain optimization, business logistic and operations management.
ORCID: 0000-0002-7384-8745

J.A. Soto-Mejía, received a BSc. Sciences in Physics in 1980 and an MSc. in physics in 1982, both from the Kharkov Maximo Gorki University, Russia. In 2002 he received a PhD in Computational Engineering from the Universidade Estadual De Campinas, Brasil. From 1991 he has worked as full professor at Universidad Tecnológica de Pereira, Colombia. His research interests include: operational research, multivariate analysis, simulation and supply chain optimization.
ORCID: 0000-0002-0205-6863

W. Sarache, received a BSc. in Industrial Engineering in 1993 from Universidad de Ibagué, Ibagué Colombia; thereafter he received an MSc. in Industrial Engineering in 1998 and a PhD in 2003 both from the Universidad Central de Las Villas, Cuba. From 1992 to 1999 he worked as operations manager in manufacturing companies. From 2000 until today he has been professor at Universidad Nacional de Colombia. His research interests include: operations management, supply chain management and business logistic.
ORCID: 0000-0003-3543-4151.

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS



DYNA
http://dyna.medellin.unal.edu.co/

Voltage sag assessment using an extended fault positions method and Monte Carlo simulation

Jorge W. Sagre [a], John E. Candelo [b] & Johny H. Montaña [c]

[a] *Grupo de Investigación en Sistemas Eléctricos de Potencia, Universidad del Norte, Barranquilla, Colombia. jsagre@uninorte.edu.co*
[b] *Departamento de Energía Eléctrica y Automática, Universidad Nacional de Colombia, Medellín, Colombia. jecandelob@unal.edu.co*
[c] *Departamento de Ingeniería Eléctrica, Universidad Técnica Federico Santa María, Valparaíso, Chile. johny.montana@usm.cl*

Received: March 5th, 2015. Received in revised form: August 4rd, 2015. Accepted: August 19th, 2015.

Abstract
In this article, we propose an extended fault positions method combined with the Monte Carlo method to evaluate voltage sags. The distribution function *SARFI* is obtained by taking into account the randomness of (i) location of faults in lines, (ii) generation dispatch, and (iii) the prefault voltage. Voltage magnitudes are calculated with power flow, while noting changes in the generation dispatch, the load, and the topology of the area of vulnerability (AOV). The method is tested in the Atlantic coast area of the National Interconnected Power System of Colombia. The distribution of the number of voltage sags per year with the magnitude in bus bars and the impact of generation on the voltage sags are determined. With a higher number of plants dispatched, voltage sags caused by faults are less severe due to the robustness of the power system and the voltage support. Operation with coupled bars had a greater impact on voltage sags compared to uncoupled bars.

Keywords: electromagnetic compatibility; fault position method; Monte Carlo; power quality; voltage sags.

Evaluación de hundimientos de tensión mediante un método extendido de posiciones de falla y simulación de Monte Carlo

Resumen
En este artículo, proponemos un método extendido de posición de falla combinado con el método de Monte Carlo para evaluar los hundimientos de tensión. La función de distribución SARFI se obtiene teniendo en cuenta la aleatoriedad de (i) los puntos de fallas en las líneas (ii) el despacho de generación y (iii) la tensión prefalla. Las magnitudes de tensión se calcularon con flujos de carga, considerando cambios en la generación, la carga y la topológica del área de vulnerabilidad (AOV). El método se probó en el área de la Costa Atlántica perteneciente al Sistema Eléctrico Interconectado Nacional de Colombia. Se calculó la distribución del número de hundimientos de tensión por año, con la magnitud en barras y el impacto de la generación de los hundimientos de tensión. Cuando se despacha un número mayor de plantas, los hundimientos de tensión causados por fallas fueron menos graves, debido a la robustez del sistema de alimentación y el soporte de tensión. La operación con las barras acopladas trae mayor impacto en los hundimientos de tensión comparada con las barras desacopladas.

Palabras clave: compatibilidad electromagnética; método de posición de falla; Monte Carlo; calidad de potencia; hundimientos de tensión.

1. Introduction

Users' requirements for better power quality have increased in the last three decades. One of the reasons is the economic impact of voltage sags in the power grid on customers and end-use equipment manufacturers [1,2].

Several factors affect power quality [1,3–5]: devices powered by electronic converters, speed drivers, and compact fluorescent lamps (CFL). Furthermore, distributed generation and renewable energy sources can create voltage variations, flickers, and harmonic distortion. Similarly, energy efficiency equipment is an important source of disturbance. All these devices are very sensitive to voltage sags because they are manufactured with narrow ranges of

operation for competitive reasons [5].

Many power quality studies have been conducted and reported on previously, including (i) measurement techniques, (ii) evaluation of voltage sags, and (iii) the economic impacts of equipment damage and losses in industrial processes.

In [6], the application of probabilistic methods was presented to predict and characterize how often events appear in the power system in order to assess their impact, demonstrating how a user can be affected.

In [7], the impact of the fault probability distribution model of transmission lines in assessing the number of voltage sags and their characteristics was analyzed.

In [8], the behavior of voltage sags using fault positions and the Monte Carlo method was evaluated. The results showed that the Monte Carlo method provides a better statistical description of voltage sags compared to the fault positions method, which offers only long-term average values, whereas Monte Carlo shows the total distribution function.

In [9], the fault positions method and a Monte Carlo simulation were compared in order to stochastically evaluate voltage sag behavior in a large transmission system. This work showed that the fault positions method cannot be used to predict the behavior of a particular year unless correction factors are used to adjust the behavior. Whereas the fault positions method gives average values, the Monte Carlo method describes the complete frequency distribution function of the voltage sags index (*SARFI*$_X$: System Average RMS Frequency Index; average number of voltage sags per year with magnitude < X%).

In [10], the Monte Carlo method and the fault positions method were applied to evaluate voltage sag indices. This approach assesses the randomness of the prefault conditions and uncertainty in failure rates.

In [11], a method for stochastic prediction of voltage sags generated by faults in the power system was presented. Furthermore, a method for determining the AOV was proposed.

In [12], the influence of generation dispatch and failure rates changing over time on the stochastic prediction of voltage sags was discussed.

In [13], the fault positions method was presented to stochastically predict the frequency and characteristics of balanced and unbalanced voltage sags in distribution systems.

In [14], an analytical method for the stochastic prediction of voltage sags in high-voltage networks was proposed. The method is based on the Z matrix and was applied to the IEEE 24-node reliability test system.

In [15], an evaluation of voltage sags based on the concept of area of severity (AOS) and the impact rankings of the lines and buses was presented. These concepts are useful for creating an efficient plan for mitigating voltage sags and evaluating the relationship between sensitive load points and system voltage sag performance.

In [16], a methodology to estimate the magnitude and frequency of voltage sags originating from faults was presented. The method took into account the application of statistical analysis such as the confidence interval and analysis of variance.

The fault positions and Monte Carlo methods have provided good results in evaluating the impact of generation on voltage sags and *SARFI* indices. Unfortunately, the behavior of generation dispatch is not randomized and the voltage profiles in the faulted bus bar and the bus bar of interest are considered to be constant parameters.

In this paper, an extended fault positions technique and the Monte Carlo method are proposed to evaluate voltage sags. Random faults, changes in generation dispatch, and load variation were taken into account in order to evaluate different operating conditions. The voltage profiles in bus bars were updated continuously using the power flow. Topology changes related to bus bars and transmission lines were also included in the simulation. Although voltage sags can be caused by lightning, disconnection of large loads, etc., this work focuses only on faults in bus bars and the overhead lines of the power system.

2. Theoretical background

2.1. Voltage sags

A voltage sag is "a decrease in r.m.s. voltage or current at the power frequency for durations of 0.5 cycle to 1 min. Typical values are 0.1 to 0.9 p.u." [17]. They are caused by events with large current flows through the network as a result of a fault at any point in the distribution or transmission network, and they affect customers [3,8,9,18].

Events that can cause voltage sags are short circuits, transformer energization, capacitor disconnection, large motors starting, and large load changes in the power system [18]. Consequences include equipment shutdown, process disruption, damage to or malfunction of electronic controllers, slight reduction in output from a capacitor bank, changes in torque and speed of induction motors [18], and others.

Voltage sag is characterized mainly by means of magnitude and duration [1,3,4,19], as shown in Fig. 1. Additionally, voltage sag can be characterized by the frequency of occurrence, phase shifting, the start point in the voltage waveform, and the shape and type of voltage sag.

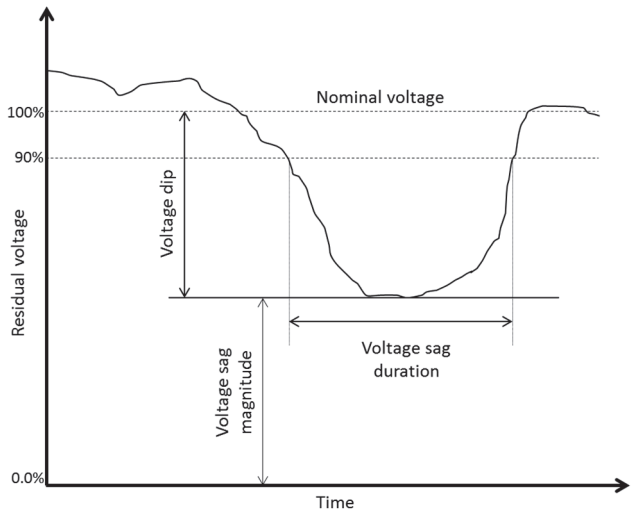


Figure 1. Characteristics of voltage sags.
Source: Adapted from [20]

Standard IEC 61000-4-30 [21] presents a calculation of a sliding reference voltage using a first-order filter for 1 minute. The filter is given by the expression presented in (1).

$$U_{sr(n)} = 0.9967 * U_{sr(n-1)} + 0.0033 * V_{(10/12)rms} \quad (1)$$

where Usr(n) is the actual value of the sliding reference voltage, Usr(n-1) is the previous value of the sliding reference voltage, and V(10/12)rms is the most recent 10/12 cycle r.m.s. value.

Vrms(1/2) is computed from the voltage samples in the time domain, as shown in (2) [2,4,7].

$$V_{rms_{(1/2)}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} v_i^2} \quad (2)$$

where *i* corresponds to each sample, *N* is the total number of samples, and v_i is the voltage value in the time domain. The value is updated every half cycle.

Most measurement devices use the lower value of *Vrms(1/2)* computed each half cycle in the time domain as the voltage sag magnitude [3].

2.2. Fault positions method

The fault positions method is a stochastic method that predicts the expected number of voltage sags in a specific node of the network. In this method, faults in different places on the network are taken into account. Electrical variables for each fault are stored: residual voltage and fault duration on the nodes of interest, as shown in Table 1.

A failure rate is assigned to each fault position. The transmission lines are divided into a specific number of fixed positions. Each position has a failure rate, which is proportional to the longitude of each section. Thus frequency, magnitude, and duration are determined for each fault position, allowing us to calculate the expected number of voltage sags per year [8,9,11,13,14].

Unlike the conventional fault positions method, the fault positions are fixed; in the proposed method in this work, the positions are randomly assigned anywhere in the transmission line for each type of fault.

2.3. Monte Carlo method

The Monte Carlo method is used to simulate the distribution function of the expected values of voltage sags (SARFIx). This method generates the stochastic variables associated with this study, as shown in Fig. 2 [8,9].

Table 1.
Failure report for many points into the network.

Item	Fault position	Failure rate	Sags magnitude	Duration (ms)
1	Bus bar 1	x faults/year	0%	120
2	…	y faults/year	30%	70
.	Bus bar n	…	…	…
.	Line 1	…	…	…
.	…	…	…	…
n	Line n	m faults/year	75%	240

Source: The authors

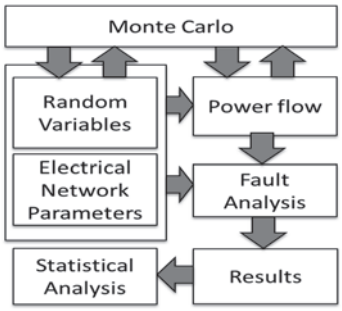


Figure 2. General structure of the Monte Carlo method to evaluate voltage sags.
Source: The authors

The following steps are part of the algorithm to implement the Monte Carlo method [8,9].

1. Select the observation node.
2. Select the years of simulation.
3. For elements in the area of vulnerability, generate a random number to define the fault time according to the probability distribution of the failure rate.
4. Compute the accumulated time.
5. Generate a random number to define a position over the transmission line according to the probability distribution of this parameter.
6. For each fault, generate a random number to define one fault type (SLG, LL, LLG, LLL) according to the probability distribution function of the fault type.
7. Compute the residual voltage as a result of the fault in the observation node.
8. If the accumulated time is less than the simulation time selected, go to step 3. Otherwise go to step 9.
9. Analyze the results statistically.

3. Methodology

This work was done using the fault positions method combined with the Monte Carlo simulation method. Fig. 3 shows the flowchart used to study the voltage sags in a power system.

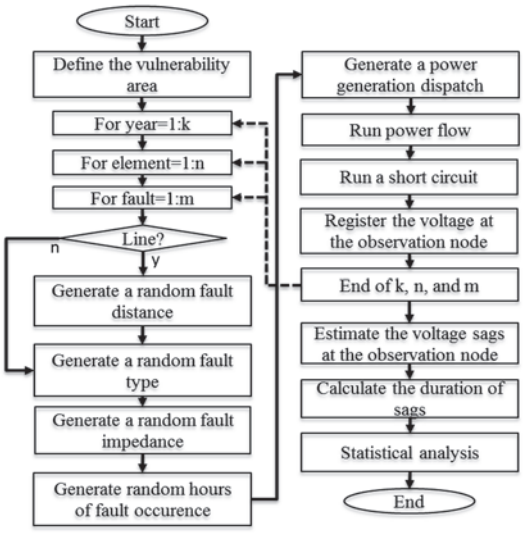


Figure 3. Flowchart of the methodology.
Source: The authors

4. Analytical procedure

4.1. Statistical analysis of results

The great volume of data generated by the Monte Carlo analysis has to be summarized by means of statistical tools. The most important results from Monte Carlo are the long-term mean values and the frequency distribution of mean values [3,8,9,16].

Calculation of the mean values and standard deviation of the voltage sags lower than the predefined magnitudes must be carried out. Their magnitudes are commonly defined between 0.1 and 0.9 in steps of 0.1. It is also important to calculate minimum, maximum, median, and quartiles 1, 2, and 3.

Once the deviation, mean value $\bar{X}$, and standard deviation s are known, the confidence intervals for the expected value can be obtained as expressed in (3) [8,9,16].

$$SARFI_X \in \left[\bar{X} - 2.05.\frac{s}{\sqrt{n}}; \bar{X} + 2.05.\frac{s}{\sqrt{n}}\right] \quad (3)$$

where $\bar{X}$ is the mean value of voltage sags per year, s is the standard deviation, *n* is the number of years of simulation, and the constant *2.05* is the critical value of *t-student* distribution for 95% confidence.

4.2. Area of vulnerability (AOV)

The area of vulnerability was identified by means of power system analysis software that takes fault analysis into account. Faults were simulated at each node of the network, considering the worst cases:

- Three-phase fault solidly grounded (R_f=0)
- Minimum generation dispatch
- Time of maximum demand

Faults were simulated sequentially by voltage level, recording the residual voltage in the node under study and the electrical distance between the node and fault point for each fault. Thus, the maximum distance was obtained for each voltage level for which the residual voltage was lower than 0.9 p.u. (critical distance).

4.3. Simulation time

According to records of meters installed in the node under study, the number of voltage sags per year was between 250 and 340. Thus, given 360 voltage sags per year and an expected error lower than 2% for a confidence of 95%, the minimum number of years to be considered in the simulation was 25.

4.4. Stochastic factors

Five factors were taken into account in the simulations in order to evaluate their impact on the voltage sags in the network.

4.4.1. Resistance

The resistance values were modeled stochastically with normal distribution and standard deviation equal to 1 [22]. The mean value of the resistance was obtained from records of phase-ground faults recorded by distance protection relays.

Historical data was used to obtain at least 20 events of voltage and the current waveform of phase-ground failures registered by a distance relay. Based on these waveforms, the fault current was calculated by

$$I_f = \frac{V_f}{Z_{sp}} = \frac{V_f}{Z_l + Z_{l2} + Z_f} \approx \frac{V_f}{Z_l + Z_f} \quad (4)$$

where

V_f = Fault voltage

I_f = Fault current

Z_{sp} = Impedance from source up to fault point

Z_l =Impedance of transmission line A – B (Fig. 4)

Therefore, Z_f can be obtained from the above equation as expressed in (5):

$$Z_f = \frac{V_f - I_f * Z_l}{I_f} \quad (5)$$

The mean value of fault impedance is the average value of fault impedances estimated by means of equation (5).

4.4.2. Type of fault

The type of fault was obtained randomly from the probability distribution of faults at each voltage level. This probability was computed from the voltage sag data measured at each bus bar. Table 2 shows the probabilities of different faults at each voltage level of the network under study.

For each voltage level, the voltage sags were classified as 1 phase, 2 phases, and 3 phases. The probability of presenting each event was calculated as a percentage from the total events of each voltage level.

4.4.3. Fault position in transmission lines

Many methods in the literature use a constant number of segments for dividing the transmission line; therefore, each segment has a fixed failure rate. In this work, the transmission

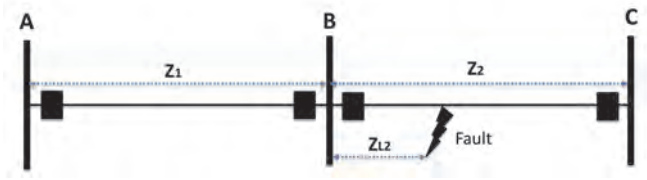


Figure 4. Transmission line representation.
Source: The authors

Table 2.
Probability of faults for each voltage level

Voltage (kV)	Probability by fault type (%)		
	1ϕ	2ϕ	3ϕ
13.8	31.5	46.2	22.3
34.5	25.3	49.0	25.7
110	79.0	15.6	5.4
Total	42.0	39.0	19.0

Source: The authors

line was not divided into segments, but the fault position was computed by means of a random number generated from the probability distribution. This random number was multiplied for the length of the overhead line, and the fault position was obtained.

4.4.4. Time period of the faults

The time of occurrence of a fault was obtained by means of a random number generated between 0 and 23. The number 0 means the time from 00:00 to 01:00, the number 1 means the time from 01:00 to 02:00, and so on. The demand of each node in the network and the prefault voltage were computed from the randomly generated time of occurrence and the load profile at each bus bar by means of a power flow.

4.4.5. Generation dispatch

Generation dispatch was also generated randomly from the probability distribution. In this study, the generation dispatch had two options: (1) all generation plants dispatched and (2) operation without generation plants connected to 110 kV bus bars. The voltage sag analysis was developed in a real power system where four generation scenarios were taken into account:

a. Baseline scenario: 30% of the time, all generators are dispatched. This is typical behavior during the dry season in Colombia (4 months).
b. Scenario 2: 40% of the time, all generators are dispatched. This occurs when the dry season is a little bit longer (5 months).
c. Scenario 3: 60% of the time, all generators are dispatched. Dry season is even longer (7 months).
d. Scenario 4: 70% of the time, all generators are dispatched. This scenario occurs during the "El Niño" phenomenon, when the dry season is even longer (8 months).

4.5. Definition of dynamic variables

The load profile at each bus bar of the network was simulated as a dynamic variable. Additionally, the failure rate of transmission lines and substations of the AOV were considered to be dynamic events.

For the failure rate, it was assumed that the time between the fault at the bus bar and the substation of the power system followed an exponential distribution. The failure rate of elements in the AOV was calculated from the real statistics of the power system under study.

4.6. Validation of the model

Validation of the model was done by comparing the records from 3 years in which the expected average number of voltage sags were found with simulations covering a longer period (25 years) [13].

The expected voltage sags were determined by the confidence intervals from the results of the simulation at the points of interest. To estimate the confidence intervals, the method of percentiles was used regardless of the probability distribution [16].

4.7. Voltage sag duration

Theoretically, the best way to calculate the duration of each voltage sag is to know the clearance time of the fault. Therefore, knowledge of the fault currents, protection relay settings, and breaker times is needed. Obtaining all the information from distribution networks is sometimes a difficult assignment because of the amount of data required to continuously update the databases. Based on [23], in this research a random generation of voltage sag duration was implemented, which took into account the distribution probability of voltage sag durations registered by measurement devices at the bus bars.

5. Power system under study

The Atlantic coast area of the National Interconnected Power System of Colombia was used to test the proposed method. General information about this power system is presented in Table 3.

Fig. 5 shows a single-line diagram of the Atlantic coast area of the National Interconnected Power System of Colombia, which was used for power quality studies.

Table 3.
Number of elements in the power system.

Voltage (kV)	Bus bar	Line	Length (km)
13.8	228	620	29779
34.5	120	121	3 100
66	33	20	290
110	102	51	1 401
220	17	34	1 870
500	10	10	1 870
Total	510	856	38 310

Source: The authors

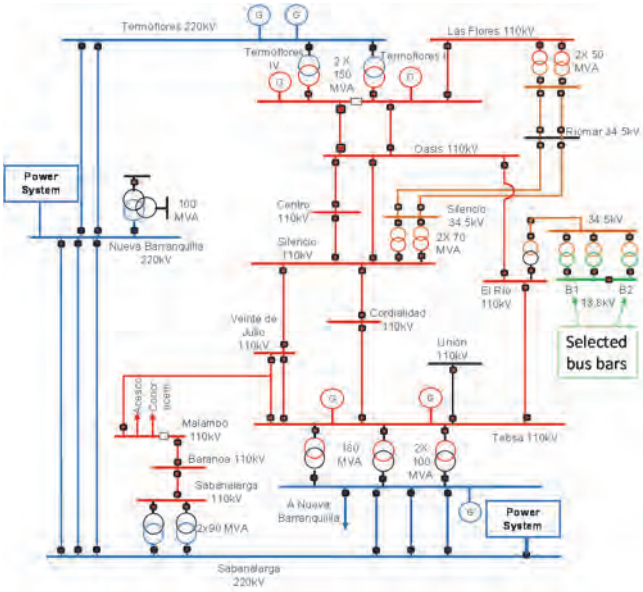


Figure 5. Single-line diagram of the power system test case.
Source: The authors

Table 4.
Number of users in the El Rio substation.

Type of users	Bus bar B1	Bus bar B2	Bus bars B1 and B2
Commercial	1 619	1 209	2 828
Industrial	95	134	229
Government	38	31	69
Residential	6 935	2 298	9 233
Total	8 687	3 672	12 359

Source: The authors

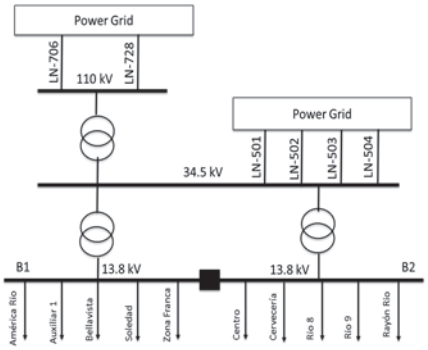


Figure 6. Single-line diagram of El Río substation.
Source: The authors

The aim of the simulation was to characterize the behavior of the power system using the number of voltage sags as a function of the magnitude and duration at a preselected point. The selected points were bus bars B1 and B2 at 13.8 kV of the El Río substation in coupled and uncoupled configurations. These bus bars can be coupled and uncoupled by a circuit breaker.

This study stochastically analyzed the behavior of the voltage sags at a bus bar with users connected. Table 4 shows information about the number of users connected to the bus bar. The simulation considered 8687 users connected to bus bar B1, 3672 users connected to bus bar B2, and 12 359 users connected to the coupled bus bars.

6. Results and analysis

6.1. Area of vulnerability

Fig. 6 shows a single-line diagram of the power system used to carry out the studies. The AOV for the selected bus bars was calculated using three-phase faults at each bus bar to evaluate the voltage magnitudes.

Based on methodology previously described, the AOV was calculated for fault impedance equal to zero with a minimum number of generation plants and maximum demand. Table 5 presents a summary of the critical distances for different voltage levels.

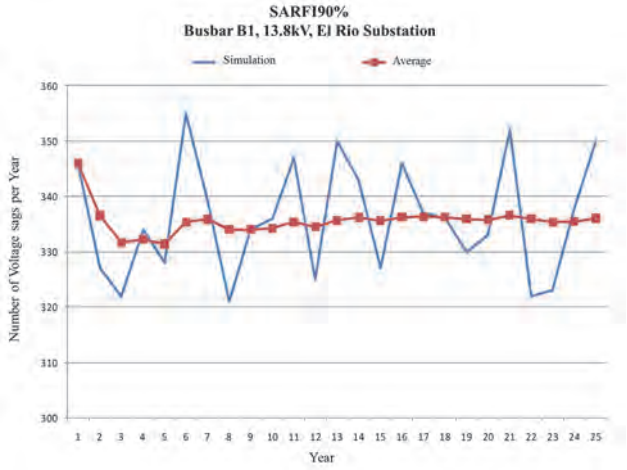
6.2. Power quality indices

Fig. 7 shows the behavior of the Monte Carlo simulation for bus bars B1 and B2 respectively. Fig. 7a and Fig. 7b show the $SARFI_{90\%}$ per year of simulation and the behavior of the voltage sags in bus bars B1 and B2 of the El Rio substation. The index $SARFI_{90\%}$ changes for the first few years but becomes stable after the eighth year.

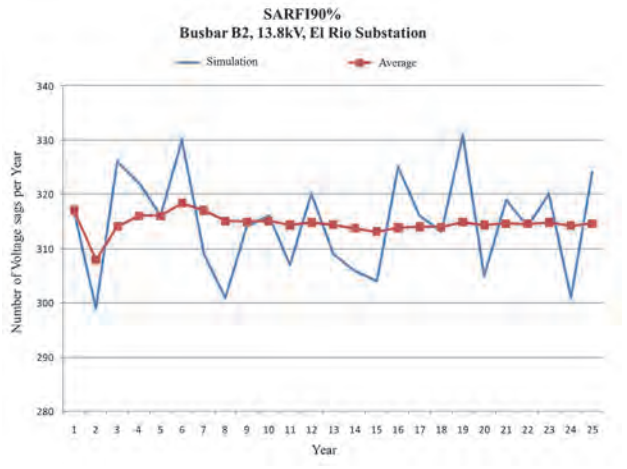
Table 5.
Results of the AOV.

Voltage (kV)	Critical Distance (km)	Substation(s)
13.8	20	Barranquilla
34.5	15	Barranquilla
110	20	Barranquilla
220	280	Atlántico, Bolivar, and Magdalena
500	700	Costa Atlántica, San Carlos, Primavera, and Ocaña

Source: The authors



(a)



(b)

Figure 7. $SARFI_{90\%}$ per year for baseline scenario: (a) bus bar B1, (b) bus bar B2.
Source: The authors

6.3. Distribution of voltage sags

Fig. 8 and Fig. 9 show the frequency distribution of $SARFI_{90\%}$ for bus bars B1 and B2 respectively.

These figures show the differences between frequencies of the *SARFIx* index of bus bars B1 and B2 for the simulated scenarios. Most of the annual voltage sags for bus bar B2 are lower than those for bus bar B1, which helps identify the best configurations for the electrical network.

Fig. 10 shows the expected number of sags per year and the *SARFIx* for ranges between 0% and 90% for bus bars B1 and B2 of the El Rio substation.

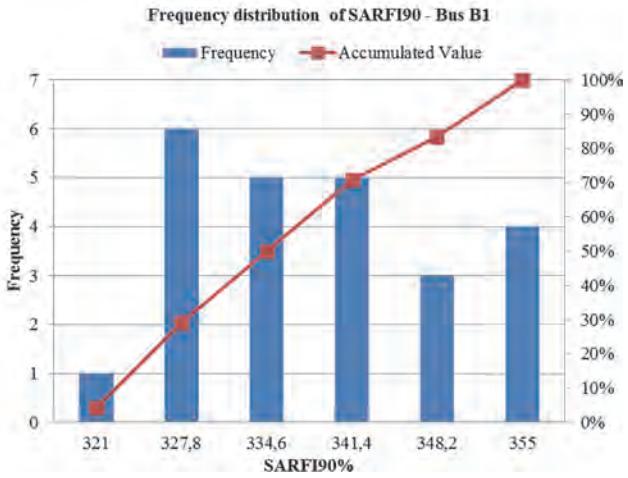


Figure 8. Frequency distribution of $SARFI_{90\%}$ for bus bar B1.
Source: The authors.

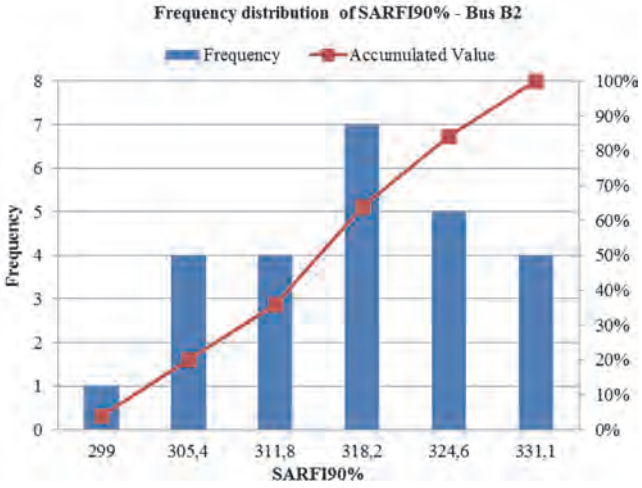


Figure 9. Frequency distribution of $SARFI_{90\%}$ for bus bar B2.
Source: The authors.

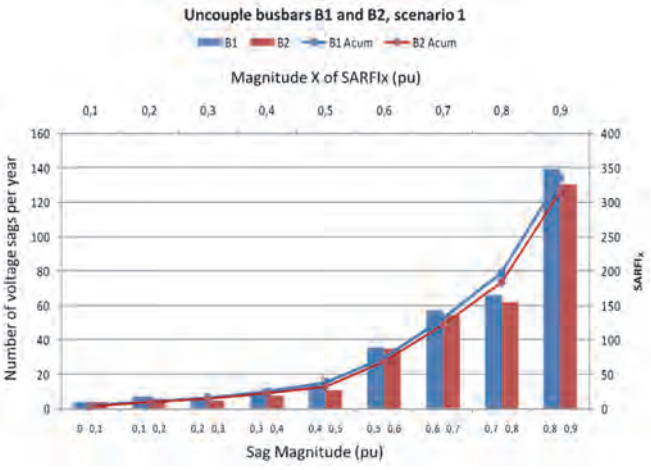


Figure 10. Number of sags and *SARFIx* vs. Sag magnitude.
Source: The authors.

6.4. *Results of the model*

Table 6 shows the number of actual voltage sags versus the number of voltage sags simulated for bus bars B1 and B2. The measurements of the actual voltage sags were carried out for the years 2010, 2011, and 2012. The results of the Monte Carlo simulation were obtained according to the four generation dispatch scenarios defined in the methodology.

In the long term, the expected number of voltage sags with magnitude ≤ 90% ($SARFI_{90\%}$) was lower for bus bar B2 than for bus bar B1 because the expected value for $SARFI_{90\%}$ in bus bar B2 was less than the corresponding percentile *P2.5%* for all scenarios.

Comparing the results of the Monte Carlo simulation for both scenarios—coupled and uncoupled bus bars—and utilizing the same failure rate of the elements in the AOV, the results show that the coupled bus bars generated a higher number of voltage sags than the uncoupled bus bars. If the number of generation plants were to increase, the problem would become less detectable.

With the exception of bus bar B2 for the year 2010, the expected values of voltage sags for B1 and B2 were contained in the corresponding confidence intervals for the two scenarios considered, compared to measurements.

6.5. *Voltage sag duration*

The probability distribution of voltage sag duration is shown in Fig. 11, calculated for 13.8 kV and 110 kV. These results were obtained by combining the magnitude and duration of the voltage sags.

Table 7 shows the accumulated voltage sags for the coupled and uncoupled bus bars B1 and B2. The voltage sag magnitudes with the coupled bus bars B1 and B2 are similar to the voltage sag magnitudes obtained with the uncoupled bus bars. The number of voltage sags is greater with the coupled bus bars.

Table 8 shows the accumulated voltage sags for different percentages of SARFI. The information is presented for the coupled and uncoupled bus bars B1 and B2.

For all *SARFI* studied, the number of accumulated voltage sags was greater for bus bar B1 than for bus bar B2. Furthermore, the coupled bus bars presented a greater number of voltage sags for all scenarios studied.

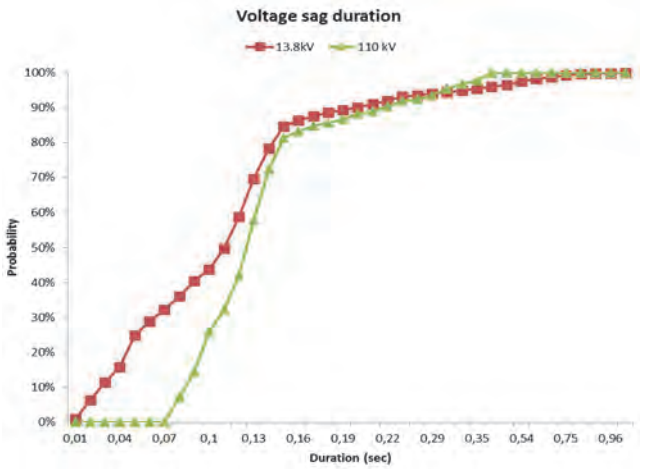


Figure 11. Accumulative distribution function of voltage sag duration.
Source: The authors.

Table 6.
Actual vs. simulated voltage sags for bus bars B1 and B2.

Bus bar	Monitoring			Confidence interval of 95%											
	2010	2011	2012	Scenario 1 (Gmax 30%)			Scenario 2 (Gmax 40%)			Scenario 3 (Gmax 60%)			Scenario 4 (Gmax 70%)		
				$P_{2.5\%}$	$P_{97.5\%}$	VE	$P_{2.5\%}$	$P_{97.5\%}$	VE	$P_{2.5\%}$	$P_{97.5\%}$	VE	$P_{2.5\%}$	$P_{97.5\%}$	VE
B1	326	313	293	321.6	353.2	336.0	306.2	342.4	323.5	282.2	314.0	298.0	272.0	302.4	286.8
B2	341	315	315	300.2	330.4	314.6	284.6	317.6	301.9	265.6	291.0	277.7	253.6	282.0	266.9
B1B2	-	-	-	366.2	397.4	379.6	342.0	373.2	356.5	296.6	334.0	314.5	276.6	314.0	295.1

Source: The authors

Table 7.
Voltage sag density according to the magnitude and the duration.

Magn. (pu)	Bus bar	Duration in milliseconds													
		0	50	100	150	200	250	300	350	400	500	600	700	800	900
0.9	B1	336.04	305.92	231.96	65.12	40.16	28.64	20.64	12.36	6.36	4.44	2.72	1.68	0.80	0.40
	B2	314.56	287.80	216.48	61.36	38.44	27.04	19.88	11.24	6.28	4.52	3.12	1.76	0.80	0.36
	B1B2	379.56	339.44	254.80	69.80	44.08	30.56	21.72	12.80	7.32	5.40	3.28	1.80	1.00	0.48
0.8	B1	196.72	185.04	140.60	39.04	23.76	16.16	11.44	6.36	2.24	1.48	0.88	0.48	0.20	0.08
	B2	184.12	175.00	131.80	36.76	22.36	15.00	11.08	5.16	2.20	1.44	0.92	0.40	0.28	0.12
	B1B2	209.80	195.80	145.40	39.88	24.56	16.60	11.60	5.88	2.72	2.00	1.48	0.84	0.48	0.36
0.7	B1	130.80	124.68	94.28	26.72	16.44	11.44	8.20	4.60	1.44	1.04	0.64	0.28	0.12	0.04
	B2	121.88	117.64	87.32	23.56	14.32	9.36	6.88	2.96	0.92	0.60	0.40	0.20	0.08	0.04
	B1B2	143.00	135.92	100.36	26.72	16.72	10.84	7.56	3.52	1.28	0.92	0.68	0.48	0.28	0.24
0.6	B1	73.52	70.00	53.80	15.16	9.08	6.36	4.28	2.44	0.68	0.48	0.32	0.12	0.04	0.04
	B2	67.52	65.04	47.60	13.12	7.84	5.08	3.84	1.76	0.52	0.36	0.24	0.04	0.04	0.04
	B1B2	88.68	84.28	62.36	17.20	10.92	7.20	5.04	2.04	0.76	0.52	0.36	0.24	0.16	0.16
0.5	B1	37.84	35.88	26.84	7.36	4.00	2.96	2.28	1.32	0.44	0.32	0.24	0.12	0.04	0.04
	B2	32.60	31.40	22.12	6.48	3.76	2.44	1.76	1.04	0.40	0.28	0.20	0.04	0.04	0.04
	B1B2	43.64	40.96	31.48	8.32	5.00	3.44	2.24	1.08	0.44	0.36	0.24	0.16	0.08	0.08
0.4	B1	24.40	23.32	17.24	4.76	2.68	2.04	1.64	1.08	0.36	0.24	0.20	0.08	0.04	0.04
	B2	22.04	21.24	14.80	4.28	2.44	1.72	1.28	0.76	0.36	0.28	0.20	0.04	0.04	0.04
	B1B2	25.56	23.84	18.00	4.64	2.68	1.96	1.28	0.60	0.24	0.20	0.12	0.08	0.04	0.04
0.3	B1	15.92	15.28	11.08	3.20	1.80	1.24	1.00	0.64	0.28	0.20	0.16	0.04	0.04	0.04
	B2	14.44	13.84	9.92	2.72	1.72	1.28	0.96	0.60	0.36	0.28	0.20	0.04	0.04	0.04
	B1B2	19.48	18.36	13.76	3.28	1.80	1.32	0.92	0.40	0.20	0.16	0.08	0.04	0.04	0.04
0.2	B1	10.96	10.44	7.40	2.32	1.28	0.92	0.72	0.44	0.24	0.20	0.16	0.04	0.04	0.04
	B2	9.60	9.20	6.64	1.84	1.16	0.80	0.64	0.40	0.32	0.24	0.16	0.04	0.04	0.04
	B1B2	13.00	12.20	9.44	2.36	1.28	1.04	0.84	0.36	0.20	0.16	0.08	0.04	0.04	0.04
0.1	B1	4.16	3.84	2.52	0.96	0.60	0.44	0.32	0.16	0.12	0.12	0.08	0.04	0.04	0.04
	B2	3.88	3.56	2.60	0.96	0.56	0.48	0.48	0.36	0.28	0.24	0.16	0.04	0.04	0.04
	B1B2	5.84	5.40	4.28	0.92	0.60	0.52	0.44	0.16	0.08	0.08	0.04	0.04	0.04	0.04

Source: The authors

Table 8.
Voltage sags evaluated for bus bars B1 and B2 with different *SARFIx*.

SARFI	**Bus bars**	**Scenarios**			
		1	**2**	**3**	**4**
90%	B1	336.04	323.52	298.00	286.76
	B2	314.56	301.92	277.68	266.92
	B1-B2	379.56	356.48	314.48	295.08
80%	B1	196.72	190.80	179.20	173.48
	B2	184.12	178.72	166.96	162.60
	B1-B2	209.80	200.24	182.60	174.68
70%	B1	130.80	124.96	112.84	107.12
	B2	121.88	116.52	104.76	100.64
	B1-B2	143.00	133.84	115.72	108.48
60%	B1	73.52	69.60	60.64	56.32
	B2	67.52	64.32	55.52	52.48
	B1-B2	88.68	82.00	67.52	60.36
50%	B1	37.84	36.00	32.24	30.76
	B2	32.60	31.12	29.24	28.04
	B1-B2	43.64	41.24	34.60	31.24

Source: The authors

7. Conclusions

This paper presented an extended fault positions method combined with the Monte Carlo method to evaluate the impact of voltage sags in a power system. Random faults in transmission lines, variation of generation dispatch, and variations of load were taken into account for the simulations. Statistical tests were conducted, finding that the greater the number of generation plants dispatched in the AOV, the lower the magnitude of the voltage sags and the index *SARFI*. The proposed method evaluated the reconfiguration of bus bars to reduce the number of voltage sags, finding that coupled bars have a greater impact on voltage sags compared to the results obtained with uncoupled bars. The results showed that voltage sags have a significant impact on power system operation. The parameters and the method utilized in this research should be included in any future power quality analysis.

Acknowledgements

This research was supported in part by the Universidad del Norte – Colombia, the Universidad Nacional de Colombia – Sede Medellín – Colombia, and the Universidad Técnica Federico Santa María – Chile. The authors wish to thank the company ELECTRICARIBE for the valuable information provided for this research.

References

[1] Dugan, R.C., McGranaghan, M.F., Santoso, S. and Beaty, H.W., Electrical Power Systems Quality. Third. McGraw-Hill Education, 2012.

[2] Gil-Montoya, F., Manzano-Agugliaro, F., Gómez-López, J. and Sánchez-Alguacil, P., Técnicas de investigación en calidad eléctrica: ventajas e inconvenientes. DYNA, 79(173), pp. 6-74, 2012.

[3] Bollen, M., Understanding power quality problems: Voltage sags and interruptions. New Jersey: Wiley-IEEE Press, 1999. DOI: 10.1109/9780470546840

[4] Bollen, M. and Gu, I., Signal processing of power quality disturbances. New Jersey: Wiley-IEEE Press, 2006. DOI: 10.1002/9780471931317

[5] Giménez-Álvarez, J. y Gómez-Targarona, J., Generación eólica empleando distintos tipos de generadores considerando su impacto en el sistema de potencia. DYNA, 78(169), pp. 95-104, 2011.

[6] Sikes, D.L., Comparison between power quality monitoring results and predicted stochastic assessment voltage sags-"real" reliability for the customer. IEEE Transactions on Industry Applications, 36(2), pp. 677-82, 2000. DOI: 10.1109/28.833787

[7] Milanovic, J.V., Aung, M.T. and Gupta, C.P., The influence of fault distribution on stochastic prediction of voltage sags. IEEE Transactions on Power Delivery, 20(1), pp. 278-285, 2005. DOI: 10.1109/TPWRD.2004.835052

[8] Olguin, G., Aedo, M., Arias, M. and Ortiz, A., A Monte Carlo simulation approach to the method of fault positions for stochastic assessment of voltage dips (sags), IEEE/PES Transmission and Distribution Conference and Exhibition: Asia and Pacific, 2005. pp. 1-6. DOI: 10.1109/TDC.2005.1546794

[9] Olguin, G., Karlsson, D. and Leborgne, R., Stochastic assessment of voltage dips (Sags): The method of fault positions versus a Monte Carlo simulation approach, IEEE Russia Power Tech, 2005. pp. 1-7. DOI: 10.1109/TDC.2005.1546794

[10] Caramia, P., Carpinelli, G., Di Perna, C., Varilone, P. and Verde, P., Fast probabilistic assessment of voltage dips in power systems, 9th International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), 2006. pp. 1-6. DOI: 10.1109/PMAPS.2006.360405

[11] Park, C.-H. and Jang, G., Stochastic estimation of voltage sags in a large meshed network. IEEE Transactions on Power Delivery, 22(3), pp. 1655-1664, 2007. DOI: 10.1109/TPWRD.2006.886795

[12] Park, C.-H., Jang, G. and Thomas, R.J., The influence of generator scheduling and time-varying fault rates on voltage sag prediction. IEEE Transactions on Power Delivery, 23(2), pp. 1243-1250, 2008. DOI: 10.1109/TPWRD.2008.915836

[13] Goswami, A.K., Gupta, C.P. and Singh, G.K., The method of fault position for assessment of voltage sags in distribution systems, IEEE Region 10 and the Third international Conference on Industrial and Information Systems (ICIIS), 2008. pp. 1-6. DOI: 10.1109/ICIINFS.2008.4798362

[14] Quaia, S. and Tosato, F., A method for analytical voltage sags prediction, IEEE Bologna Power Tech Conference Proceedings, 4, 2003. pp. 181-186. DOI: 10.1109/PTC.2003.1304720

[15] Park, C.-H., Hong, J.-H. and Jang, G., Assessment of system voltage sag performance based on the concept of area of severity. IET Generation, Transmission and Distribution, 4(6), pp. 683-693, 2010. DOI: 10.1049/iet-gtd.2009.0492

[16] De Oliveira, T., Carvalho-Filho, J.M., Abreu, J.P. and Chouhy-Leborgne, R., Voltage sags: Statistical evaluation of monitoring results based on predicted stochastic simulation. 12th International Conference on Harmonics and Quality of Power, 2006.

[17] IEEE Std. 1346. Recommended practice for evaluating electric power system compatibility with electronic process equipment. 1998.

[18] Jaramillo-Matta, A.A., Guasch-Pesquer, L. and Trujillo-Rodriguez, C.L., Classification of voltage sags according to the severity of the effects on the induction motor. DYNA, 82(190), pp. 96-104, 2015. DOI: 10.15446/dyna.v82n190.43286

[19] IEC Std. 61000-2-8 Voltage dips and short interruptions on public electric power supply systems with statistical measurement results. 2002.

[20] IEEE Std. 1159. IEEE recommended practice for monitoring electric power quality. 1995.

[21] IEC Std. 61000-4-30, Testing and measurement techniques—Power quality measurement methods. 2003.

[22] Avendano-Mora, M., Milanović, J.V., Patel, B. and Zhang, Y., The influence of model parameters and uncertainties on assessment of network wide costs of voltage sags, 10th International Conference on Electrical Power Quality and Utilisation (EPQU), 2009. pp. 1-7. DOI: 10.1109/EPQU.2009.5318857

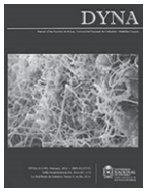
[23] Wämundsson, M., Calculating voltage dips in power systems, using probability distributions of dip durations and implementation of the moving fault node method, MSc. Thesis, Department of Energy and Environment, Chalmers University of Technology, Göteborg, Sweden, 2007.

J. Sagre, received his BSc. in Electrical Engineering in 1978 from the Universidad Nacional de Colombia, Bogotá, Colombia and a MSc. in Electrical Engineering in 2014 from Universidad del Norte, Barranquilla, Colombia. His employment experiences include: the Corporación Eléctrica de la Costa Atlántica – CORELCA -, TRANSELCA S.A. E.S.P. and ELECTRICARIBE S.A. E.S.P. He is now a Consultant Engineer. His research interests include: planning, operation and control of power systems, power quality and regulation.
ORCID: 0000-0001-8036-0644

J. Candelo, received his BSc. in Electrical Engineering in 2002 and his PhD in Engineering with emphasis in Electrical Engineering in 2009 from Universidad del Valle, Cali, Colombia. His employment experiences include the Empresa de Energía del Pacífico EPSA, Universidad del Norte, and Universidad Nacional de Colombia, Medellín, Colombia. He is now an assistant professor of the Universidad Nacional de Colombia, Medellín, Colombia. His research interests include: planning, operation and control of power systems, artificial intelligence and smart grids.
ORCID: 0000-0002-9784-9494

J. Montaña, received his BSc. in Electrical Engineer, MSc. in High Voltage, and his PhD. in 1999, 2002, and 2006, all of them from Universidad Nacional de Colombia,. He is a full-time professor with the Department of Electrical Engineering, Universidad Federico Santa María, Valparaiso, Chile. He worked as a research assistant with Universidad Nacional de Colombia from 2000 to 2005, as junior design engineer with Siemens S.A. Colombia, from 2006 to 2009, and a fulltime professor with Universidad del Norte, Barranquilla, Colombia, from 2010 to 2013. His research interests include lightning protection systems, lightning location systems, and grounding systems.
ORCID: 0000-0002-9999-2366

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS



DYNA
http://dyna.medellin.unal.edu.co/

Conversion of an existing electrostatic precipitator casing to Pulse Jet Fabric filter in fossil power plants

Francisco Manzano-Agugliaro [a] & Javier Carrillo-Valle [a]

[a] *Departamento de Ingeniería. Universidad de Almería. Almería, España. fmanzano@ual.es; javier.carrillo@endesa.es*

Received: March 19th, 2015. Received in revised form: October 26th, 2015. Accepted: November 3rd, 2015

Abstract
The combustion process of power generation plants originates particulates. There are different technologies to collecting particulate such as electrostatic precipitators (ESPs) or fabric filters. Currently, these ESPs take 25 or 35 years in service and if the performance expectations of their Plants are positives, improving investments required which can adapt to the new particulate emission limits becoming more stringent. This paper analyzes an alternative means great savings in investment costs; Conversion of the existing ESP casing to a Pulse Jet fabric filter. This study also presents a real case, implementing this conversion with good results in unit of 660 MW power plants of Italy.

Keywords: particulate, opacimeter, electrostatic precipitator, fabric filter, pulse-jet

Transformación de un precipitador electrostático a un filtro de mangas Pulse Jet en centrales térmicas convencionales

Resumen
El proceso de combustión de las Centrales Térmicas convencionales origina emisiones de partículas. Existen distintas tecnologías de captación de partículas, como los precipitadores electrostáticos (PE) o los filtros de mangas. Hoy en día, estos precipitadores electrostáticos llevan 25 o 35 años en servicio y si las expectativas de funcionamiento de sus Centrales son positivas, se requieren inversiones de mejora que permitan adaptarse a los nuevos límites de emisión de partículas cada vez más estrictos. El presente artículo analiza una alternativa que supone grandes ahorros en los costes de inversión, transformando el precipitador electrostático en un filtro de mangas con sistema de limpieza por impulsos de aire a presión (pulse-jet), aprovechando la estructura existente. Este estudio presenta también un caso real, poniendo en práctica esta conversión con buenos resultados en una unidad de 660 MW de una central térmica de Italia.

Palabras clave: partícula, opacímetro, precipitador electrostático, filtro de mangas, pulse-jet

1. Introducción

El suministro de energía de fuentes diversificadas y seguras, de forma económicamente admisible y medioambientalmente compatible resulta esencial para el desarrollo sostenible de la sociedad [1-3]. En este escenario es imprescindible la participación del carbón como fuente de energía primaria, para poder responder de forma efectiva a las elevadas demandas energéticas de las economías maduras y a las demandas crecientes de los países en desarrollo [4,5]. En España, el parque de generación eléctrica que emplea carbón representa un 11% de la potencia instalada, y a la hora de dar cobertura a la demanda de energía eléctrica en el año 2014, esta tecnología representó el 19% de la producción total anual [6], contribuyendo así a la imprescindible diversificación, garantía de suministros y estabilidad de precios.

Actualmente existen en España 20 centrales térmicas de producción de energía eléctrica a carbón con una potencia instalada de 11.700 MW [7]. Un hecho destacable es que la mayoría de estas centrales emplean como sistema de captación de partículas el precipitador electrostático, acumulando por tanto más de 30 años de vida en funcionamiento, por lo que se requiere eficientes inversiones que mejoren su rendimiento.

El precipitador electrostático está formado por una estructura soporte externa dividida en una serie de campos o secciones que recibe los gases de combustión a un temperatura de 100°-220°C y utiliza la fuerza electrostática para separar las partículas de los gases de combustión que las arrastran [8]. Los gases de combustión se hacen pasar por una cámara que contiene placas de acero (electrodos colectores) colocadas verticalmente, en la dirección paralela al flujo de los gases, formando entre ellas una serie de pasillos. En cada pasillo existen un conjunto de alambres verticales (electrodos de descarga\emisores) situados en un plano paralelo y equidistantes de las placas que están soportados por una estructura única [9].

Esta estructura se apoya en aisladores para que quede aislada eléctricamente del resto de los componentes, los cuales están conectados a tierra. La estructura aislada que soporta todos los electrodos de descarga está alimentada con una tensión continua negativa, que puede alcanzar valores de 45/70kV, desde un conjunto transformador-rectificador [10]. Como los electrodos colectores o placas están conectadas a tierra, se crea un campo eléctrico intenso entre los electrodos de descarga y las placas.

El gas está formado normalmente por moléculas neutras y su movimiento no se ve afectado por el campo eléctrico, sin embargo, siempre existe alguna molécula ionizada, ya sea por radiación electromagnética o como consecuencia de elevadas temperaturas [9]. Estas moléculas ionizadas (iones gaseosos) y los electrones separados de ellas se mueven hacia los electrodos de polaridad opuesta, donde se descargan. Es decir, se crea una corriente eléctrica, pero ésta es tan débil que casi no puede ser detectada por los aparatos de medida. Si se aumenta el voltaje aplicado a los electrodos, la intensidad del campo en las proximidades del emisor aumenta hasta el punto que los iones y electrones son acelerados pudiendo provocar la ionización de otras partículas neutras por impactos con ellas, y si se repite con una determinada frecuencia de forma que se generen una gran cantidad de iones y electrones libres en la zona próxima al electrodo emisor, se produce una descarga eléctrica, denominada efecto corona [11]. El gas se ioniza en esta zona y se forma una gran cantidad de iones positivos y negativos. Los iones positivos son atraídos y atrapados inmediatamente por los electrodos de descarga, cargados negativamente. Los iones negativos, sin embargo, tienen que atravesar todo el espacio que hay entre los electrodos de descarga y las placas (electrodos colectores, polo positivo). Por lo tanto hay un flujo de iones negativos desde los electrodos de descarga. En el camino hacia las placas, los iones negativos chocan con las partículas de polvo y se adhieren a ellas [12]. Estas partículas, por tanto, quedan cargadas eléctricamente y comienzan a moverse hacia las placas en la misma dirección que los iones negativos. La fuerza eléctrica que actúa sobre cada partícula es mucho mayor que la fuerza gravitatoria y por tanto su "velocidad de migración" hacia las placa es mucho mayor que la "velocidad de sedimentación" y, si se diseña adecuadamente el precipitador, también será adecuada para competir con la velocidad de arrastre de los gases. El polvo se adhiere a las placas colectoras y mediante golpeo periódico, se hace que la capa de polvo depositada sobre ellas, se desprenda y se deslice hacia una tolva situada en la parte inferior [13].

2. Motivos para de reemplazar o convertir el PE existente en un Filtro de Mangas

Existen varios factores que inciden de manera significativa sobre el rendimiento de los PE, y que justifican la opción de reemplazarlos o convertirlos en un Filtro de Mangas, y que se detallan a continuación.

2.1. Envejecimiento del equipo

EL PE instalado en una central térmica típica de carbón se espera que permanezca en funcionamiento durante varias décadas, sobre todo teniendo en cuenta que la vida útil media de una central son 30 años [13]. A lo largo de los años, es normal realizar intervenciones de mantenimiento programado para remplazar componentes internos, tales como martillos del sistema de golpeo, electrodos de descarga, etc. Sin embargo, llega un momento en el que la mayoría de las placas colectoras necesitan ser sustituidas e incluso los transformadores y/o el sistema de control, con lo que la inversión se hace tan elevada que llegamos al punto en el que la conversión a otra alternativa es viable [14].

2.2. Límites de emisión más restrictivos

El consumo de combustibles fósiles se traduce en un aumento de la contaminación ambiental debido a la emisión de CO_2 y otros gases que causan el calentamiento global a través de lo que se conoce como el efecto invernadero [15,16]. En relación a los limites de emisión de partículas, hasta el pasado día 6/01/2011 ha estado en vigor la Directiva Europea 2001/80/CE de Grandes Instalaciones de Combustión (GIC's) de fecha 23/10/2001, la cual fijaba el límite en 50 mg/Nm^3, a las instalaciones con una potencia superior a 50 MW térmicos [17]. Con fecha 17/12/2010 se publicó la nueva Directiva 2010/75/UE de Emisiones Industriales (IED, en inglés), la cual entró en vigor el día 6/01/2011. Esta nueva directiva modificará los términos de emisiones permitidos por la Directiva 2001/80/CE a partir del 01/01/2016. Esta normativa, impone unos valores diferentes en función de la fecha de concesión de permiso y potencia térmica nominal total. Para las instalaciones de combustión a las que se haya concedido permiso antes del 7 de enero de 2013, o para las que sus titulares hayan presentado una solicitud de permiso completa antes de dicha fecha, a condición de que dichas instalaciones hayan entrado en funcionamiento a más tardar el 7 de enero de 2014, los valores límites de emisión de partículas, para una potencia térmica nominal superior a 300 MW serán 20 mg/Nm^3. Para las instalaciones no incluidas en las indicadas anteriormente, los valores límites de emisión, con carácter general, para una potencia térmica nominal superior a 300 MW serán 10 mg/Nm^3 [18].

$$CaCO_3 + SO_2 + 2H_2O + ½\ O_2 \rightarrow CaSO_4 \cdot 2H_2O + CO_2 \quad (1)$$

(caliza) + (dióxido de azufre) + (agua) + (oxigeno) → (yeso) + (dióxido de carbono)

2.3. Cambio del tipo de carbón utilizado

Las Centrales térmicas en muchas ocasiones se ven obligadas a cambiar el tipo de carbón que están quemando, por diversos motivos:

- Deterioro de los parámetros de calidad del carbón suministrado por razones geológicas, o simplemente por el cierre de las minas de origen.
- Por estrategias de mercado, de forma que les permitan disminuir los costes de producción [19].
- El fuerte crecimiento de China e India en el mercado del carbón [20].

Estos nuevos carbones pueden penalizar el rendimiento del PE. Normalmente esto ocurre debido a un aumento del porcentaje de cenizas, a un aumento de la resistividad de las cenizas o un aumento de la fineza de las partículas, las cuáles son más difíciles de captar.

2.4. Instalación de Desulfuraciones de los Gases de Combustión (DGCs)

Para el control de las emisiones de los SOx, las grandes instalaciones cuentan con desulfuradoras de los gases de combustión (DGCs), y en la mayoría de los casos son húmedas [21]. Estas desulfuraciones húmedas están normalmente ubicadas aguas a bajo de los PE, y por consiguiente se ven afectadas por la eficacia de éste. Las reacciones químicas que se crean en un sistema DGC con caliza húmeda están caracterizadas por una serie de fases, aunque dichas fases se pueden simplificar con la siguiente reacción global [22], (eq. 1):

El objetivo de las plantas es crear yeso de carácter comercial, cumpliendo con los criterios de calidad de la industria del cemento, para poder venderlo [23]. Si el rendimiento de captación del PE no es adecuado, existirá una gran cantidad de cenizas en el tanque de reacción del absorbedor de la DGC, donde se mezclan con la lechada de caliza existente, por lo que afectan a los parámetros de calidad del yeso (blancura, pureza>80%, humedad libre<10%) [22] que pueden impedir su venta como subproducto, generando problemas de almacenamiento y/o costes de esmaltimento a las centrales.

2.5. Emisiones de metales pesados: Mercurio

Durante la combustión, los metales pesados que contienen la mayoría de los combustibles sólidos son liberados y algunos de ellos pueden llegar a la atmósfera integrados en partículas o en forma de vapor, particularmente el mercurio (Hg) por su elevada volatilidad [19,25].

En relación con el mercurio u otros metales condensables, existe una preocupación creciente, principalmente en Estados Unidos [26], sobre sus potenciales efectos adversos sobre la salud [27], por lo que la Administración norteamericana ha comenzado a legislar para restringir el uso del mercurio en los procesos industriales y reducir los niveles de concentración emitidos a la atmósfera. En Europa se ha comenzado a tomar conciencia de la problemática [28], y también es presumible la aparición de limitaciones en las emisiones de mercurio, sin embargo los PE muestran una caída importante de rendimiento en el entorno de la fracción de partículas de 1 μm, como en el caso de partículas enriquecidas con metales pesados. Para muchas centrales de carbón, la mejor opción actualmente es instalar filtros de mangas con los que se puede conseguir una eficiencia media aproximada de eliminación del Hg del 40% [29]. Esta eficiencia de captura se puede aumentar acondicionando el gas de combustión con pequeñas cantidades de aditivos que se añade al flujo de gas antes de que atraviese el filtro de mangas, alcanzando así valores de eliminación del Hg > 90% [30].

2.6. Control de PM2,5

Las previsiones para un futuro no muy lejano pasan por la introducción de limitaciones no sólo en partículas totales PM_{10} como actualmente, sino en las fracciones más finas. Las partículas más pequeñas de 2,5 micras $PM_{2,5}$, presentan un mayor riesgo para la salud [31], además de permanecer mayor tiempo suspendidas en el aire y transportadas a largas distancias [32,33]. A parte de esto, es sobre la superficie de las partículas más finas donde suelen depositarse algunos compuestos condensables como los gases ácidos y los metales, lo que incrementa su riesgo potencial [34].

Las mejoras de rendimiento necesarias en los internos de los PE existentes, para alcanzar los límites de emisión $PM_{2,5}$ podrían ser tan importantes que la conversión a filtro de mangas seria una opción viable. Debido a las consideraciones actuales y futuras expuestas anteriormente, la transformación de un precipitador electrostático en un filtro de mangas, es un proyecto que ofrece atractivas ventajas y se adapta a las exigencias medioambientales cada vez más estrictas.

Bajo este contexto se ha elaborado el presente artículo, que tiene como objetivo establecer los aspectos fundamentales de diseño e instalación a tener en cuenta a la hora de realizar una completa especificación para transformar un precipitador electrostático en un filtro de mangas, reutilizando la estructura existente y aplicando un sistema de limpieza por impulsos de aire a presión (pulse-jet). Finalmente se ilustra su aplicación, analizando un caso real en una unidad de 660 MW de una Central Térmica de Italia.

3. Propuesta tecnológica

Los filtros de mangas consisten en estructuras metálicas cerradas en cuyo interior se disponen elementos filtrantes textiles en posición vertical. Según el diseño pueden adoptar formas tubulares, y se denominan mangas, o formas rectangulares, y se denominan bolsas. Se montan sobre una cámara o compartimento que acaba en su parte inferior en una tolva de recogida de partículas. Los gases cargados de partículas entran en cada compartimento a través de un difusor, donde pierde velocidad y turbulencia, para distribuirse en toda la longitud del mismo, dejando las partículas más pesadas que se desprenden por gravedad en las tolvas. A continuación, el gas asciende y es forzado a pasar a través de las filas de mangas, atravesándolas del exterior hacia el interior, depositando las partículas sobre la superficie de las mismas. La filtración se produce como resultado de la formación de una capa de polvo primaria en la superficie de

las mangas (llamada torta) y de una acumulación de partículas de polvo en el interior del material filtrante, que dependen de la concentración de cenizas en los gases, del estado de las mangas y del caudal de gases. Una vez formada la capa primaria, la penetración se hace muy baja y la filtración se produce por tamizado (filtración superficial). El proceso de filtración continúa hasta que la caída de presión se hace tan importante que requiere la limpieza del sistema [35].

Cada manga es instalada en una estructura (jaula portamangas) para que éstas no colapsen por depresión y permanezcan abiertas durante el proceso de filtrado. Finalmente, todas las jaulas portamangas con sus respectivas mangas, son soportadas verticalmente desde su parte superior en la placa portamangas. Esta placa es el límite de separación entre el gas sucio y el gas limpio, por lo que el único camino del gas para pasar a la zona de gas limpio (walk in plenum), es a través de las mangas. Por tanto, cuando se trata de placa portamangas instalada en un filtro de mangas con reutilizo de la estructura existente del PE es necesario una especial consideración ya que se pueden presentar dilataciones térmicas diferenciales entre los compartimentos y la placa, si son de grandes dimensiones o son de diferentes materiales. Para impedir cualquier fuga de gas sucio a través de la placa y las mangas, los fabricantes de las mismas proporcionan datos para un buen "sellado", aplicando el sistema "snap-band" (pellizco o pliego normal), y su diseño permite la fácil sustitución de las mangas [36]. La selección de las características del tejido filtrante de las mangas es el parámetro más importante a considerar desde el punto de vista de diseño, ya que la operación del filtro está limitada a la temperatura de los gases que resista el material filtrante (actualmente se puede trabajar hasta temperaturas del orden de 250°C); al grado de humedad que puede generar costras sobre las mangas; y al posible ataque químico por sustancias ácidas o alcalinas presentes en los gases. Todo ello disminuye la vida útil de las mangas, aumentando los costes de O&M, con la necesidad de recambio de las mismas. Principalmente existen dos tipos de configuración de los medios filtrantes:

a) Tejido hilado (TH): se trata de fibras que son extruidas en diferentes tamaños de hilo y se disponen en una saca tejida. Aquí la captación de partículas es superficial de dentro a fuera; están construidas de fibra de vidrio y/o poliéster, el método de limpieza es aire inverso/vibración y la velocidad de filtrado es baja.

b) Fieltro (F): se trata de fibras de ryton, acrilicos, y poliamidas entrelazadas mecánicamente; para mejorar su estabilidad, la mayoría de los fieltros cosidos tienen un elemento soporte en su centro. La captación de partículas es superficial de fuera a dentro; el método de limpieza es Pulse Jet y la velocidad de filtrado es alta.

En EE.UU, el tejido más comúnmente adoptado en centrales térmicas es la fibra de vidrio hilada, aunque con tendencia a la implantación de fieltros cosidos o membranas de politetrafluoroetileno (PTFE), tales como Gore-Tex sobre el vidrio hilado. En Canadá, Australia y Europa, las experiencias principales son con tejidos de fieltro (acrílicos como Dralon T, PPS o Ryton, Nomex, Huyglas, poliamida P84 y teflón). Estos materiales son aplicables para los filtros con método de limpieza por impulsos de aire a presión (Pulse Jet), y se han considerado más adecuados por su coste y fiabilidad [37].

Existen tres tipos de filtros de mangas atendiendo al método de limpieza utilizado:

- Filtro de mangas con limpieza por vibración
- Filtro de mangas con limpieza por aire a contracorriente
- Filtro de mangas con limpieza por impulsos de aire a presión (Pulse Jet).

Sin embargo, la opción más viable a la hora de transformar un precipitador electrostático en un filtro de mangas, reutilizando la estructura existente, es el sistema de limpieza "Pulse Jet", porque ocupan menor área en planta que los sistemas convencionales, adecuando la superficie requerida por las mangas de filtrado en el espacio disponible del precipitador electrostático.

Este sistema durante el ciclo de limpieza, el polvo se desaloja utilizando un impulso de aire comprimido que se inyecta en la parte superior de las mangas filtrantes. Este impulso a alta presión interrumpe el flujo normal del gas a través del filtro y crea una onda que al desplazarse a lo largo de la manga hace que esta se flexione, rompiendo la capa de polvo que cae hacia la tolva. La parte inferior de las mangas está cerrada y el impulso es lo suficiente grande para que pueda desplazarse a lo largo de la manga y regresar a la parte superior.

El impulso de aire de limpieza puede ser suministrado por un compresor o bien por un soplador retráctil, y típicamente el soplado se hace a una fila completa de mangas, a través de una lanza o tubería de soplado que contiene pequeños orificios "boquillas o venturis" para inyectar el aire a alta velocidad en la parte superior de las mangas [37]. Cada lanza está conectada a un tanque de aire a través de una válvula de diafragma. El tanque se mantiene normalmente 2 puntos por encima del valor nominal de la presión de trabajo, y a través de la lógica del ciclo de limpieza, que puede ser por un tiempo preestablecido o por una medida de ΔP; se energiza la válvula solenoide y se libera el paso a través de la válvula de diafragma, permitiendo el paso de aire a la lanza de soplado. A su vez, el sistema Pulse Jet, se subdivide en 3 tipos de lógica de funcionamiento, que se caracterizan por la presión y el volumen del aire utilizado para la limpieza:

- Alta Presión/Bajo Volumen (HP/LV)
- Media Presión/Medio Volumen (IP/IV)
- Baja Presión/Alto Volumen (LP/HV)

El método Pulse Jet de alta presión y bajo caudal de aire (HP/LV) es mucho más energético, lo que permite relaciones de caudal de gases /superficie de tejido, A/T (m3min-1/m2) o velocidad de filtrado (m/min), mucho mayores que en los sistemas convencionales. Este es el parámetro principal a la hora de dimensionar el filtro, ya que influye sobre el límite de las emisiones, la pérdida de carga del sistema, la frecuencia de los ciclos de limpieza (y por tanto, en la vida útil de las mangas) y en el tamaño necesario del filtro.

Teniendo en cuenta que se pretende reutilizar la estructura existente del precipitador electrostático, un valor elevado de A/T es fundamental, ya que conlleva ocupar menos área de filtrado, asegurando que el espacio disponible es suficiente para albergar las mangas y con ello minimizando las posibles modificaciones de la estructura que impactarían de modo sustancial sobre el coste inicial de la inversión.

4. Aplicación Práctica

Se ilustra la tecnología propuesta, aplicándola al caso real de un precipitador electrostático en la Unidad 4 (660MW) de una Central Térmica de Italia, de modo que nos familiaricemos con el tipo de información requerida y las áreas claves normalmente incluidas en una especificación de este tipo. La conversión fue completada en 12 semanas, en junio del año 2012, aprovechando una Revisión General de Turbina (RGT) de la Unidad.

4.1. Antecedentes

El precipitador electrostático de la unidad 4, es fabricado en el año 1990, y en el año 2002 se realiza una importante intervención de mantenimiento, principalmente en la parte mecánica. Después de una mejora en la eficiencia de captación, ha sufrido un degrado de las prestaciones que no garantizaban la funcionalidad de la unidad en general.

a) El Precipitador electrostático existente está formado por dos cuerpos en paralelo separados físicamente; cada cuerpo está compuesto de 14 campos eléctricos. Cada cuerpo está dotado de cámara superior al interno de la cual están ubicados los alimentadores de alta tensión (transformadores-convertidores). En la tabla 1 se muestran las principales características técnicas del mismo.

b) Los Ventiladores de tiro inducido (VTI) tienen: Potencia (7.000 kW), Tensión (6.000 V), N° de polos (8), Intensidad (804,1 A) y Factor de Potencia (0,86).

4.2. Descripción funcional de la transformación a PJFF

El objetivo principal de esta transformación es la de llegar a adecuar la estructura mecánica y la envolvente externa del actual precipitador electrostático, con el fin de reconvertir este sistema de captación de partículas en un filtro de mangas con sistema de limpieza pulse-jet (HP/LV). Los sistemas eléctricos auxiliares serán ubicados en la sala eléctrica del PE, previo desmontaje de los equipos no utilizables. Por otro lado, para poder compensar la mayor pérdida de carga introducida por el filtro, se sustituye el ventilador de tiro inducido (VTI) para aumentar la aspiración. El filtro es proyectado para una vida útil de 25 años, con 8.000 horas de funcionamiento y 50 arranques/paradas al año. El combustible de diseño es 100% carbón, cuya composición puede variar fuertemente en función del origen del suministro en el mercado spot, pero como referencia será carbón sudafricano. El uso de fuel será permitido solo durante el proceso de arranque de la unidad (calentamiento caldera).

Tabla 1.
Parámetros del PE.

Superficie de captación	139.212 m^2
SCA (Specific Colleting Area)	140 $m^2/m^3/s$
Presión del gas al ingreso	-3,92 kPa (-400 mmca)
Espacio entre electrodos y placas	300 mm
Velocidad medio de gases al interno	1,2 m/s
N° de campos eléctricos en serie	7
N° de transformadores-convertidores	28 (14 por cuerpo)
N° total de placas	4.116
Dimensiones de las placas colectoras	14,07 x 1,18 m

Fuente: Elaboración Propia

4.2.1. Descripción funcional de la transformación a PJFF

La estructura externa de los 2 cuerpos de PE es reutilizada, para ello es inicialmente previsto un lavado interno con agua, para a continuación proceder con las actividades de desmontaje de los siguientes elementos:

-Cámara Superior: sistema de calefacción con vapor, alimentadores de alta tensión (T/R) con toda la conexión eléctrica a los electrodos de descarga, aisladores soporte, techo "caliente" y "frío".

- Internos: electrodos de captación (placas completas), electrodos de descarga con la estructura soporte y el sistema de golpeo (martillos, yunques, motoreductores, etc). En la figura 1 muestra un detalle del desmontaje.

La siguiente fase es la construcción de los compartimentos internos a través de la instalación de paredes de separación (con los respectivos refuerzos), que dividen longitudinalmente cada cuerpo/semisección en 2 compartimentos (4 en total). El filtro podrá funcionar también con un compartimento excluso por anomalía o mantenimiento. La interceptación del compartimento se obtiene con la instalación de compuertas de accionamiento neumático al ingreso y a la salida. También, cada compartimento es dotado de un sistema de by-pass realizado a través de un conducto interno, gobernado con una compuerta neumática que mete en conexión el conducto de ingreso de gases con el de salida de los mismos, ver figura 2.

Externamente son reutilizados los conductos existentes al ingreso, ver figura 10, provenientes de los precalentadores de aire hasta las 4 bocas de entrada, para instalar a continuación:

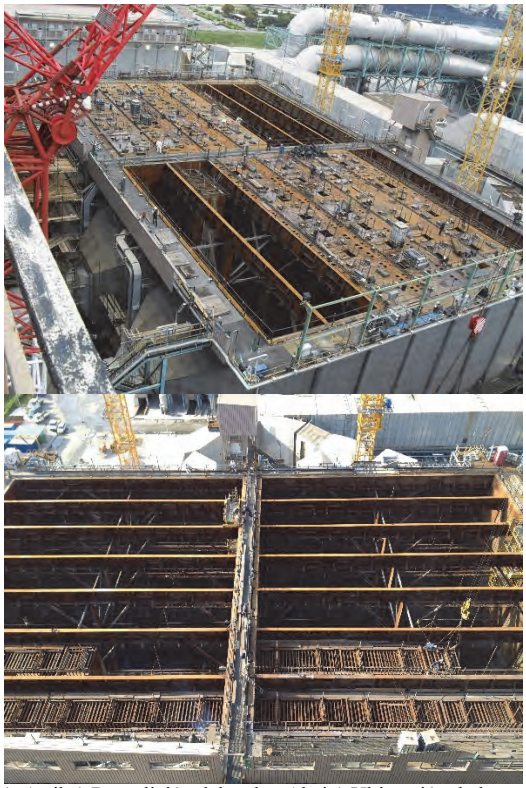


Figura 1. Arriba) Demolición del techo. Abajo) Ultimación de la extracción de todas las placas colectoras.
Fuente: Fotos propias.

- Un conducto de transición con rectificadores de flujo, para adquirir la mejor uniformidad del gas al ingreso
- Dos conductos de equilibrio, uno al ingreso y otro a la salida, para meter en comunicación a través de 2 compuertas de regulación neumáticas (posicionables manualmente), los compartimentos centrales B-C que pertenecen respectivamente al cuerpo 2 y 1 del filtro.

4.2.2. Elementos filtrantes: limpieza y producción de aire comprimido

El material base, constituyente de las mangas filtrantes empleadas es fieltro P84/PPS con armadura en PPS y tratamiento superficial PTFE. Este material consiente de operar a una temperatura nominal de 180°C con puntas de corta duración de 200°C, y posee una excelente resistencia a la abrasión, a los ácidos y alcalinos. En la figura 4, se muestran las mangas ya instaladas.

La limpieza de las mangas es de tipo pulse-jet, a alta presión y bajo volumen (HP/LV) basado en un sistema de producción y secado de aire comprimido, constituido de 4 compresores y 4 torres de secado (3 en servicio y uno en reserva).

Figura 2. Arriba) Inserción de las paredes divisorias para crear los compartimentos. Abajo) Montaje de los conductos de by-pass.
Fuente: Fotos propias.

Figura 9. Instalación del conducto de equilibrio del ingreso.
Fuente: Fotos propias.

El sistema cíclico de limpieza de las mangas debe ser en servicio cuando la unidad térmica es en funcionamiento, con la sola excepción de la fase de arranque, en la cual viene utilizado gasóleo o fuel y por tanto conviene mantener las mangas revestidas de calcio y/o cenizas para evitar fenómenos de ataque químicos en las mismas. La secuencia de limpieza prevé la posibilidad de utilizar dos velocidades:

- Con baja frecuencia, en la cual el tiempo de pausa entre dos impulsos sucesivos de limpieza (es decir, el tiempo que transcurre entre la limpieza de dos filas de mangas sucesivas) es alto, 12 segundos.
- Con alta frecuencia, en la cual el tiempo de pausa disminuye a 5 segundos.

La duración de cada singular impulso es de 100 milisegundos y, normalmente, cada manga recibe de 2,77 a 6,66 ciclos de limpieza cada hora, en función del intervalo de tiempo establecido entre un ciclo y otro (12 ò 5 segundos, respectivamente). En caso necesario (ΔP del singular compartimento es elevada y no disminuye), el operador puede establecer manualmente, tramite contraseña, una frecuencia de limpieza más alta hasta un máximo de 3,11 segundos de pausa entre un ciclo y otro. Las secuencias sobre los 4 compartimentos del filtro se desarrollan independientemente entre ellos, en base a la presión diferencial del respectivo compartimento, iniciando del lado de entrada de gases y excitando las primeras 8 electroválvulas de los primeros 8 tanques de aire (entendiendo como los primeros 2 de cada compartimento en el sentido perpendicular del flujo de gases). Después del tiempo de pausa establecido, serán excitadas las primeras 8 electroválvulas de los segundos 8 tanques, y así sucesivamente; terminados los 12 x 4 tanques de los compartimentos, el ciclo reparte con las segundas 8 electroválvulas de los primeros 8 tanques y así sucesivamente. En la figura 5 se muestra un detalle del compartimento D, con las electroválvulas y los tanques de aire.

Figura 4. Instalación de las mangas. a) Vista interna del compartimento b) Detalle del sistema snap-ring y de la jaula portamangas, visto desde la parte superior "walk in plenum".
Fuente: Fotos propias.

Figura 5. a) Detalle de un tanque de aire con 18 válvulas de diafragma b) Detalle de las tuberías de soplado (24 mangas/tubería) con las boquillas de inyección del aire.
Fuente: Fotos propias.

4.2.3. Sistemas de protección: alta T° y ΔP

El filtro de mangas en cada uno de los dos conductos de ingreso de distribución de gases a los compartimentos se instala un sistema de protección contra la altísima temperatura del gas, causada por transitorios en caldera. Dicho sistema está compuesto por un ventilador y tres válvulas tipo "poppets" de comando neumático.

Una vez que llegamos a 180°C el sistema genera una señal de alarma de altísima T° e inicia a abrir automáticamente la primera válvula poppets, transcurridos un tiempo (30 s) y si la T° media no ha bajado, vendrá abierta la segunda válvula y así hasta llegar a abrir las tres al 100% para inyectar lo que se denomina "aire falso" para refrigerar el compartimento. En el caso que no sea suficiente, y la temperatura de los gases llegue al valor de 200°C, se produce el disparo de la semisección con indisponibilidad de la misma, con cierre inmediato de las compuertas de ingreso y salida de los compartimentos de dicha semisección, con el fin de proteger las mangas de cualquier posibilidad de daño y apertura automática de las compuertas de by-pass.

Al igual que para la temperatura, se instalan 3 presostatos diferenciales para la revelación de la perdida de carga aguas arriba/debajo de cada compartimento del filtro, de modo que si se alcanza el valor de alarma de 25 mbar, el operador podrá actuar de modo manual (ya que el fenómeno del aumento de presión diferencial es lento) abriendo las compuertas de equilibrio de ingreso y salida. Si continua y se llega a un valor ΔP=30 mbar, se produce el disparo de la semisección con indisponibilidad de la misma, con cierre inmediato de las compuertas de ingreso y salida de los compartimentos de dicha semisección, con el fin de proteger las mangas de cualquier posibilidad de daño y apertura automática de las compuertas de by-pass.

4.2.4. Sistemas de protección: alta T° y ΔP

El sistema de ventilación "walk in plenum" se compone de 4 ventiladores (uno por compartimento) y es el que permite realizar intervenciones de mantenimientos (sustitución de mangas) en un compartimento excluido con el filtro de mangas en servicio, manteniendo una atmósfera transitable.

Una vez interceptado el paso de gases en el compartimento con el cierre de las compuertas de ingreso y salida, se abren las válvulas rompedoras de vacío (3 por compartimento) que equilibran la presión interna-externa, y se mete en servicio en ventilador dedicado al correspondiente compartimento en mantenimiento y se abre la compuerta neumática dedicada al walk in plenum de dicho compartimento.

4.2.5. Recubrimiento de las mangas

El correcto funcionamiento de un filtro de mangas requiere respetar algunas recomendaciones, en particular en las fases de paradas y arranques de la unidad. En caso contrario es posible un daño irreversible en las mangas, con una rápida reducción de las prestaciones del filtro en términos de mayores pérdidas de carga y emisiones. En particular se debe evitar:

- En el arranque de la unidad, las partículas finas de gasoil/fuel penetran al interno de la tela de las mangas, determinando así un saturación irreversible de la misma.
- Formación de condensados en contacto directo con las mangas, puesto que pueden cementar las cenizas sobre la tela y, si es de naturaleza ácida, atacar químicamente la misma fibra.

Durante el funcionamiento normal del filtro, es la costra de cenizas que se forma sobre las mangas la que realiza la función filtrante y de protección de la tela de la formación de condensados, absorbiéndolos. Durante la fase de arranque de la unidad, en la cual no hay cenizas, el estrato de protección debe ser creado artificialmente usando productos de adecuada basicidad y granulometría. Esta operación es la que se conoce como "precoating" (recubrimiento). En el filtro se ha instalado un sistema de precoating de las mangas, utilizando normalmente calcio hidratado, aunque también es posible el uso de cenizas almacenadas.

En el caso de una parada breve de la unidad (48-72 horas), es posible evitar el precoating al arrancar con la condición de mantener el estrato de cenizas sobre las mangas y evitar la formación de condensados, por tanto se debe:

- Parar el sistema de limpieza de las mangas cuando se pare el grupo, para evitar el desprendimiento de las cenizas de las mangas

- Aislar los compartimentos del filtro manteniendo cerradas las compuertas de ingreso y salida de gases y manteniendo elevada la temperatura interior utilizando los calentadores de las tolvas inferiores

En el caso de paradas de larga duración (>72 horas), no se puede en ningún caso dejar el estrato de cenizas sobre las mangas. De hecho las cenizas, siendo higroscópicas, absorberían la humedad del aire hasta cementar sobre las mismas mangas. Por tanto, se debe meter en servicio el sistema de limpieza de las mangas y la evacuación de las tolvas durante unas horas, sucesivamente a la parada del grupo, con el objetivo de desprender la mayor cantidad posible de cenizas de las mangas.

La correcta secuencia del precoating, se obtiene metiendo en servicio los dos ventiladores de tiro inducido a máximo caudal, para obtener en los compartimentos una velocidad ascensional del aire óptima para el revestimiento de las mangas, minimizando las perdidas/caídas del agente químico (calcio hidratado) a las tolvas. Simultáneamente el agente es suministrado e impulsado mediante camiones autobotes que a través de un circuito isométrico, y se introduce en los conductos de ingreso de gas de cada semisección del filtro. La inyección de dicho agente se realiza hasta obtener un valor de $\Delta P \sim 2 \div 3$ mbar, que indica el correcto revestimiento de las mangas.

4.2.6. Recubrimiento de las mangas

A la hora de instalar el filtro de mangas, se plantea necesario realizar una estimación de la pérdida de carga total del sistema (incluidos los conductos), para establecer si los actuales ventiladores de tiro inducido (VTIs) son adecuadamente dimensionados para las nuevas condiciones de trabajo. Los datos para el dimensionamiento de los mismos son, tabla 2:

- Asumir como dato de partida del diagrama piezomètrico de la instalación, la actual perdida del circuito: 68 mbar

- Las entradas de aire a considerar en los precalentadores de aire (lujngstrom) es el 15%@2.600.000 Nm3/h

- Las ΔP debidas a los conductos y a las compuertas de entrada y salida, se establecen típicamente en un rango de 51 - 67 mbar

- Las ΔP debidas a la placa portamangas, normalmente está establecida en un rango de 152-220 mbar durante el funcionamiento normal del filtro (todos los compartimentos en servicio), pero se debe prever que dichos valores se pueden incrementar si algún compartimento está aislado por limpieza “off-line” y/o mantenimiento

- Por tanto, la ΔP total del sistema del filtro de mangas, se estima en 20 mbar**.**

Tabla 2.
Datos de dimensionamiento del VTI.

Caudal (Nm^3/h)	Presión (mbar)	T (°C)
2.800.000	89	130

Fuente: Elaboración Propia.

Tabla 3.
Condiciones del gas al ingreso del filtro de mangas.

Caudal total de gases	2,6 x 106 Nm3/h
Temperatura de gases	130°C
Presión estática	- 40 mbar
Cantidad de partículas	15 g/Nm3
Tamaño medio de partículas	4 micron

Fuente: Elaboración Propia.

Así, se modificaron los motores de los actuales ventiladores de tiro inducido (VTI), sustituyéndolos por nuevos motores adaptados para ser accionados por Convertidores Electrónicos de Frecuencia (CEF).

4.3. Resultados

Los resultados de la unidad se analizan a través de varios tests de Rendimiento, los cuales se realizan, a los 6 meses y 18 meses, después de la fecha de puesta en marcha de la instalación.

Dichos tests se realizan de acuerdo a la normativa vigente de aplicación para la determinación de la cantidad de partículas presentes en los gases de combustión:

- UNI EN 13284 – 1 Enero 2003: Determinación de la concentración de partículas a la salida del filtro de mangas.

- UNI EN 13284 – 1 ò ASME PTC 38: Determinación de la concentración de partículas a la entrada del filtro de mangas.

Las condiciones del gas al ingreso del filtro se muestran en la tabla 3. El parámetro fundamental a verificar, es la cantidad de partículas en los conductos de salida del walk in plenum, donde están instalados 2 instrumentos de control por cada compartimento. El dato debe ser referido a plena potencia de la instalación.

Tras los tests se obtuvo una máxima concentración de partículas a la salida del filtro a mangas, con todos los compartimentos en servicio de 20 mg/Nm^3 (gas seco@6% O2) de concentración de media horaria, cumpliéndose así las expectativas esperadas.

5. Conclusiones y Recomendaciones

En este trabajo se han mostrado los principales problemas que afectan al funcionamiento de los precipitadores electrostáticos en las grandes instalaciones de combustión, que sumado a los límites de emisión más restrictivos hacen que la transformación a un filtro de mangas pulse-jet, aprovechando la estructura existente sea una opción particularmente interesante a considerar a la hora de afrontar importantes inversiones de mejora o sustitución de los precipitadores existentes, ya que se adaptan al espacio disponible, evitando costosos reemplazos completos, y alcanzando mayores rendimientos de captación de partículas.

Sin embargo, cada caso es diferente y va estudiado en detalle ya que depende de las condiciones específicas de cada instalación. Por ello, queremos realizar una serie de recomendaciones, deducidas de la aplicación presentada en este artículo y que deberían tenerse en cuenta para eliminar los factores que puedan contribuir al mal funcionamiento del equipo:

- El tiempo de instalación requerido varía entre 12-14 semanas. Por tanto, la viabilidad económica del proyecto sólo se puede justificar si la construcción del filtro de mangas se combina en una típica revisión general de turbina o caldera de la unidad de producción. En este caso es fundamental durante la fase de diseño maximizar en el ensamblaje de componentes y realizar un detallado programa de demolición y montaje.

- Debe existir suficiente área en planta, para adaptar el espacio requerido por las mangas. En este caso es fundamental un valor elevado de la velocidad de filtrado A/T, reduciendo las modificaciones estructurales, a la construcción de los compartimentos.

- Realizar un estudio fluido-dinámico durante la fase de diseño que permita conocer la velocidad del gas y la distribución de partículas en el interior de los compartimentos, para establecer los deflectores y desviadores adecuados de modo que el flujo sea uniforme, y alargue la vida de las mangas (3-4 años). Además, dicho estudio nos permitirá estimar las nuevas pérdidas de carga introducidas, para decidir si podemos conservar los ventiladores inducidos existentes.

Referencias

[1] Baños, R., Manzano-Agugliaro, F., Montoya, F.G., Gil, C., Alcayde, A. and Gómez, J., Optimization methods applied to renewable and sustainable energy: A review. Renewable and Sustainable Energy Reviews, 15(4), pp. 1753-1766, 2011. DOI: 10.1016/J.RSER.2010.12.008

[2] Fernández-García, A., Rojas, E., Pérez, M., Silva, R., Hernández-Escobedo, Q. and Manzano-Agugliaro, F., A parabolic-trough collector for cleaner industrial process heat. Journal of Cleaner Production, 89, pp. 272-285, 2015. DOI: 10.1016/ j.jclepro.2014.11.018

[3] Gallego-Vega, L.E. and Duarte-Velasco, O.G., The effect of electric transmission constraints on how power generation companies bid in the Colombian electrical power market. Ingeniería e Investigación, [Online]. 30(2), pp. 62-77. 2010. Available at: http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120-56092010000200007&lng=en&nrm=iso. ISSN 0120-5609.

[4] Manzano-Agugliaro, F. and Carrillo-Valle, J., Location of air in-leakage in power plants condensers by helium test. DYNA, 86(2), pp. pp. 173-181, 2011. DOI: 10.6036/3916

[5] Maioli, A., Pulfer, J. and Mitjans, F., Generating electricity during peak hours in Asuncion, Paraguay, through anaerobic digestion of cultivated water hyacinths. Ingeniería e Investigación, [Online]. 31, pp. 66-70, 2011 Available at: http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120-56092011000500010&lng=en&nrm=iso. ISSN 0120-5609.

[6] Montoya, F.G., Aguilera, M.J. and Manzano-Agugliaro, F., Renewable energy production in Spain: A review. Renewable and Sustainable Energy Reviews, 33, pp. 509-531, 2014. DOI: 10.1016/j.rser.2014.01.091

[7] de Miera, G.S., del Río-González, P. and Vizcaíno, I., Analysing the impact of renewable electricity support schemes on power prices: The case of wind electricity in Spain. Energy Policy, 36(9), pp 3345-3359, 2008. DOI: 10.1016/j.enpol.2008.04.022

[8] Wen, T.-Y., Wang, H.-C., Krichtafovitch, I. and Mamishev, A.V., Novel electrodes of an electrostatic precipitator for air filtration. Journal of Electrostatics, 73, pp. 117-124, 2015. http://dx.doi.org /10.1016/j.elstat.2014.11.002

[9] Gutiérrez-Ortiz, F.J., Navarrete, B. and Cañadas, L., Assessment of plate-wire electrostatic precipitators based on dimensional and similarity analyses. Fuel, 90(9), pp. 2827-2835, 2011. DOI: 10.1016/j.fuel.2011.04.006

[10] Gutiérrez-Ortiz, F.J., Navarrete, B. and Cañadas, L., Dimensional analysis for assessing the performance of electrostatic precipitators. Fuel Processing Technology, 91(12), pp. 1783-1793, 2010. DOI: 10.1016/j.fuproc.2010.07.013

[11] Yan, C., Liu, G. and Chen, H., Effect of induced airflow on the surface static pressure of pleated fabric filter cartridges during pulse jet cleaning. Powder Technology, 249, pp. 424-430, 2013. DOI: 10.1016/j.powtec.2013.09.017

[12] Dang, X.-Q., Pang, M., Li, X., Zhang, J and Li, Q., Discussion on influencing factors of the pulse-jet performance of fabric filter. 2011 International Conference on Electric Technology and Civil Engineering, ICETCE 2011 - Proceedings, art. 5776073, pp. 1167-1170, 2011. DOI: 10.1109/ICETCE.2011.5776073

[13] Ait-Said, H., Nouri, H. and Zebboudj, Y., Effect of air flow on corona discharge in wire-to-plate electrostatic precipitator. Journal of Electrostatics, 73, pp. 19-25, 2015. DOI: j.elstat.2014.10.004

[14] Jdrusik, M., Gajewski, J.B. and Wierczok, A.J., Effect of the particle diameter and corona electrode geometry on the particle migration velocity in electrostatic precipitators. Journal of Electrostatics, 51-52 (1-4), pp. 245-251, 2001. DOI: S0304-3886(01)00047-X

[15] Vélez, F., Chejne, F. and Quijano, A., Thermodynamic analysis of R134a in an Organic Rankine Cycle for power generation from low temperature sources. DYNA, 81(185), pp. 153-159, 2014. DOI: 10.15446/dyna.v81n185.37598

[16] Vélez, F., Selecting working fluids in an organic Rankine cycle for power generation from low temperature heat sources. DYNA, 81(188), pp. 173-180, 2014. DOI: 10.15446/dyna.v81n188.41666

[17] Directiva de Grandes Instalaciones de Combustión (GICs) 2001/80/CE, 2001.

[18] Directiva de Emisiones Industriales (DEI) 2010/75/UE, 2010.

[19] Pavlish, J.H., Sondreal, E.A., Mann, M.D., Olson, E.S., Galbreath, K.C., Laudal, D.L. and Benson, S.A., Status review of mercury control options for coal-fired power plants. Fuel Processing Technology, 82(2-3), pp. 89-165, 2003. DOI: 10.1016/S0378-3820(03)00059-6

[20] Wara, M., Is the global carbon market working? Nature, 445 (7128), pp. 595-596, 2007. DOI: 10.1038/445595a

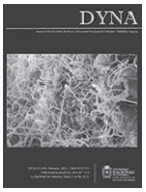
[21] Córdoba, P., Status of Flue Gas Desulphurisation (FGD) systems from coal-fired power plants: Overview of the physic-chemical control processes of wet limestone FGDs. Fuel, 144, pp. 274-286, 2015. DOI: 10.1016/j.fuel.2014.12.065

[22] Song, K., Jang, Y.-N., Kim, W., Lee, M.G., Shin, D., Bang, J.-H., Jeon, C.W.and Chae, S.C., Factors affecting the precipitation of pure calcium carbonate during the direct aqueous carbonation of flue gas desulfurization gypsum. Energy, 65, pp. 527-532, 2014. DOI: 10.1016/j.energy.2013.11.008

F. Manzano-Agugliaro, Obtuvo su título de Dr. Ingeniero en 1995, por la Universidad de Córdoba, España. Es profesor de la Escuela Superior de Ingeniería de Universidad de Almería, España, desde 1996, Profesor Titular de Universidad del departamento de Ingeniería desde 2001, y desde 2011 está acreditado como catedrático de Universidad por la Rama de Ingeniería y Arquitectura. Su investigación está centrada en torno a la energía, el medio ambiente, y la geomática.
ORCID: 0000-0002-0085-030X

J. Carrillo-Valle, es Ing. Industrial por la Universidad de Málaga, España. En 2006 obtuvo el MSc. en "Tecnologías de Generación de Energía Eléctrica" por la Escuela Técnica Superior de Ingeniería Industrial de la Universidad Politécnica de Madrid, España. Actualmente realiza estudios en el programa de doctorado de Ingeniería Ambiental por la Universidad de Almería, España. Ha trabajado en empresas del sector energético, principalmente en el mantenimiento y operación de centrales térmicas de carbón, en diversos países como España, Italia o Chile.
ORCID: 0000-0003-2406-9427

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS



DYNA
http://dyna.medellin.unal.edu.co/

Design of an open source-based control platform for an underwater remotely operated vehicle

Luis M. Aristizábal [a], Santiago Rúa [b], Carlos E. Gaviria [c], Sandra P. Osorio [d], Carlos A. Zuluaga [e], Norha L. Posada [f] & Rafael E. Vásquez [g]

Escuela de Ingenierías, Universidad Pontificia Bolivariana, Medellín Colombia.
[a] luismi911@gmail.com, [b] santiago.ruape@upb.edu.co, [c] cgaviriacelis@gmail.com, [d] sandrapatricia.osoriog@gmail.com, [e] carlos.zuluaga@upb.edu.co, [f] norha.posada@upb.edu.co, [g] rafael.vasquez@upb.edu.co

Received: March 25th, de 2015. Received in revised form: August 31th, 2015. Accepted: September 9th, 2015

Abstract
This paper reports on the design of an open source-based control platform for the underwater remotely operated vehicle (ROV) Visor3. The vehicle's original closed source-based control platform is first described. Due to the limitations of the previous infrastructure, modularity and flexibility are identified as the main guidelines for the proposed design. This new design includes hardware, firmware, software, and control architectures. Open-source hardware and software platforms are used for the development of the new system's architecture, with support from the literature and the extensive experience acquired with the development of robotic exploration systems. This modular approach results in several frameworks that facilitate the functional expansion of the whole solution, the simplification of fault diagnosis and repair processes, and the reduction of development time, to mention a few.

Keywords: open-source hardware; ROV control platforms; underwater exploration.

Diseño de una plataforma de control basada en fuente abierta para un vehículo subacuático operado remotamente

Resumen
Este artículo presenta el diseño de una plataforma de control basada en fuente abierta para el vehículo subacuático operado remotamente (ROV) Visor3. Primero se describe la plataforma de control original del vehículo con arquitecturas cerradas de hardware y software. La modularidad y la flexibilidad se establecen como guías para el diseño propuesto, dadas las limitaciones de la infraestructura previa. El nuevo diseño incluye las arquitecturas de hardware, software, firmware y control. Se usan plataformas abiertas de hardware y software para el desarrollo de la nueva arquitectura del sistema, con soporte en la literatura y la extensa experiencia adquirida en el desarrollo de sistemas robóticos de exploración. Esta aproximación modular arroja varias plataformas que facilitan, entre otros: la expansión funcional de la solución completa, la simplificación de los procesos de diagnóstico y reparación de fallos, y la reducción del tiempo de desarrollo.

Palabras clave: hardware de fuente abierta; plataformas de control de ROVs; exploración subacuática.

1. Introduction

Several underwater operations in aquaculture, port security, archeology, marine biology, offshore industry, to name but a few, are performed nowadays using underwater remotely operated vehicles (ROVs) [1]. The NORSOK U-102 standard [2] classifies such vehicles as follows:

- Class I – Pure observation
- Class II – Observation with payload option
- Class III – Work class vehicles
- Class IV – Seabed-working vehicles
- Class V – Prototype or development vehicles

A typical ROV system is comprised of an underwater vehicle, connected by a tether cable to a surface station where different tasks are specified through a mission control software interface.

Correa et al. [3] described the architecture for the conceptual design of underwater exploration vehicles; they state that several mechanical and hardware components are required for the appropriate operation of the system. The hardware subsystem constitutes the ROV's nucleus since it concentrates information provided by propulsion elements, a vision system, instrumentation, a power supply, a control system, among others [4].

The brain of the ROV is the on-board processor since it executes navigation, guidance, and control algorithms, and manages information exchange with the surface station using fiber optic communications, etc. [5]. Requirements such as robustness, reliability, processing speed, memory space, and flexibility are commonly specified during the control platform's design process [6].

Deep-water ROV operations demand navigation systems with high performance computation requirements since they use, for instance, Kalman filter-based algorithms to estimate the position, orientation, and velocity of the vehicle from measurements provided by different sensors (IMU, magnetometer, depth meter, DVL, USBL, among others) [7-12]. Such computation requirements are commonly met by using high-cost embedded processors with real-time operating systems such as the ones presented in [13,14].

Nowadays, several underwater vehicles and robotic platforms have been developed with low cost components [15-19], for instance, control platforms that use Arduino®-based hardware [20-25]. These open-hardware platforms can be used in ROVs that are intended for inspection tasks which are performed at low depths, and that do not require high performance real-time computations. Typically, Arduino®-based hardware is used to perform low-level tasks, such as data acquisition and communication with sensors and actuators, while the implementation of control algorithms and surface communication systems is done with open-source single-board computers such as Beaglebone® or Rasperry PI® [26]. Arduino platforms simplify the process of developing firmware by using a high level programming language, which implies a trade-off when trying to optimize code.

The use of low-cost open-hardware robotic platforms can be useful to motivate learning activities in people with no previous high-tech education; hence, they can be used for teaching tasks in programs related to science, technology, engineering, mathematics (STEM), ocean engineering, and marine sciences [27,28]. An interesting example is the MIT Sea Grant college program, developed with help from the Office of Naval Research (ONR) in the United States, and within that the Sea Perch ROV [29]. This ROV is comprised of low-cost equipment in order to easily teach students how to build an underwater vehicle [30].

This work addresses the design of an open-source control platform for the underwater remotely operated vehicle Visor3 [13], which has been used as a test platform for the development of robotic technology for the underwater exploration of Colombian seas. In this work, expansion capability and modularity of hardware, firmware, and software are established as general guidelines for the design process, since they are fundamental for open-source solutions.

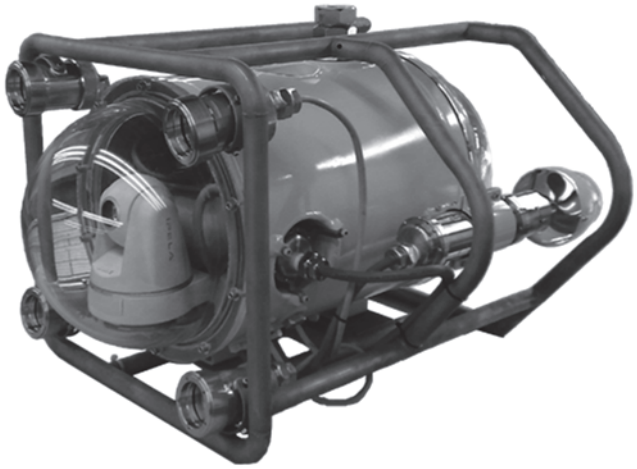


Figure 1. Underwater remotely operated vehicle Visor3.
Source: [13]

Section two of the paper describes the ROV Visor3 and details its original control platform and associated problems. The third section describes the approach followed to develop the new open-hardware architecture. Section four addresses software architecture development for each stage of the vehicle; including firmware, vehicle software, and surface station software. The fifth section describes the vehicle's control architecture and its relation with the hardware architecture elements. Finally, some conclusions are provided.

2. Remotely operated vehicle Visor3

Visor3, Fig. 1, is a Class I ROV that was designed to acquire visual information for the surveillance and maintenance of ships hulls and underwater structures in port facilities, and oceanographic research tasks [13]. The original hardware architecture of this ROV system was divided into two main parts: the surface control station that is comprised of an industrial joystick as command input, and a computer used as a Human-Machine Interface (HMI), connected to the vehicle using a fiber optic link; and the vehicle which is comprised of an IP camera, fiber optic communications devices, an embedded processor, multiple sensors (inertial measurement unit, thermometer, compass, etc.), thrusters, drivers, and DC power units [31], Fig. 2.

2.1. Original control platform

Visor3's original control platform was conceived, within a general framework for the control of unmanned vehicles [32]. This infrastructure relies on a central PC/104 form factor embedded processor that would allow one to use a real-time operating system (RTOS) in order to guarantee execution times for critical tasks, commonly present in systems such as in some unmanned vehicles with fast dynamics.

As is shown in Fig. 2, a CAN network was used to connect sensors, actuators, and other devices inside the vehicle. The network was designed as a modular tower board

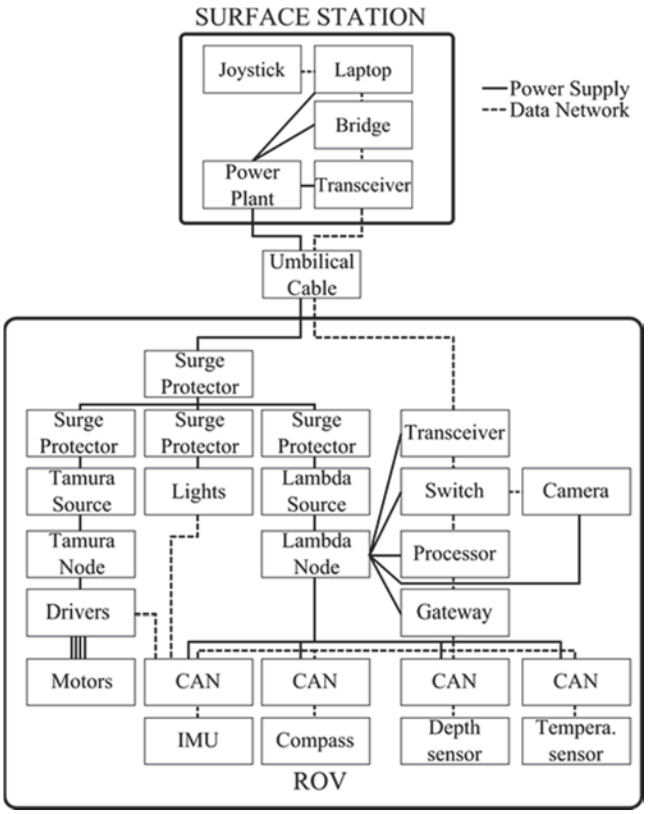


Figure 2. Hardware infrastructure in Visor3.
Source: The authors

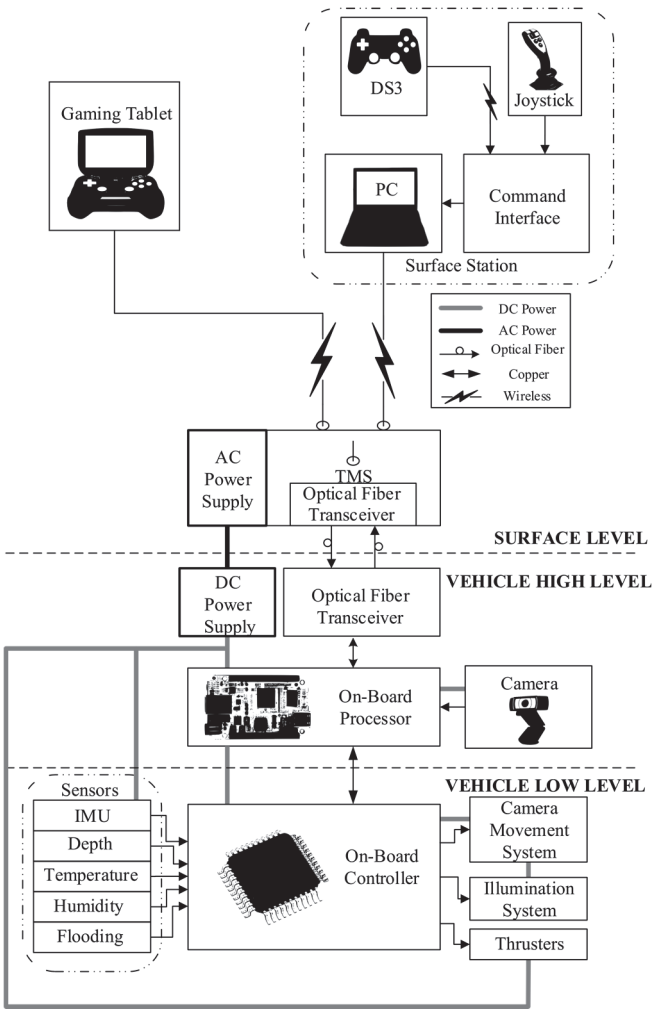


Figure 3. Proposed hardware architecture.
Source: The authors.

platform, with six independent boards used to integrate sensors and actuator signals. Such general control architecture resulted in a complex bulky system with a lack of reliability; a negative characteristic for an underwater exploration system intended to be used in maintenance activities and oceanographic research.

After several deployments and operations on both laboratory and field environments with Visor3, the failure rate for both hardware and software increased. Additionally, due to the complexity of the control architecture, diagnostics and repair processes were delayed. Furthermore, the availability of technical information for several devices, e.g. the embedded processor, and the proprietary nature of software and hardware, restricted the implementation of improvements in order to increase the system's reliability.

3. New hardware architecture

The new open source-based hardware architecture has been designed by taking Visor3's original requirements and the experience acquired with its previous implementation into consideration. Additional considerations were taken into account, for instance, the use of low cost, highly available components, and preferably with a wide support community.

The hardware architecture is based on a hierarchical structure, with three defined levels. A higher, or surface level, that has all the components that are located outside the vehicle, i.e., in the surface control station, and includes the tether, power supply, and optical fiber communication devices. Middle and lower levels are located inside the ROV, so they can be identified as a vehicle's high and low-level respectively, as seen in Fig. 3.

3.1. Surface level

From a functional perspective, the surface level is comprised of two elements: a Tether Management System (TMS) and an operator interface. TMS is in charge of ensuring the ROV's power demand, and communication with the operator's interface, which includes a set of tools used by the pilot to command and drive the ROV.

Communication between the TMS and the operator interface is made through a wireless high speed WiFi link, thus allowing one to use a variety of open-source software enabled devices to control the ROV, e.g., personal computers (PC) equipped with joysticks and buttons, gaming tablets based on Android Operating Systems (OS) [33,34], and so on.

For this ROV, two operator devices are used. The first one is the surface station, and is comprised of the following devices: a computer for processing data coming from the

vehicle, including video streaming and visualization; a command interface, i.e., a device that integrates control elements, such as joysticks, buttons, indicators and a Global Positioning System (GPS), that gathers information and sends it to the computer; and an auxiliary wireless gamepad (DualShock3® or DS3) that can be connected through the command interface, and offers a portable alternative for the main control elements. The second device is wholly based on a commercial gaming tablet computer running Android OS, in which a Human-Machine Interface (HMI) can be implemented, allowing one to integrate video stream visualization, touch screen capabilities, analog joysticks, digital buttons and directional pads in a small, handheld device. The latter option was chosen mainly because Android OS based systems have been increasingly used in automation and remote control of vehicles [33], but it is not within the scope of this work.

3.2. Vehicle level

Regarding the on-board hardware system, it was devised as a two-layer structure, with each layer being a centralized subsystem controlled by a processing unit. The upper layer's central unit is an on-board open source embedded computer that is responsible for data reception from the surface station, video acquisition from a high-definition USB web camera, processing and streaming acquired images over a TCP/IP network, along with the vehicle status data, and, in short, every high level processing task that must be executed on-board. The lower layer's processing unit is an Arduino-compatible, 32-bit Micro Controller Unit (MCU) running at 84 MHz bus clock which has direct access to the ROV's devices, such as motor drivers and thrusters, lights, camera movement, and sensors, which include fault detection sensors (internal temperature and flooding are measured within the vehicle) and navigation sensors (Inertial Measurement Unit IMU and a depth meter). A communication link has been established between both layers to integrate all the ROV devices; this link allows one to treat the hardware system as a single unit. The integration of physical devices is made through a custom-made circuit board, designed to be compatible with additional components that allow the ROV extend its functionality.

Additional efforts are being made in order to maintain the architecture's flexibility to consider future growth and expansion; specifically on the ROV instrumentation, e.g., installing underwater positioning devices: an Ultra-Short Base Line (USBL) for positioning, a Doppler Velocity Log (DVL) for speed measurements, among others. The designed data acquisition system is equipped with appropriate hardware for reading additional sensor measurements, including standard connections used for instrumentation; for example, 4-20 mA signals or RS-232 serial communication. Moreover, an Arduino shield socket has been included in the system for the fast integration of Arduino shield-type expansion cards.

4. Software architecture

To define the appropriate software architecture for an ROV, some requirements must be taken into account. Standards matching and portability are important aspects that need to be considered; in addition, the complexity of such applications requires modularity, reusability, and ease of integration between every element [35]. Hence, modularity is the main aspect that defines the guidelines for the final design of this software architecture.

A modular architecture confers a significant advantage over other approaches, as it supports massive reuse of existing code through different solutions; this allows one to reduce the development time of new applications, as has been demonstrated by several software platforms used for robotic devices and unmanned vehicles [36-39]. Although these software implementations are usually aimed at high level processing platforms, the designed architecture covers both high and low-level platforms making use of object-oriented programming, which has been recently enabled and optimized for microcontroller systems [40].

The general software architecture, as shown in Fig. 4, is composed of three layers. The bottom layer is firmware-based, and is in charge of low-level control of most devices in the ROV. The middle layer runs on the embedded on-board processor, and acts as middleware between the surface station and the low level controllers; it is in charge of communicating both layers and integrating a video signal. Finally, the upper level implements the user interface, including command devices, and every program and routine that cannot be executed on-board due to processing power limitations.

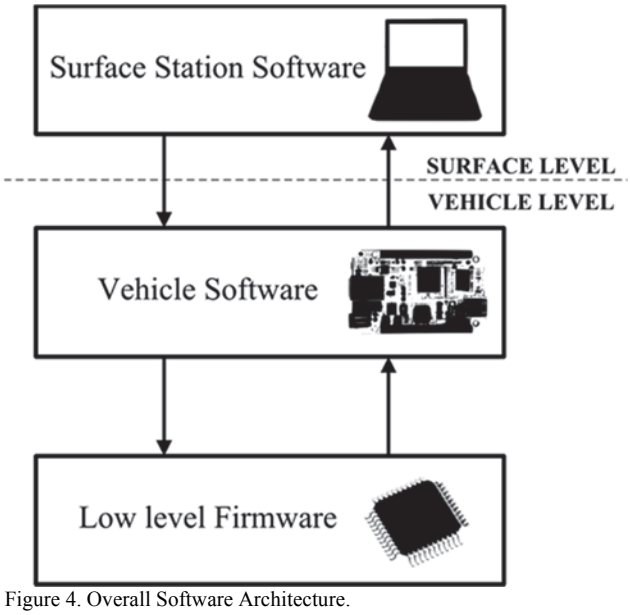


Figure 4. Overall Software Architecture.
Source: The authors.

4.1. Low level firmware

Following the established guidelines of modularity, and making use of an available open-source Object-Oriented Programming environment for microcontrollers, a firmware class structure was defined using C++ language, see Fig. 5. It is based on modules that can be added or removed without

significant modifications to the program. For this implementation, modules are treated as devices, mainly because of their specific functions related to the vehicle.

Every device depends on a group of core modules that cover basic functions, such as coordination and communication, between devices and with upper level elements of the architecture.

Device manager. This is the main core module of the firmware, designed as a general abstract base class that serves as a template for other devices. Every new device is a derivation of the device manager, allowing one to take advantage of polymorphism, and greatly simplifying the process of incorporating devices to the main program's workflow by providing a single method for executing every device's routine, regardless of the number of devices. Nevertheless, the maximum number of devices is limited, due to physical factors, e.g., memory or processing capacity, by the target implementation device.

Shared memory. In this particular application, communication between devices is a requirement due to functionality and safety of the vehicle; for instance, the fault detection module that depends on information from sensors, or the motor controller that needs information about energy consumption levels. For this reason, a shared memory module serves this purpose, allowing devices to share data seamlessly.

Communications module. This module handles communications with upper layers of the architecture, collecting data from devices, encapsulating them, and sending messages periodically according to the message's priority. The communication module is also in charge of receiving, parsing, and storing incoming data from higher level devices, making it available for the recipient device.

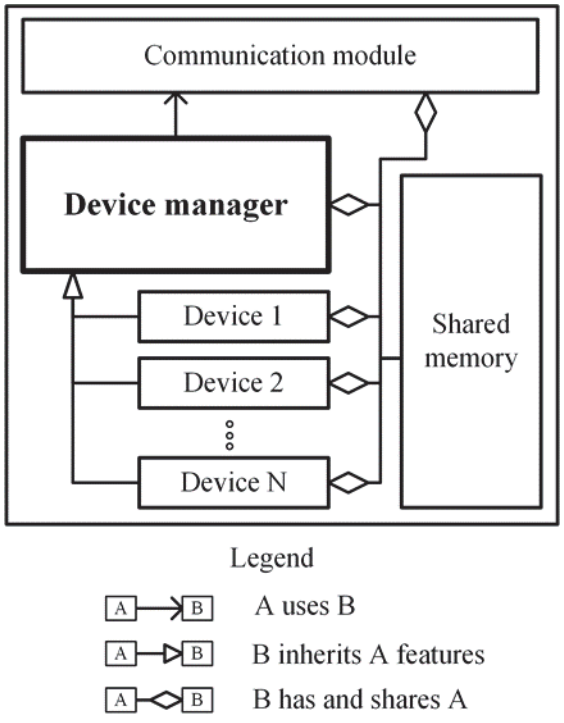


Figure 5. Low-level firmware general class structure.
Source: The authors

In the program flow, each device's routine is executed sequentially, each of their instances have been declared in the same order. Timer-based loops are defined for ensuring the time execution requirements of each device, e.g., a 10-millisecond loop must be used for acquiring IMU samples, and an additional 1000-millisecond loop is used for reading slower sensors, such as temperature, pressure, and flooding.

4.2. Vehicle's software

The vehicle's software is implemented in a Beaglebone embedded computer. This processor runs an Ångström operating system which is a Linux distribution used in embedded devices with built-in components. From the point of view of functionality, the vehicle's onboard processor is responsible for executing two main processes: the first one is focused on running a video streaming server that deals with camera operation, receives frames, compresses and sends them to the surface station; and the second process handles communication between the surface station and the vehicle, i.e., it allows data to be exchanged between the surface level and the vehicle level.

For communication with sensors, according to the proposed hardware architecture, the BeagleBone uses a serial link with the microcontroller that handles low-level processes. Then, the processor is responsible for sending all data acquired from sensors to the surface station, via UDP protocol. Both communication tasks are performed by a routine developed on a programming environment based on JavaScript called Node.js, which allows one to create highly scalable network programs such as web servers (streaming video, TCP / UDP, etc.). This framework enables the development of modular software to import and export modules developed by third parties, implementing several middleware utilities to be used in web applications, such as express frameworks [41].

Further Node.js comes with an API covering low-level networking, basic HTTP server functionality, file system operations, compression and many other common tasks. These libraries are used to adjust the communication protocol and send data frames to the surface station via Ethernet on UDP protocol [42].

4.3. Surface station's software

A similar approach to the one used in the firmware architecture was applied on the surface station, i.e., easy integration of specific functions in the form of modules, through a framework that handles essential tasks to ensure cohesion between all parts of the solution. The main difference between software and firmware is the target hardware platform; firmware runs on a device with limited processing resources compared to a personal computer, enabling the latter to be used with higher level developing environments.

To take advantage of the capabilities of the surface hardware platform, which in this case is a personal computer (PC) running Windows® OS, National Instruments™ LabVIEW® graphical dataflow programming environment was selected in order to allow one use parallel execution code through processor multithreading [43].

The proposed software structure uses parallel execution with three loops designed to run simultaneously; each loop takes care of a group of tasks, as shown in Fig. 6. The first loop, top down, is in charge of receiving and assembling incoming data from the vehicle and the command interface (see Fig. 3); then, it stores each message in a buffer for further reading and processing. The middle loop takes assembled frames as they become available and extracts, classifies, and stores data, making them available to other modules in the program. The bottom loop is responsible for collecting command data and transmitting them to the ROV; this is the designated space for adding functional modules to the program. This loop allows one to add and remove functional modules, as long as their time requirements meet the designated loop time characteristics; they are executed sequentially. For example, if the navigation algorithm's execution time is at least 100 ms, and the transmission loop time is defined to 50 ms, a delay will occur in the data exchange with the ROV. To solve this issue, additional loops with specific timings, different from the existing loops can be created, allowing one to add functional modules with longer or shorter execution times.

Caution measures must be taken when designing functional modules regarding execution time, due to limitations on the processing power of the host machine, so undesired delays can be avoided.

5. Open-source control platform

The navigation, guidance and control (NGC) system for an underwater vehicle can have different degrees of sophistication, depending on the type of operation that is to be performed, and the autonomy levels that need to be achieved [44, 45].

One of the important vehicle design parameters is the number of degrees of freedom needed to perform the planned operations, because they represent the number of independent

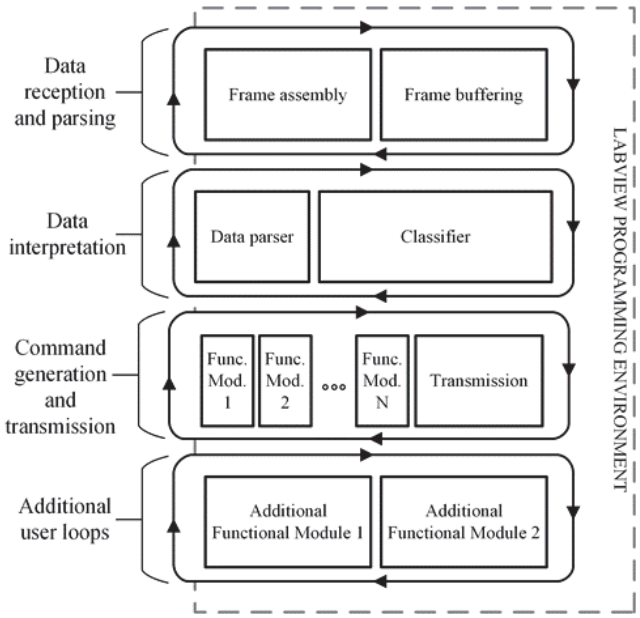


Figure 6. Surface Station software structure.
Source: The authors.

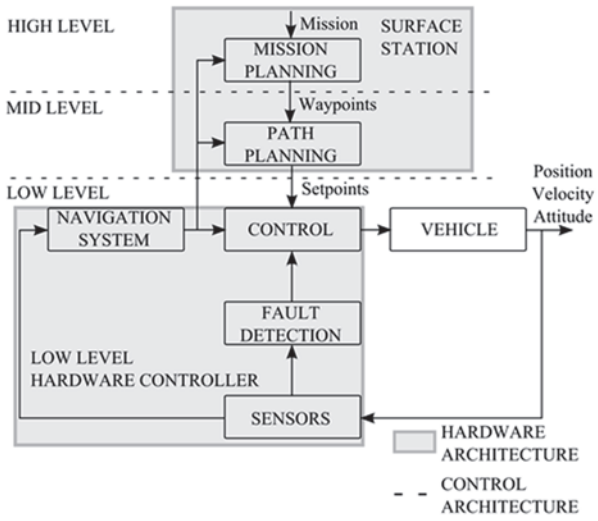


Figure 7. Control and hardware architecture of the Visor3 ROV.
Source: The authors.

movements that the vehicle can achieve in the three-dimensional space. Additionally, the tasks that are to be performed determine the instrumentation (sensors, actuators, complementary systems, among others) required to control the vehicle. Fig. 7 shows a three-level hierarchical NGC structure for an underwater vehicle; this kind of structure is useful to control and stabilize the vehicle. Additionally, each element of the control architecture is related to a specific group of hardware that executes the corresponding tasks.

The high and middle levels in the control structure are executed in the surface station. The objective is to allow the operator define the vehicle's mission using a high level language and translate it into commands to the low-level control system. Through the use of the PC, joystick and buttons, the operator is capable of moving the ROV in accordance with the independent movements that the vehicle can make. The joystick's three degrees-of-freedom are used to move in surge, sway, and yaw directions; heave motions are commanded by using 2 buttons, with increments and decrements of the speed in this direction (up/down). The graphical user interface (GUI) allows the operator to follow the mission's progress and the position of the vehicle.

The low-level control structure is executed in the ROV's hardware. The navigation system, with information provided by all sensors, allows one to estimate the position, velocity, and attitude of the vehicle with respect to an inertial frame located in the surface station. The control component contains the algorithms that stabilize the state of the vehicle, in order to follow the commands given by the operator. The Beaglebone embedded computer executes different algorithms in order to achieve the desired movements.

6. Conclusions

This paper addressed the design of an open source-based control platform for an underwater exploration ROV. Previous experiences obtained through the development and

deployment of the ROV Visor3 have been compiled and analyzed, and shortcomings of using closed-source elements have been identified.

A major part of the selected hardware and software elements are distinguished by their open-source features; the main benefits are the high availability of information and extended support communities, differentiating them from closed-source solutions with limited information and, in some cases, poor support.

A modular approach for developing software architectures for robotic platforms has been presented. The results show that several frameworks facilitate, among others: functional expansion, simplification of fault diagnosis and repair processes, and reduction of development time. The software platforms selected and used in the design have comprehensive, available information and extensive support communities, factors that are key for a rapid and successful development. Furthermore, programming languages such as C++, JavaScript and LabVIEW® have proved to be adequate tools for developing modular software frameworks for robotic devices. Additionally, software development has been undertaken using Python, a high-level programming language that allows one to implement control algorithms that require complicated matrix computations

Finally, it has been shown how each component of the hardware architecture has a role in the control platform. The use of a compact hardware architecture can help one to increase reliability, maintenance and durability, without compromising the controllability of the vehicle.

Even though, there are observation-class ROVs with high-technology equipment such as the VideoRay such ROVs are not manufactured with open-source philosophies, therefore, they are difficult to repair or improve with changing operation requirements. In comparison with the OpenROV, which is another open source underwater vehicle solution; Visor3 has better instrumentation and the appropriate robustness to accomplish observation tasks in harsher environments.

Acknowledgements

This work was developed with the funding from the Fondo Nacional de Financiamiento para la Ciencia, la Tecnología y la Innovación, Francisco José de Caldas; the Colombian petroleum company, ECOPETROL; the Universidad Pontificia Bolivariana – Sede Medellín, UPB; the Universidad Nacional de Colombia – Sede Medellín, UNALMED; through the Strategic Program for the Development of Robotic Technology for Offshore Exploration of the Colombian Seabed, project 1210-531-30550, contract 0265 – 2013.

References

[1] Christ, R. and Wernli, R., The ROV Manual. A User guide for remotely operated vehicles, 2nd Ed. Elsevier, 2013. DOI: 10.1016/B978-0-08-098288-5.00033-6

[2] NORSOK, NORSOK STANDARD U-102, Norwegian Technology Centre Std., 2nd Ed., [Online]. 2012 [Date of reference: January 20 of 2015]. Available at: http://www.standard.no/petroleum

[3] Correa, J.C., Vásquez, R.E., Ramírez-Macías, J.A., Taborda, E.A., Zuluaga, C.A., Posada, N.L. and Londoño, J.M., Una arquitectura para el diseño conceptual de vehículos para exploración subacuática Ingeniería y Ciencia, 11(21), pp. 73-97, 2015. DOI: 10.17230/ingciencia.11.21.4

[4] Bai, Y. and Bai, Q., Subsea engineering handbook, in Subsea Engineering Handbook, Eds. Boston: Gulf Professional Publishing, 2010, pp. 3-25. DOI: 10.1016/B978-1-85617-689-7.10001-9

[5] Kim, T. and Yuh, J., Development of a real-time control architecture for a semi-autonomous underwater vehicle for intervention missions, Control Enginering Practice, 12(12), pp. 1521-1530, 2004. DOI: 10.1016/j.conengprac.2003.12.015

[6] Yoshida, H., Underwater Vehicles. In Tech, no. 29, ch. Fundamentals of Underwater Vehicle Hardware and Their Applications, 2009, pp. 557-582. DOI: 10.5772/6721

[7] Caccia, M., Casalino, G., Cristi, R. and Veruggio, G., Acoustic motion estimation and control for an unmanned underwater vehicle in a structured environment. Control Engineering Practice, 6(5), pp. 661-670, 1998. DOI: 10.1016/S0967-0661(98)00057-4.

[8] Caccia, M. and Veruggio, G., Guidance and control of a reconfigurable unmanned underwater vehicle. Control Engineering Practice, 8(1), pp. 21-37, 2000. DOI: 10.1016/S0967-0661(99)00125-2.

[9] Blain, M., Lemieux, S. and Houde, R., Implementation of a ROV navigation system using acoustic/Doppler sensors and Kalman ltering. In Proceedings of the OCEANS 2003, pp. 1255-1260, 2003. DOI: 10.1109/OCEANS.2003.178033.

[10] Loebis, D., Sutton, R., Chudley, J. and Naeem, W., Adaptive tuning of a Kalman lter via fuzzy logic for an intelligent AUV navigation system. Control Engineering Practice, 12(12), pp. 1531-1539, 2004. DOI: 10.1016/j.conengprac.2003.11.008.

[11] Lee, P.M. and Jun, B.H., Pseudo long base line navigation algorithm for underwater vehicles with inertial sensors and two acoustic range measurements. Ocean Engineering, 34(3-4), pp. 416-425, 2007. DOI: 10.1016/j.oceaneng.2006.03.011.

[12] Geng, Y. and Sousa, J., Hybrid derivative-free EKF for USBL/INS tightly-coupled integration in AUV. In Proceedings of the 2010 International Conference on Autonomous and Intelligent Systems (AIS), pp. 1-6, 2010. DOI: 10.1109/AIS.2010.5547035.

[13] Gutiérrez, L.B., Zuluaga, C.A., Ramírez, J.A., Vásquez, R.E., Flórez, D.A., Taborda, E.A. and Valencia, R.A., Development of an underwater remotely operated vehicle (ROV) for surveillance and inspection of port facilities, in Proceedings of the ASME IMECE2010, 2010. DOI: 10.1115/IMECE2010-38217

[14] Huang, H., Wan, L., Pang, Y. and Qin, Z., Control architecture of SY-II ship inspection remote operated vehicle, in Computer Science Service System (CSSS), 2012 International Conference on, pp. 2322-2325, 2012. DOI: 10.1109/CSSS.2012.576

[15] Wang, B., Wu, C. and Ge, T., A low cost compact control system for the Hippo ROV, Applied Mechanics and Materials, 190-191, pp. 627-633, 2012. DOI: 10.4028/www.scientific.net/AMM.190-191.627

[16] Wang, B., Wu, C. and Ge, T., Development of a remotely operated vehicle test-bed, Sensors & Transducers, [Online]. 153, pp. 45-52, 2013. Available at: http://www.sensorsportal.com/HTML/DIGEST/P_1218.htm

[17] Bonarini, A., Matteucci, M., Migliavacca, M. and Rizzi, D., R2P: An open source hardware and software modular approach to robot prototyping, Robotics and Autonomous Systems, 62(7), pp. 1073-1084, 2014. DOI: 10.1016/j.robot.2013.08.009

[18] Ahmed, Y., Yaakob, L. and Bong, S., Design of a new low cost ROV vehicle, Jurnal Teknologi 69(7), pp. 1-11, 2014. DOI: 10.11113/jt.v69.3262

[19] Sinha, P., Stiehl, K.R., Huo, E.S., Oyebode, O.A., Dokov, R.P., Chin, S.J., Price, R.E. and Larson, R.W., Design of a modular, compact, multi-role remotely operated vehicle for sheltered water operations, in IEEE OCEANS 2007, pp. 1-7, 2007. DOI: 10.1109/OCEANS.2007.4449392

[20] Busquets, J., Busquets, J.V., Tudela, D., Perez, F., Busquets-Carbonell, J., Barbera, A., Rodriguez, C., Garcia, A.J. and Gilabert, J., Low-cost AUV based on Arduino open source microcontroller board for oceanographic research applications in a collaborative long term deployment missions and suitable for combining with an USV as autonomous automatic recharging platform, in Autonomous Underwater Vehicles (AUV), 2012 IEEE/OES, pp. 1-10, 2012. DOI: 10.1109/AUV.2012.6380720

[21] Johansson, P. and Bernhard, J., Advanced control of a remotely operated underwater vehicle, Department of Electrical Engineering, Linköpings

universitet, Sweden, Tech. Rep., [Online]. 2012. Available at: http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-79364
[22] Fittery, A., Mazumdar, A., Lozano, M. and Asada, H., Omni-Egg: A smooth, spheroidal, appendage free underwater robot capable of 5 dof motions, in IEEE OCEANS 2012, pp. 1-5, 2012. DOI: 10.1109/OCEANS.2012.6404986
[23] Hassan, O.I.H., Black-box identification and control for autonomous underwater vehicles, PhD. Dissertation, School of Engineering and Information Technology, The University of New South Wales, Australian Defence Force Academy, [Online]. 2013. Available at: http://handle.unsw.edu.au/1959.4/52888
[24] Portugal, D. and Rocha, R.P., Distributed multi-robot patrol: A scalable and fault-tolerant framework, Robotics and Autonomous Systems, 61(12), pp. 1572-1587, 2013. DOI: 0.1016/j.robot.2013.06.011
[25] Faugel, H. and Bobkov, V., Open source hard- and software: Using Arduino boards to keep old hardware running, Fusion Engineering and Design, 88(6-8), pp. 1276-1279, 2013. DOI: 10.1016/j.fusengdes.2012.12.005
[26] Grimmett, R., Mastering beagle bone robotics. Packt Publishing, 2014
[27] Quintián, H., Calvo, J.L. and Fontenla, O., Aplicación de un robot comercial de bajo coste en tareas de seguimiento de objetos. DYNA, 79(175), pp. 24-33, 2012. ISSN 2346-2183.
[28] Álvarez-Chavarría, J.S., Jiménez-Builes, J.A. and Ramírez-Patiño, J.F., Design cycle of a robot for learning and the development of creativity in engineering. DYNA, 78(170), pp. 51-58, 2011.
[29] VanCott, R., Wilbur, B.M., Chryssostomidis, C., Soroka, M. and Shroyer, K., STEM education through open hardware at MIT sea grant, Open Hardware Journal, [Online]. 1, pp. 9-15, 2011. Available at: https://archive.org/details/OpenHardwareJournal
[30] Schneider, D., Build your own robosub [Hands On], IEEE Spectrum, 48(9), pp. 24-26, 2011. DOI: 0.1109/MSPEC.2011.5995890
[31] Tirado, D., Asistencia a la investigación en la educación e implementación del hardware (joystick, IMU, profundímetro, interfaz con actuadores), para el proyecto ROV. Trabajo de grado, Facultad de Ingeniería Eléctrica/Electrónica, Universidad Pontificia Bolivariana, Medellín, Colombia, 2008.
[32] Franco, J., Frameworks para sistemas de control embebidos de tiempo real. Trabajo de grado, Especialización en Automática, Universidad Pontificia Bolivariana, Medellín, Colombia, 2007.
[33] Nadvornik, J. and Smutny, P., Remote control robot using android mobile device, in Control Conference (ICCC), 15th International Carpathian, pp. 373-378, 2014. DOI: 10.1109/CarpathianCC.2014.6843630
[34] Speers, A., Forooshani, P.M., Dicke, M. and Jenkin, M., Lightweight tablet devices for command and control of ROS-enabled robots, in Advanced Robotics (ICAR), 16th International Conference on, 2013. DOI: 10.1109/ICAR.2013.6766481
[35] Bertolotti, I.C. and Hu, T., Modular design of an open-source, networked embedded system, in Computer Standards & Interfaces, 37, pp. 41-52,2015. DOI: 10.1016/j.csi.2014.05.004.
[36] Gerkey, B.P., Vaughan, R.T. and Howard, A., The player/stage project: Tools for multi-robot and distributed sensor systems, in: Proceedings of the 11th International Conference on Advanced Robotics, 2003, pp. 317-323.
[37] Bruyninckx, H., Open robot control software: The OROCOS project, Robotics and Automation, 2001. Proceedings 2001 ICRA. IEEE International Conference on, 3, pp.2523-2528 2001. DOI: 10.1109/ROBOT.2001.933002
[38] Quigley, M., Conley, K., Gerkey, B.P., Faust, J., Foote, T., Leibs, J., Wheeler, R. and Ng, A.Y., ROS: An open-source robot operating system, in: ICRA Workshop on Open Source Software, 2009.
[39] Huang, A., Olson, E. and Moore, D., LCM: Lightweight communications and marshalling, in: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2010, pp. 4057-4062. DOI: 10.1109/IROS.2010.5649358
[40] Becker, L.B., Gergeleit, M. and Nett, E., Approach for implementing object-oriented real-time models on top of embedded targets, in OMER-2 – Workshop on O-O Modeling of Embedded RT Systems, 2001.
[41] Tilkov, S. and Vinoski, S., Node.JS: Using JavaScript to build high-performance network programs, internet computing, IEEE, 14(6), pp. 80-83, 2010. DOI: 10.1109/MIC.2010.145
[42] Ojamaa, A. and Duuna, K., Assessing the security of Node.JS platform, 2012 International Conference for Internet Technology and Secured Transactions, IEEE, pp. 348-355, 2012.
[43] National Instruments™, Benefits of Programming Graphically in NI LabVIEW, [Online]. 2013. Available at: http://www.ni.com/white-paper/14556/en/pdf.
[44] Roberts., G.N., Trends in marine control systems, Annual Reviews in Control, 32(2), pp. 263-269, 2008. DOI: 10.1016/j.arcontrol.2008.08.002
[45] Chyba, M., Haberkorn, R.S., and Choi, S., Design and implementation of time efficient trajectories for autonomous underwater vehicles. Ocean Engineering, 35(1), pp. 63-76, 2008. DOI: 10.1016/j.oceaneng.2007.07.007

L.M. Aristizábal, IEO, received a BSc. in Electronic Engineering in 2013 from the Universidad Pontificia Bolivariana (UPB), Medellín, Colombia. He is currently studying for a Master's degree in the field of Engineering with emphasis in Automation at the UPB. His research interests are hardware and software architectures, robotics, and automation systems.
ORCID: 0000-0003-0607-9613

S. Rúa, MSc., received a BSc. in Electronic Engineering in 2013, and an MSc. in Engineering with emphasis in Automation in 2015, both from the Universidad Pontificia Bolivariana (UPB), Medellin, Colombia. He is currently a PhD student in Engineering at the UPB, in the area of control and navigation systems. His research interests are control of dynamic systems, navigation systems, machine learning, and embedded systems.
ORCID: 0000-0003-0067-8316

C.E. Gaviria, IEO, received a BSc. in Electronic Engineering in 2011 from the Universidad de Antioquia, Medellin, Colombia. He is currently an MSc. student in Engineering at the UPB, in the area of control and automation systems.
ORCID: 0000-0003-2038-581X

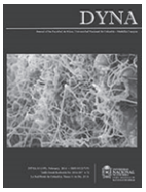
S.P. Osorio, IEO, received a BSc. in Electronic Engineering in 2014 from the Universidad Pontificia Bolivariana (UPB), Medellín, Colombia. She is currently studying for a Master's degree in the field of Engineering with emphasis in Automation at the UPB.
ORCID: 0000-0003-1985-499X

C.A. Zuluaga, MSc., received a BSc. in Electronic Engineering in 1999, and an MSc. in Engineering with emphasis in Automation in 2006, both from the Universidad Pontificia Bolivariana (UPB), Medellin, Colombia. He is currently professor in the Department of Electrical Engineering at the UPB. He currently coordinates the graduate programs in automation and participates in several research projects related to control engineering. For several years his research work has been related to autonomous underwater vehicle control (AUV) and remotely operated vehicles (ROV).
ORCID: 0000-0002-1773-767X

N.L. Posada, MSc., received a BSc. Eng in Instrumentation and Control Engineering in 2000 from the Politecnico Colombiano Jaime Isaza Cadavid, and an MSc. in Engineering with emphasis in Automation in 2010 from the Universidad Pontificia Bolivariana (UPB), Medellín, Colombia. Norha is currently professor in the Department of Mechanical Engineering at the UPB, in the area of systems and control. Her research interests are instrumentation, process automation, and state observer design.
ORCID: 0000-0002-7607-0756

R.E. Vásquez, PhD., received a BSc. in Mechanical Engineering in 2002, and an MSc. in Engineering with emphasis in Automation in 2007, both from the Universidad Pontificia Bolivariana (UPB), Medellin, Colombia. He received his PhD. in Mechanical Engineering from the University of Florida, USA in 2011. Rafael is currently professor in the Department of Mechanical Engineering at the UPB, in the area of dynamics, systems and control. His research interests are theory of mechanisms; robotics; design, analysis, and control of dynamic systems; tensegrity systems; new technologies for energy harvesting; and engineering education. He is a member of the American Society of Mechanical Engineers (ASME) since 2005.
ORCID: 0000-0003-4871-8823

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS



DYNA
http://dyna.medellin.unal.edu.co/

Probabilistic seismic hazard analysis using a new ground motion intensity measure

Robespierre Chávez-López [a] & Edén Bojórquez-Mora [b]

[a] *Facultad de Ingeniería, Universidad Autónoma de Sinaloa, Culiacán, México. robespierre_chavez@hotmail.com, est.robespierre_chavez@uas.edu.mx*
[b] *Facultad de Ingeniería, Universidad Autónoma de Sinaloa, Culiacán, México. eden_bmseg@hotmail.com, eden@uas.edu.mx*

Received: April 16th, 2015. Received in revised form: July 29th, 2015. Accepted: August 18th, 2015.

Abstract
The main objective of this work is to compute the probabilistic seismic hazard analysis for a region of Mexico using a new ground motion intensity measure which is based on the spectral acceleration and a parameter proxy of the spectral shape named N_p. The motivation of using this new ground motion intensity measure is because recently it has demonstrated its potential in predicting the response of buildings subjected to earthquakes. In fact, it was demonstrated that intensity measures based on N_p are more efficient compared with other parameter of the literature. It is important to mention that this is the first time that a probabilistic seismic hazard analysis is performed using this new intensity measurement.

Keywords: probabilistic seismic hazard analysis; ground motion intensity measure, spectral acceleration.

Análisis de peligro sísmico probabilístico usando una nueva medida de intensidad sísmica

Resumen
El objetivo principal de este trabajo es realizar un análisis de peligro sísmico probabilístico para una región de México utilizando una nueva medida de intensidad sísmica basada en la aceleración espectral y en un parámetro para caracterizar la forma espectral llamado N_p. La motivación para usar esta nueva medida de intensidad sísmica es porque recientemente se ha demostrado su potencial en la predicción de la respuesta de edificaciones sometidas a sismos. De hecho, se ha demostrado que las medidas de intensidad basadas en N_p son más eficientes en comparación con otros parámetros propuestos en la literatura. Es importante decir que esta es la primera vez que se realiza un análisis de peligro sísmico probabilístico utilizando esta nueva medida de intensidad sísmica.

Palabras clave: análisis de peligro sísmico probabilístico; medida de intensidad sísmica; aceleración espectral.

1. Introducción

El análisis de peligro sísmico probabilístico es un procedimiento para estimar la probabilidad de que un determinado nivel de intensidad sísmica sea excedido en un sitio y durante un periodo de tiempo. Debido a la importancia del análisis de peligro sísmico probabilístico, varios estudios de este tipo han sido desarrollados para diferentes sitios alrededor del mundo [1-5]. El uso de conceptos probabilísticos ha permitido considerar explícitamente las incertidumbres en el tamaño, localización y tasa de recurrencia sísmica, así como la variación de las características del movimiento sísmico con el tamaño y la ubicación del terremoto. El análisis de peligro sísmico probabilístico permite identificar, cuantificar y combinar en una manera racional estas incertidumbres, proporcionando una evaluación más completa del peligro sísmico. Comúnmente, en los códigos de diseño sísmico, el movimiento sísmico es expresado en términos de una medida de la intensidad. Actualmente, las medidas de intensidad más utilizadas para el análisis de peligro sísmico probabilístico son la aceleración máxima del suelo (A_{ms}) y la aceleración espectral en el primer modo de vibración ($Sa(T_1)$). Sin embargo, varios estudios han demostrado la limitación de

estos dos parámetros para representar el potencial del movimiento de suelo durante un terremoto, de tal manera que la evaluación del peligro sísmico requiere el uso de medidas de intensidad más sofisticadas.

Como se mencionó anteriormente, para poder evaluar el peligro sísmico en un sitio es de vital importancia contar con un parámetro que pueda describir la severidad de un movimiento sísmico, uno de los objetivos principales de dicho parámetro que se conoce como medida de intensidad sísmica, es estar íntimamente relacionado con la demanda sísmica, lo cual permite reducir las incertidumbres en la respuesta estructural de edificios. Algunas características de dicha medida son su relación con la respuesta estructural y su capacidad para desacoplar las incertidumbres sismológicas y estructurales. Debido a la gran importancia que representa contar con una medida de intensidad sísmica apropiada, se han desarrollado varios estudios orientados en encontrar un parámetro que refleje el potencial destructivo de un movimiento sísmico [6-9].

La evidencia existente muestra que los esfuerzos en plantear parámetros para definir una medida de intensidad sísmica apropiada se concentran en la forma espectral debido a su relación con la respuesta estructural. Lo anterior indica que el primer paso es tratar de encontrar un parámetro capaz de representar con la mejor aproximación la forma espectral. Por ello, utilizando este razonamiento, recientemente una medida de intensidad sísmica conocida como I_{Np} fue propuesta por Bojórquez y Iervolino [10] cuyo objetivo es capturar con buena aproximación la forma espectral, a partir de un parámetro conocido como N_p. Ellos observaron que medidas escalares y vectoriales basadas en $Sa(T_1)$ y en la variable N_p, presentan una mejor relación con la respuesta sísmica no lineal de las estructuras. La motivación para usar esta nueva medida de intensidad sísmica (I_{Np}) para realizar el análisis de peligro sísmico para una región de México es porque Buratti [11] demostró el gran potencial de la medida propuesta por Bojórquez y Iervolino [10] y su gran eficiencia, además concluyó que una medida de intensidad sísmica eficiente resulta, al mismo tiempo, suficiente.

2. Sismología de México

La República Mexicana está situada en una de las regiones sísmicamente más activas del mundo, enclavada dentro del área conocida como el Cinturón Circumpacífico donde se concentra la mayor actividad sísmica del planeta [12].

La alta sismicidad en el país, es debido principalmente a la interacción entre las placas de Norteamérica, la de Cocos, la del Pacífico, la de Rivera y la del Caribe, así como a fallas locales que corren a lo largo de varios estados aunque estas últimas menos peligrosas. La Placa Norteamericana se separa de la del Pacífico pero roza con la del Caribe y choca contra las de Rivera y Cocos, de aquí la incidencia de sismos.

El estudio de la actividad sísmica de México comenzó a principios del siglo XX; sin embargo, los antecedentes históricos de grandes sismos del país fueron registrados en un gran número de documentos. En 1910 se inauguró la red sismológica mexicana y desde esa fecha hasta nuestros días se ha mantenido una observación continua de los temblores cuyos registros se conservan en la Estación Sismológica de Tacubaya y otras instalaciones del Instituto de Geofísica de la UNAM, encargada de operar el Servicio Sismológico Nacional -SSN- y su red de estaciones sismológicas. El SSN reporta en el país, en promedio, la ocurrencia de 4 sismos por día de magnitud M > 3.0.

Existe otro grupo de trabajo en el Centro de Investigaciones y de Educación Superior de Ensenada, B.C. (CICESE), que enfoca su estudio entre otros aspectos, a la actividad sísmica asociada tanto al Golfo de California como a la falla de San Andrés, al igual que la Red Sismológica del Noroeste (RESNOR). Adicionalmente, algunas instituciones de enseñanza superior en el interior del país, hacen estudios de sismicidad regional. Los diferentes grupos mantienen comunicación para dar a conocer sus avances.

3. Peligro sísmico

El análisis de peligro sísmico probabilístico considera todos los posibles eventos sísmicos, junto con sus probabilidades asociadas de ocurrencia, con el fin de encontrar el nivel de intensidad de movimiento del suelo que sea igual o mayor que cierta tasa de excedencia [13]. En esta sección se muestra la información requerida para aplicación de esta metodología.

3.1. Sitios en estudio

En este estudio fueron considerados cuatro sitios para realizar los análisis de peligro sísmico probabilístico, estos sitios se encuentran ubicados en el estado de Sinaloa, que es una región del noroeste de México de sismicidad moderada a alta. En la Fig. 1 se muestra la localización de cada uno de estos sitios, que se encuentran en las cuatro principales ciudades del Estado de Sinaloa que son Los Mochis, Guasave, Culiacán y Mazatlán.

3.2. Catalogo sísmico homogeneizado

El catalogo sísmico utilizado para la regionalización sismo tectónica de México fue sugerido por Zúñiga y aplicado en un estudio reciente [14]. Es un catálogo a primera aproximación de terremotos que ocurrieron en México entre

Figura 1. Ubicación de los lugares en estudio.
Fuente: https://www.google.com.mx/maps/@23.8534243,-108.7385837,678294m/data=!3m1!1e3.

1899 y 2007, y comprende 60,606 eventos. Para esta compilación se tomaron en cuenta catálogos sísmicos previamente publicados [15-20]. Además fueron consultados catálogos producidos por reconocidas agencias como: El Centro Internacional de Sismología (ISC), Servicio Geológico de los E.U. (PDE reportados por NEIC), La Administración Nacional Oceánica y Atmosférica de Estados Unidos (NOAA), el Servicio Sismológico Nacional (SSN) y la Red Sísmica del Noroeste de México (RESNOM). Se documentaron sismos históricos y estudios paleo-sismológicos. La mayoría de los eventos con magnitud superior a 7.0 fueron estudiados individualmente por varios autores, de estos estudios, se incluyeron las estimaciones de la magnitud consideradas como las más fiables.

Se consideraron las localizaciones y magnitudes para datos anteriores a 1988 en la siguiente prioridad: 1) Estudios individuales para sismos grandes, 2) Catalogo ISC, 3) Catalogo PDE, 4) Catalogo SSN y RESNOM, 5) Catálogos históricos y otros catálogos, debido a la precisión en estos datos. Para datos posteriores a 1988 se consideraron básicamente los datos de localización y magnitud del SSN y la RESNOM; sin embargo, para los sismos mayores que 6.5, únicamente las magnitudes fueron tomadas del catálogo CMT [21].

En general, el catalogo sísmico se compilo usando estimaciones de magnitud y/o momento, así como de localizaciones consideradas óptimas, ya que se tomaron en cuenta los errores de localización [14]. Además, se homogeneizo la magnitud para todos los eventos.

El catalogo instrumental se complementa con los terremotos históricos más importantes reportados en una compilación reciente de la historia sísmica de México [22]. Se estimó una magnitud basada sobre datos de intensidad [23] para los eventos más grandes y estos datos fueron usados para definir los límites de las regiones sismotectónicas descritas a continuación, y en la estimación de la magnitud máxima aproximada de las provincias. Los ejemplos más claros son La Faja Volcánica Mexicana y la costa sureste del Golfo de México ya que son regiones que han mostrado una historia de grandes eventos históricos, pero el catalogo instrumental es relativamente pobre. La discusión de importantes sismos históricos se incorpora a continuación en la descripción y definición de las regiones sismotectónicas.

3.3. Selección de regiones sismotectónicas

Los catálogos instrumentales e históricos descritos anteriormente fueron utilizados por Zúñiga et al. [24] como base para la división del territorio de la República Mexicana en unidades o regiones sismotectónicas. Las zonas en que fue dividido México se presentan en la Fig. 2.

Esta regionalización no es de ningún modo única; sin embargo, proporciona una división coherente y sistematizada que incorpora la mayor parte de los conocimientos sobre características generales de los sismos en diferentes partes del país. Además considera el potencial destructivo de los eventos más importantes que han ocurrido en el pasado. Esta regionalización se destina a ser utilizada como una herramienta en la definición de una zonificación de primer orden del peligro sísmico de México [24].

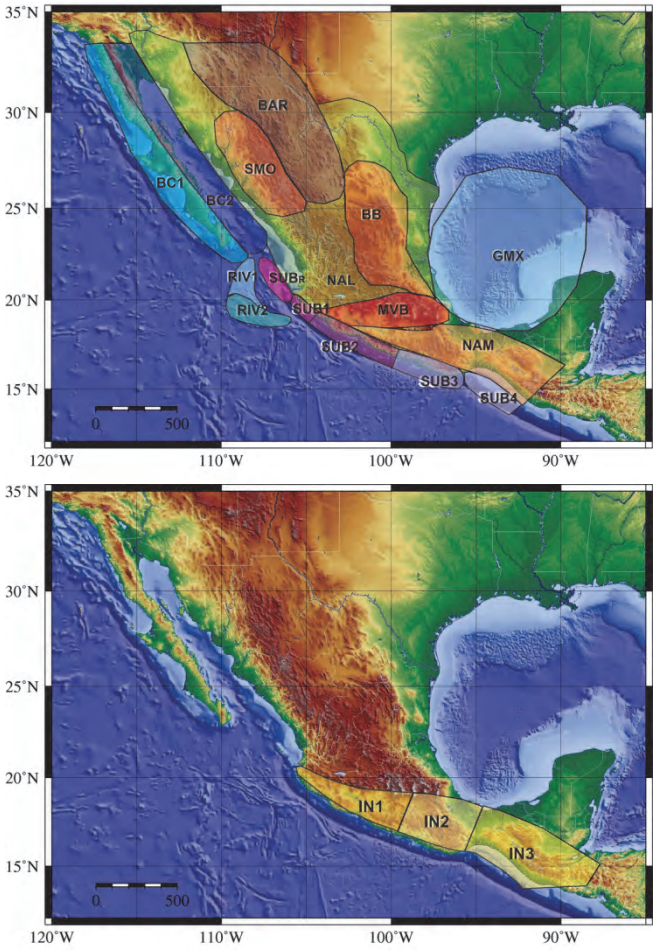


Figura 2. Zonas sismotectónicas de México. Arriba: Regiones correspondientes a sismos de profundidad somera. Abajo: Regiones de sismos de profundidad intermedia.
Fuente: Zúñiga et al., 1997.

3.4. Selección de relaciones de atenuación sísmica

El Servicio Sismológico Nacional en México inició sus operaciones en 1910, con la información recabada se han podido desarrollar varias relaciones de atenuación para la zona centro del país [25-28], pero en caso contrario, para la parte noroeste de México no se cuenta con la información suficiente para desarrollar una relación de atenuación. Por lo tanto, es necesario seleccionar una relación de atenuación de las existentes en la literatura, extraída de datos estadísticamente significativos y que comprenda un amplio intervalo de magnitud y distancia.

Para seleccionar la relación de atenuación se consideraron principalmente dos criterios: a) que debería ser derivada de una extensa base de datos y de preferencia desarrollada para sitios localizados con características sismotectónicas similares, b) que la variable independiente debería ser dada en términos de la aceleración máxima del suelo y ordenadas espectrales para un amplio intervalo de periodos de vibración. Considerando los criterios antes mencionados, se optó por seleccionar la relación de atenuación de Boore y Atkinson [29], la cual se define en la siguiente expresión:

$$\ln Y = F_M(M) + F_D(R_{JB}, M) + F_S(V_{S30}, R_{JB}, M) + \varepsilon\sigma_T \quad (1)$$

En la ec. (1), F_M, F_D y F_S representan el escalamiento de la magnitud, función de distancia y amplificación del sitio, respectivamente. M es la magnitud momento, R_{JB} es la distancia de Joyner-Boore (definida como la distancia más cercana a la superficie de proyección de la falla, que es aproximadamente igual a la distancia epicentral para eventos de $M \leq 6$), y V_{S30} es la velocidad promedio de las ondas de cortante en los primeros 30 metros de profundidad en el sitio. ε es el número fraccional de desviaciones estándar de un solo valor predicho de $ln\ Y$ lejos del valor medio de $ln\ Y$ (e. g., ε = -1.5, seria 1.5 desviaciones estándar menor que el valor medio). Todos los términos, incluyendo el coeficiente σ_T, son periodo dependientes. σ_T es calculado usando la ec. (2).

$$\sigma_T = \sqrt{\sigma^2 + \tau^2} \quad (2)$$

donde σ es la incertidumbre aleatoria intra-evento y τ es la incertidumbre aleatoria inter-evento.

3.5. Análisis de peligro sísmico probabilístico para I_{Np}

Las medidas escalares tradicionales de intensidad sísmica (e.g. $Sa(T_1)$, A_{ms}) tienen ciertas deficiencias como la insuficiencia o limitada eficiencia, las cuales han sido superadas por otras medidas de intensidad sísmica tanto escalaras como vectoriales. Una medida de intensidad sísmica vectorial basada en $Sa(T_1)$ y en combinación con el parámetro N_p ya ha sido investigada y se describirá enseguida. Bojórquez *et al*. [30] han explorado el potencial del vector $<Sa, N_p>$ en comparación con otras medidas de intensidad.

El vector $<Sa, N_p>$ resulta mucho más eficiente que las medidas de intensidad frecuentemente utilizadas. Sin embargo, el uso de una medida de intensidad vectorial con la cual se pretenda desarrollar análisis probabilístico de la demanda sísmica de una estructura, requerirá la estimación de distribuciones condicionales de N_p dada $Sa(T_1)$. Por otro lado, el análisis probabilístico de la demanda sísmica se facilita por medio del uso de medidas de intensidad escalar, y la relación con la respuesta estructural también se vuelve más clara. Por esta razón, Bojórquez y Iervolino [10] propusieron una medida de intensidad sísmica escalar basada en los parámetros $Sa(T_1)$ y N_p con características similares a la medida de intensidad propuesta por Córdova *et al.* [31], la cual se define de la siguiente manera:

$$I_{Np} = Sa(T_1) \cdot N_p{}^{\alpha} \quad (3)$$

$$N_p = \frac{Sa_{avg}(T_1 \ldots T_N)}{Sa(T_1)} \quad (4)$$

En la ec. (3), I_{Np} es la medida de intensidad sísmica escalar, y el valor α es un parámetro que debe ser calibrado, en función de la estructura y el parámetro de demanda sísmica seleccionado. Buratti [11] demostró que I_{Np} resulta más eficiente que un gran de número de medidas de intensidad sísmica encontradas en la literatura [6,8,9,31-37].

En la ec. (3) se puede observar lo siguiente: a) la aceleración espectral en el primer modo de vibrar o el periodo de la estructura es un caso particular de I_{Np}, y este ocurre cuando α es igual a cero; b) la media geométrica $Sa_{avg}(T_1 \ldots T_N)$ representa el caso particular de I_{Np} cuando $\alpha=1$; y c) la medida de intensidad vectorial propuesta por Córdova et al. [31], denominada $<Sa, R_{T1,T2}>$, también corresponde a un caso particular cuando solo dos puntos del espectro son tomados en consideración, por ejemplo, T_1 y T_N. Bojórquez y Iervolino [10] sugieren que los valores óptimos de α están en un intervalo que va desde cero hasta uno, lo cual significa dotar a las ordenadas espectrales con un nivel diferente de importancia cuando se analice I_{Np} más allá del primer modo.

Debido a la posibilidad actual de calcular el peligro sísmico que es de gran utilidad para cualquier medida de intensidad propuesta, en esta sección se muestra como puede ser empleado I_{Np} para análisis de peligro sísmico probabilístico con herramientas actualmente disponibles para otras medidas de intensidad sísmica. La formulación general para el cálculo de peligro puede obtenerse de Bojórquez y Iervolino [10].

$$E\left[\ln\left(I_{Np}\right)\right] = (1-\alpha)E\left\{\ln\left[Sa(T_1)\right]\right\} + \frac{\alpha}{N}\sum_{i=1}^{N} E\left\{\ln\left[Sa(T_1)\right]\right\} \quad (5)$$

$$\begin{aligned} Var\left[\ln\left(I_{Np}\right)\right] &= \alpha^2 Var\left\{\ln\left[Sa_{avg}(T_1 \ldots T_N)\right]\right\} \\ &+ (1-\alpha)^2 Var\left\{\ln\left[Sa(T_1)\right]\right\} \\ &+ 2\alpha(1-\alpha)\rho_{\ln[Sa_{avg}(T_1 \ldots T_N)],\ln[Sa(T_1)]}\sigma_{\ln[Sa_{avg}(T_1 \ldots T_N)]}\sigma_{\ln[Sa(T_1)]} \end{aligned} \quad (6)$$

$$Var\left\{\ln\left[Sa_{avg}(T_1 \ldots T_N)\right]\right\} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left[\rho_{\ln[Sa(T_i)],\ln[Sa(T_j)]}\sigma_{\ln[Sa(T_i)]}\sigma_{\ln[Sa(T_j)]}\right] \quad (7)$$

$$\rho_{\ln[Sa_{avg}(T_1 \ldots T_N)],\ln[Sa(T_1)]} = \frac{\sum_{i=1}^{N}\rho_{\ln[Sa(T_i)],\ln[Sa(T_1)]}\sigma_{\ln[Sa(T_i)]}}{\sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N}\left[\rho_{\ln[Sa(T_i)],\ln[Sa(T_j)]}\sigma_{\ln[Sa(T_i)]}\sigma_{\ln[Sa(T_j)]}\right]}} \quad (8)$$

donde $E[(\,)]$ representa la esperanza de un valor, en las ec. (7)-(8), $\rho_{\ln[Sa(T_i)],\ln[Sa(T_j)]}$ es la correlación y puede obtenerse con varias formulaciones propuestas en la literatura, en este estudio se obtiene mediante la correlación de Baker y Jayaram [38] que se muestra en la ec. (9):

$$\begin{aligned} si\ T_{\max} < 0.109 \quad & \rho_{\ln[Sa(T_i)],\ln[Sa(T_j)]} = C_2 \\ si\ T_{\min} > 0.109 \quad & \rho_{\ln[Sa(T_i)],\ln[Sa(T_j)]} = C_1 \\ si\ T_{\max} < 0.2 \quad & \rho_{\ln[Sa(T_i)],\ln[Sa(T_j)]} = \min(C_2, C_4) \\ de\ otra\ manera \quad & \rho_{\ln[Sa(T_i)],\ln[Sa(T_j)]} = C_4 \end{aligned} \quad (9)$$

En la ec. (9) C_1, C_2 y C_4 son calculados mediante la ec. (10):

$$C_1 = 1 - \cos\left(\frac{\pi}{2} - 0.366\ln\left(\frac{T_{\max}}{\max(T_{\min}, 0.109)}\right)\right)$$

$$C_2 = \begin{cases} 1 - 0.105\left(1 - \frac{1}{1+e^{100T_{\max}-5}}\right)\left(\frac{T_{\max}-T_{\min}}{T_{\max}-0.0099}\right) & si\ T_{\max} < 0.2 \\ 0 & de\ lo\ contrario \end{cases} \quad (10)$$

$$C_3 = \begin{cases} C_2 & si\ T_{\max} < 0.109 \\ C_1 & de\ lo\ contrario \\ C_4 = C_1 + 0.5\left(\sqrt{C_3} - C_3\right)\left(1 + \cos\left(\frac{\pi T_{\min}}{0.109}\right)\right) & \end{cases}$$

Finalmente, porque los valores de $\ln[Sa(T_1)]$ son conjuntamente Gausiano, las ec. (5)-(6) pueden ser obtenidas de relaciones de atenuación existentes, y estas ecuaciones son suficientes para describir la distribución completa de I_{Np}, y con esto realizar el análisis de peligro sísmico probabilístico como se hace para un solo valor de aceleración espectral.

El resultado de un análisis de peligro sísmico probabilístico puede ser expresado en muchas diferentes formas. Todas implican algunos niveles de cálculos probabilísticos a combinar las incertidumbres en tamaño del terremoto, localización, frecuencia, y efectos para estimar el peligro sísmico. Una aproximación común implica el desarrollo de curvas de peligro sísmico, que indican la probabilidad anual de excedencia de diferentes valores de un parámetro de movimiento sísmico seleccionado. Las curvas de peligro sísmico pueden entonces ser usadas para calcular la probabilidad de excedencia de un parámetro de movimiento sísmico seleccionado en un periodo específico de tiempo.

Las curvas de peligro sísmico pueden ser obtenidas para zonas fuentes individuales y combinadas para expresar el peligro total en un sitio particular. El concepto básico de los cálculos requeridos para desarrollar curvas de peligro sísmico es bastante simple. La probabilidad de exceder un valor particular, y^*, de un parámetro de movimiento de suelo, Y, es calculado para un posible terremoto en un lugar posible de la fuente y entonces multiplicado por la probabilidad de que un terremoto de magnitud particular podría ocurrir en esa ubicación particular. El proceso es entonces repetido para todas las magnitudes y localizaciones posibles con la probabilidad de cada suma. La formulación básica para el cálculo de las curvas de peligro sísmico se muestra en la ec. (11).

$$\lambda_{y^*} = \sum_{i=1}^{N_S}\sum_{j=1}^{N_M}\sum_{k=1}^{N_R} \nu_i P\left[Y > y^* \mid m_j, r_k\right] P\left[M = m_j\right] P\left[R = r_k\right] \quad (11)$$

donde λ_{y^*} es la tasa anual de excedencia, ν_i es la tasa media anual de excedencia de una magnitud asociada a una fuente sísmica, $P\left[Y > y^* \mid m_j, r_k\right]$ es la probabilidad que algún parámetro del movimiento del suelo Y estimado para un sismo de una magnitud m_j y una distancia r_k, exceda cierto valor y^*, $P\left[M = m_j\right]$ es la probabilidad que una magnitud M sea igual a una magnitud m_j y $P\left[R = r_k\right]$ es la probabilidad que una distancia R sea igual a una distancia r_k.

Las curvas de peligro sísmico pueden ser fácilmente combinadas con el modelo de Poisson para estimar probabilidades de excedencia en intervalos de tiempo finito. La probabilidad de excedencia de y^* en un periodo de tiempo T es dado por la ec. (12):

$$P\left[Y_T > y^*\right] = 1 - e^{-\lambda_{y^*}T} \quad (12)$$

4. Estimación del peligro sísmico

Para el análisis de peligro sísmico se representó la distribución de distancias de la fuente al sitio por un histograma. Para ello se divide la zona fuente en una fina malla de 0.1 grado de separación entre puntos coordenados, conocido el intervalo total de distancia se agrupa en intervalos de 20 km, para caracterizar la distribución de distancias de la fuente al sitio, se determina que tantos segmentos forman parte de cada intervalo de distancias. Las ordenadas del histograma normalizado representan la frecuencia relativa que deberá ser igual a la probabilidad si un infinito número de segmentos fue usado, pero que es una aproximación a la probabilidad en este caso. De la misma manera se realiza para cada fuente sismotectónica.

Con fin de trazar espectros de peligro uniforme para periodos de retorno de 475 y 2475 años se obtienen curvas de peligro sísmico considerando un amplio intervalo de periodos que van de 0.2 a 3 segundos en intervalos de 0.2 segundos.

Los resultados se clasificaron según el tipo de suelo de acuerdo a la velocidad promedio de las ondas de cortante (V_{S30}), las cuales son un indicador del tipo de suelo en un sitio. Es decir, si se desea diseñar una estructura se debe realizar un estudio de mecánica de suelos en el sitio de interés para poder conocer la V_{S30}, una vez conocida la V_{S30} es posible trazar las curvas de peligro sísmico y el espectro de peligro uniforme para un periodo de retorno deseado, para un sitio en específico. La Tabla 1 muestra la relación entre la velocidad de ondas de cortante y el tipo de suelo.

La ventaja de los resultados obtenidos es que las curvas están asociadas a distintos valores de la velocidad de ondas de cortante, por lo tanto bastará con seleccionar la curva adecuada al tipo del suelo del sitio. En las Fig. 3-6 se muestran las curvas de peligro sísmico para los cuatro sitios en estudio para un periodo estructural de un segundo y distintas velocidades promedio de ondas de cortante. Se observa como a medida que la velocidad promedio de ondas de cortante disminuye, se incrementa I_{Np} para una tasa de excedencia en particular. Por ejemplo, para un periodo de retorno de 475 años, la Tabla 2 indica los valores de I_{Np} para distintas velocidades de ondas de cortante para el sitio FIC-UAS.

Tabla 1.
Clasificación de suelos NEHRP.

Tipo de suelo	Descripción	Velocidad promedio de las ondas de cortante (V_{S30}) (m/s)
A	Roca dura	> 1500
B	Roca	760 - 1500
C	Suelo muy denso o roca blanda	360 - 760
D	Suelo rígido	180 - 360
E	Suelo blando	< 180

Fuente: NEHRP.

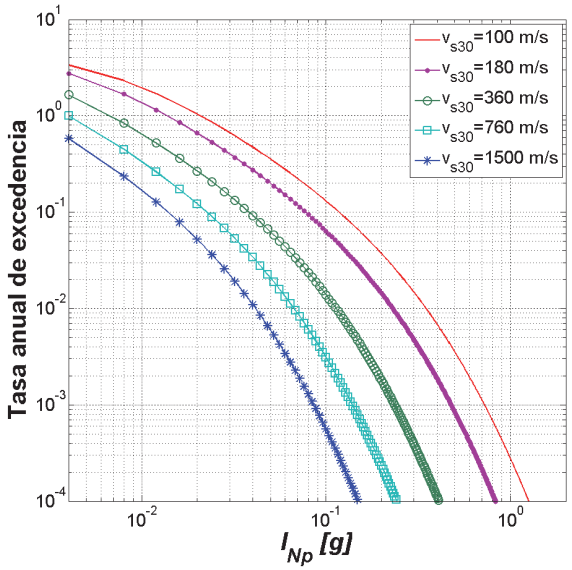


Figura 3. Curvas de peligro sísmico para un periodo estructural de un segundo para distintas velocidades de ondas de cortante para el sitio FIC-UAS.
Fuente: Propia.

Tabla 2.
Valores de la intensidad en términos de I_{Np} para el sitio FIC-UAS y periodo estructural de un segundo para distintas velocidades de ondas de cortante para un periodo de retorno de 475 años.

Velocidad de ondas de cortante (V_{S30})	I_{Np} [g]
100	0.59
180	0.39
360	0.19
760	0.11
1500	0.07

Fuente: Propia.

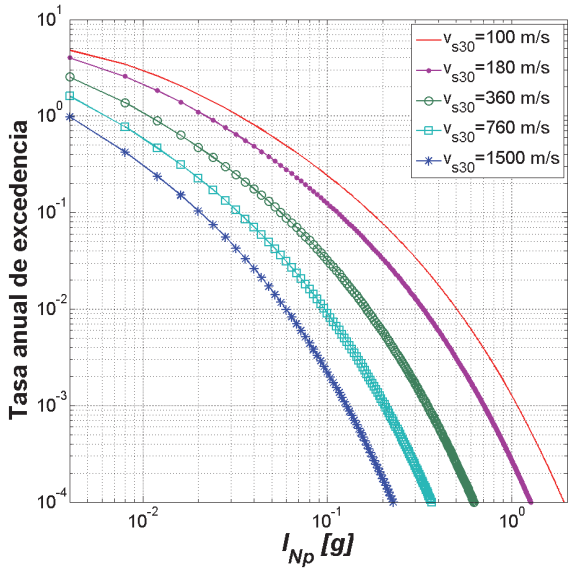


Figura 4. Curvas de peligro sísmico para un periodo estructural de un segundo para distintas velocidades de ondas de cortante para el sitio FIM-UAS.
Fuente: Propia.

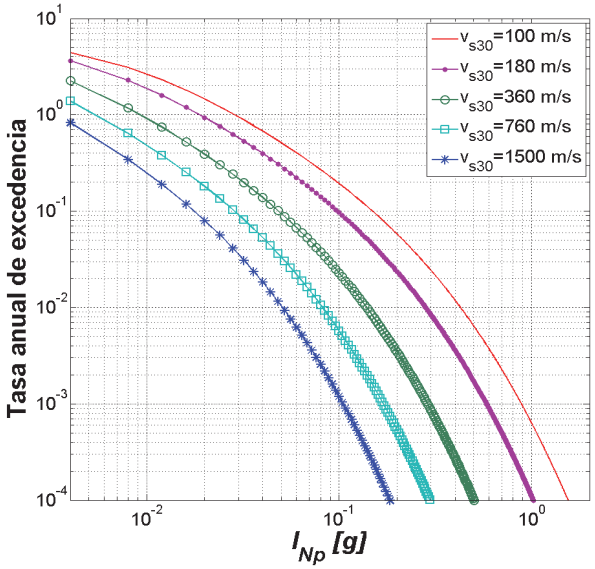


Figura 5. Curvas de peligro sísmico para un periodo estructural de un segundo para distintas velocidades de ondas de cortante para el sitio GSV.
Fuente: Propia.

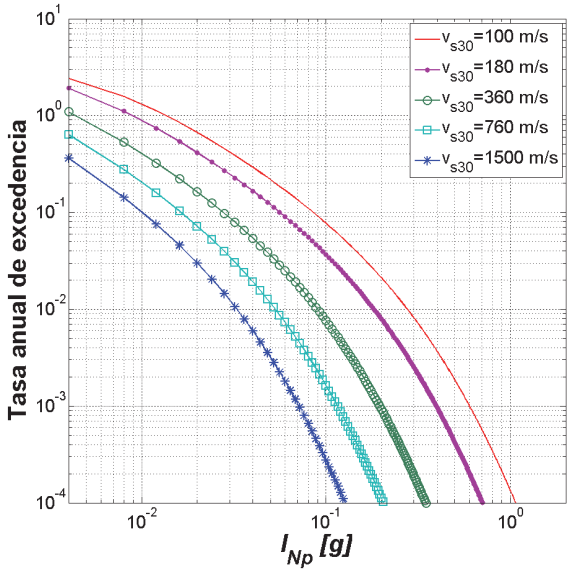


Figura 6. Curvas de peligro sísmico para un periodo estructural de un segundo para distintas velocidades de ondas de cortante para el sitio MZT-UAS.
Fuente: Propia.

Con el fin de mostrar cómo influye la velocidad promedio de las ondas de cortante sobre I_{Np} se elaboraron las Figs. 7-9, donde se grafican algunas curvas de peligro sísmico asociadas a periodos de 0.2, 0.6, 1.0, 2.0 y 3.0 segundos para el sitio FIC-UAS para velocidades promedio de ondas de cortante de 100, 360 y 1500 m/s. En las Figs. 7-9 se observa claramente como I_{Np} se ve afectada por la velocidad promedio de ondas de cortante, al ir incrementando la velocidad en un sitio en específico, I_{Np} va disminuyendo. En otras palabras a medida que la rigidez del suelo crece, los valores de I_{Np} tienden a reducirse.

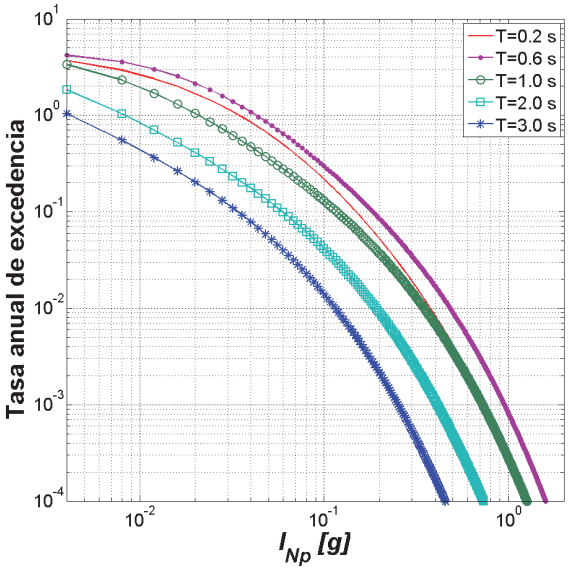


Figura 7. Curvas de peligro sísmico para el sitio FIC-UAS para una velocidad de ondas de cortante de 100 m/s.
Fuente: Propia.

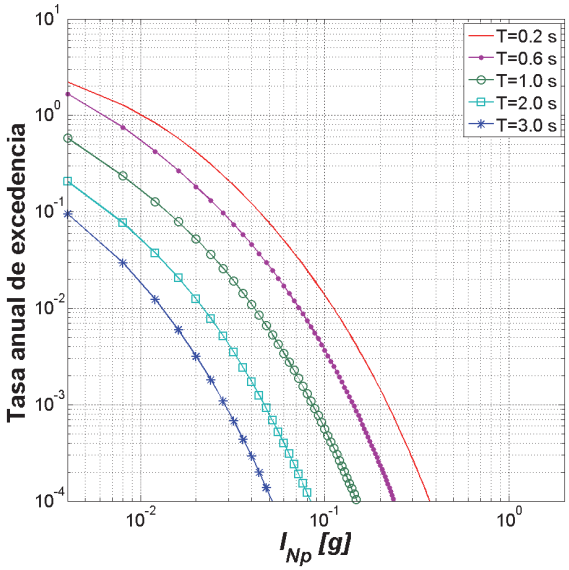


Figura 9. Curvas de peligro sísmico para el sitio FIC-UAS para una velocidad de ondas de cortante de 1500 m/s.
Fuente: Propia.

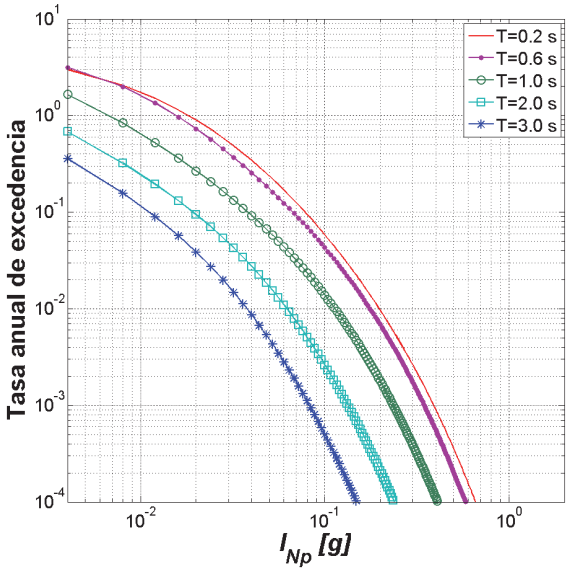


Figura 8. Curvas de peligro sísmico para el sitio FIC-UAS para una velocidad de ondas de cortante de 360 m/s.
Fuente: Propia.

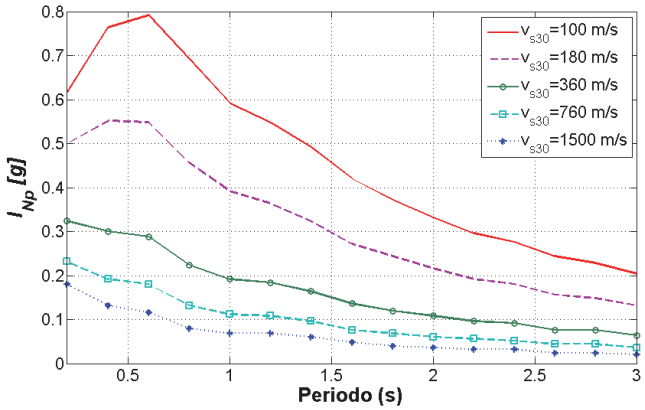


Figura 10. Espectros de I_{Np} y peligro uniforme para el sitio FIC-UAS para diferentes velocidades de ondas de cortante para un periodo de retorno de 475 años.
Fuente: Propia.

Un objetivo común del análisis de peligro sísmico probabilístico es obtener un espectro de respuesta de diseño a utilizar para el análisis estructural. Un enfoque para el desarrollo de un espectro de diseño es calcular un espectro de peligro uniforme. Este espectro se obtiene realizando los cálculos anteriores involucrados en el análisis de peligro sísmico para un intervalo de períodos, después se elige una tasa de excedencia objetivo, y para cada período se identifica la amplitud de la intensidad correspondiente a esa tasa. Estas amplitudes de intensidad se grafican contra sus respectivos periodos, este espectro es llamado espectro de peligro uniforme porque cada ordenada tiene una probabilidad igual de ser excedida.

En la Fig. 10 se ilustran los espectros de I_{Np} con peligro uniforme para un periodo de retorno de 475 años, y velocidades promedio de ondas de cortante 100, 180, 360, 760 y 1500 m/s, para el sitio FIC-UAS. Se observa que a medida que la velocidad de ondas de cortante disminuye, se incrementa la intensidad en términos de I_{Np}. También se muestra que la forma del espectro de peligro uniforme se va suavizando conforme se incrementa la velocidad de las ondas de cortante pero tiende a mantener su forma, y las intensidades máximas en términos de I_{Np} se concentran entre los periodos estructurales de 0.2 a 0.6 segundos.

El resumen de los resultados obtenidos en los sitios de interés se muestra en las Figs. 11-13, donde se comparan los espectros de peligro uniforme para los sitios FIC-UAS, GVE, MZT-UAS y FIM-UAS utilizando diferentes velocidades de ondas de cortante y diferentes periodos de retorno. Es evidente

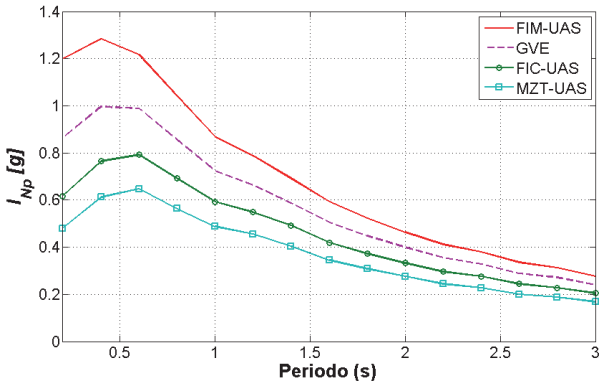


Figura 11. Espectros I_{Np} de peligro uniforme para los sitios FIC-UAS, GVE, MZT-UAS y FIM-UAS para una velocidad de ondas de cortante de 100 m/s para un periodo de retorno de 475 años.
Fuente: Propia.

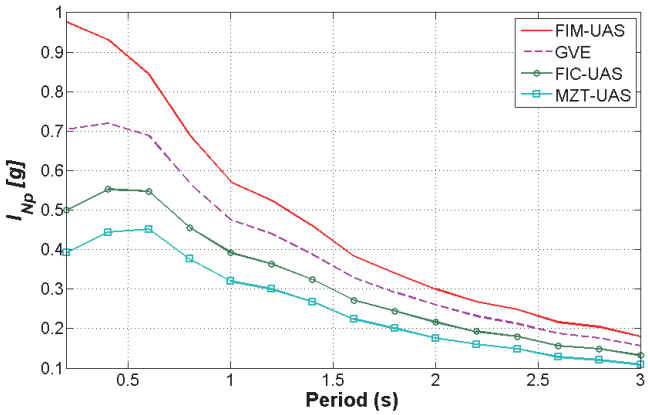


Figura 12. Espectros I_{Np} de peligro uniforme para los sitios FIC-UAS, GVE, MZT-UAS y FIM-UAS para una velocidad de ondas de cortante de 180 m/s para un periodo de retorno de 475 años.
Fuente: Propia.

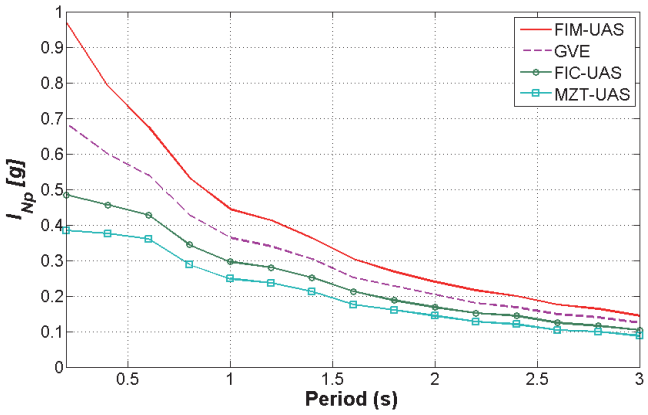


Figura 13. Espectros I_{Np} de peligro uniforme para los sitios FIC-UAS, GVE, MZT-UAS y FIM-UAS para una velocidad de ondas de cortante de 360 m/s para un periodo de retorno de 2475 años.
Fuente: Propia.

que el sito FIM-UAS es el que presenta mayor peligro sísmico y va decreciendo para los sitios GVE y FIC-UAS, hasta llegar el sitio MZT-UAS; es decir, en la parte norte de región de Sinaloa es donde se encuentran los sitios de mayor peligro sísmico, y conforme se analizan sitios de la parte sur del estado, se observa como el peligro sísmico va decreciendo. Lo anterior implica que los coeficientes sísmicos de diseño estructural son mayores en la parte norte del Estado de Sinaloa.

Una de las ventajas de utilizar I_{Np} como medida de intensidad es su eficiencia como lo demostró Buratti [11]; además, como se mencionó anteriormente él concluyó que una medida de intensidad sísmica eficiente resulta, al mismo tiempo, suficiente. La suficiencia se refiere a que la respuesta estructural depende solamente de la medida de intensidad utilizada y no de las características sísmicas de la fuente, tales como la magnitud del sismo y la distancia al sitio de interés. Por otro lado, una medida de intensidad sísmica eficiente es aquella que presenta una buena relación con la respuesta estructural.

Es de gran importancia mencionar que esta es la primera vez que se trazan espectros para I_{Np}, estos espectros son de gran utilidad para seleccionar acelerogramas. Supongamos que un código de diseño sísmico de alguna región nos pide un cierto número de registros sísmicos para el análisis estructural tomando como medida de intensidad $Sa(T_1)$, al utilizar I_{Np} se necesitan menos registros, esto es debido a la alta eficiencia de I_{Np}, lo cual a su vez se traduce en un diseño más eficaz y a la vez se ahorra tiempo de cómputo debido al uso de menos registros.

El trabajo aquí presentado solamente se llevó acabo para una región específica de la República Mexicana; sin embargo, eso no quiere decir que únicamente se aplicable para esta región. La formulación desarrollada por Bojórquez y Iervolino [10] puede ser empleada en cualquier parte del planeta siempre y cuando se cuente con los elementos necesarios para llevar acabo el análisis de peligro sísmico probabilístico.

5. Conclusiones

La estimación del peligro sísmico para la región de Sinaloa se efectuó tomando en cuenta la nueva medida de intensidad sísmica I_{Np} basada en el parámetro N_p, la característica principal de este parámetro es el uso de la forma espectral.

Se muestra la posibilidad de obtener tanto curvas de peligro sísmico como de espectros de peligro uniforme de manera sencilla utilizando la formulación desarrollada por Bojórquez y Iervolino [10]; además, se ilustra la importancia de la velocidad promedio de ondas de cortante, mientras mayor sea la velocidad, las curvas de peligro sísmico presentan menor intensidad en términos de I_{Np} para un sitio en específico, de igual manera se ven afectados los espectros de peligro uniforme por la velocidad de las ondas de cortante ya que estos son obtenidos a partir de las curvas de peligro.

De los resultados obtenidos se observa que el lugar que experimenta mayor peligro sísmico es el sitio FIM-UAS (es decir, la Ciudad de Los Mochis), lo cual indica que las ordenadas espectrales de diseño (coeficiente sísmico requerido) son mayores para este sitio. El segundo sitio con mayor peligrosidad sísmica corresponde a GVE. Para el sitio

FIC-UAS el peligró sísmico es menor que para el sitio GVE, y en este trabajo el lugar con menor peligro sísmico que se obtuvo corresponde al sitio MZT-UAS. Tomando en cuenta el análisis de peligro sísmico probabilístico desarrollado y la distribución geográfica de los sitios en estudios se concluye de manera preliminar que la parte norte de la región de Sinaloa es la que sufre de un mayor peligro sísmico, y conforme se avanza a la parte sur de la región de Sinaloa el peligro tiende a disminuir (al menos hasta la Ciudad de Mazatlán). Sin embargo, es importante resaltar que en todos los sitios considerados el peligro sísmico está presente, por lo tanto, las edificaciones deben diseñarse tomando en cuenta el efecto de los sismos. La distribución del peligro sísmico observada se debe a la cercana distancia entre la región de Sinaloa y la región sismotectónica BC2, y a la actividad sísmica de dicha región ya que es mayor en comparación con el resto de las otras fuentes que afectan al Estado. Por tal motivo, la región de Sinaloa se ve afectada considerablemente por las intensidades sísmicas que se generan a partir de eventos ocurridos en la región sismotectónica BC2.

Agradecimientos

Este trabajo se desarrolló con el apoyo económico brindado por El Consejo Nacional de Ciencia y Tecnología a través del Proyecto CB-2011-01-167419 y la Beca otorgada al primer autor. Se agradece el apoyo de la Universidad Autónoma de Sinaloa dentro del proyecto PROFAPI 2014/032.

Referencias

[1] Romeo, R., Paciello, A. and Rinaldis, D., Seismic hazard maps of Italy including site effects. Soil Dynamics and Earthquake Engineering, 20, pp. 85-92, 2000. DOI: 10.1016/S0267-7261(00)00040-3

[2] Meletti, C., Galadini, F., Valensise, G., Stucchi, M., Basili, R., Barba, S., Vannucci, G. and Boschi, E., A seismic source zone model for the seismic hazard assessment of the Italian territory. Tectonophysics, 450, pp. 85-108, 2008. DOI: 10.1016/j.tecto.2008.01.003

[3] Leyton, F., Ruiz, S. y Sepulveda, S.A., Reevaluación del peligro sísmico probabilístico en Chile central. Andean Geology, 37(2), pp. 455-472, 2010.

[4] Secanell, R., Goula, X., Susagna, T., Fleta, J. and Roca, A., Seismic hazard zonation of Catalonia, Spain, integrating random uncertainties. Journal of Seismology, 8, pp. 25-40, 2004. DOI: 10.1023/B:JOSE.0000009516.91044.51

[5] Sierra, C.A. y Jaramillo, J.D., IE - RISS: Una herramienta computacional para la estimación de pérdidas por sismo en edificaciones. DYNA, 79(175), pp. 34-42, 2012.

[6] Housner, G.W., Spectrum intensities of strong motion earthquakes. Proceedings of Symposium on Earthquake and Blast Effects on Structures, Earthquake Engineering Research Institute, 1952.

[7] Von-Thun, J.L., Rochin, L.H., Scott, G.A. and Wilson, J.A., Earthquake ground motions for design and analysis of dams. Earthquake Engineering and Soil Dynamics II – Recent Advance in Ground-Motion Evaluation, Geotechnical Special Publication 20 ASCE, New York, 1988, pp. 463-481

[8] Shome, N., Probabilistic seismic demand analysis of nonlinear structures, PhD dissertation, Department of Civil and Environmental Engineering, Stanford University, USA, 1999.

[9] Baker, J.W. and Cornell, C.A., A vector-valued ground motion intensity measure consisting of spectral acceleration and epsilon. Earthquake Engineering & Structural Dynamics, 34, pp. 1193-1217, 2005. DOI: 10.1002/eqe.474

[10] Bojórquez, E. and Iervolino, I., Spectral shape proxies and nonlinear structural response. Soil Dynamics and Earthquake Engineering, 31(7), pp. 996-1008, 2011. DOI: 10.1016/j.soildyn.2011.03.006

[11] Buratti, N., A comparison of the performances of various ground–motion intensity measures. Proceedings of The 15th World Conference on Earthquake Engineering, Lisbon, Portugal, pp. 24-28, 2012.

[12] SERVICIO GEOLÓGICO MEXICANO. [En línea]. Disponible en: http://portalweb.sgm.gob.mx/museo/es/riesgos/sismos/sismologia-de-mexico

[13] Reiter, L., Earthquake hazard analysis - issues and insights, Columbia University Press, New York, 1990, 254 P.

[14] Mendoza, A., Estudio de la estabilidad del valor b para regiones sismotectónicas de México, MSc Thesis, Programa de Posgrado en Ciencias de la Tierra, UNAM, México, 2012.

[15] Gutenberg, B. y Richter, C.F., Seismicity of the earth (and associated phenomena). Princeton University Press, Princeton, NJ, EUA, 1954.

[16] Duda, S.J., Secular seismic energy release in the circun-Pacific belt. Tectono-physics, 2, pp. 409-452, 1965. DOI: 10.1016/0040-1951(65)90035-1

[17] Figueroa, J., Catálogo de sismos ocurridos en la República Mexicana, Reporte No. 272. Serie: Investigación y desarrollo (azul), Coordinación: Sismología e Instrumentación Sísmica, Instituto de Geología, UNAM, 1970.

[18] Miyamura, S., Aspectos importantes de la sismicidad en Costa Rica, Rev. Avances de Inv., Inst. Inv. Soc., Univ. De Costa Rica, pp. 10-18, 1976.

[19] Abe, K., Magnitude of large shallow earthquakes from 1904 to 1980. Phys. of the Earth and Planet. Int., 27, pp. 72-92, 1981. DOI: 10.1016/0031-9201(81)90088-1

[20] Singh, S.K., Rodriguez, M. and Espíndola, J.M., A catalogue of shallow earthquakes of Mexico from 1900 to 1981. Bulletin of the Seismological Society of America, 74, pp. 267-279, 1984.

[21] The Global Centroid-Moment-Tensor. [Online]. Available at: http://www.globalcmt.org/

[22] García-Acosta, V. y Suárez, G., Los sismos en la historia de México, vol. 1, Eds. Científicas Universitarias: Serie texto científico universitario. Universidad Nacional Autónoma de México, 1996.

[23] Malagón, A., Calibración y estimación de magnitudes para sismos históricos de México, MSc Thesis, Universidad Nacional Autónoma de México, México. 1989.

[24] Zúñiga, F.R., Suárez, G., Ordaz, M. y García-Acosta, V., Proyecto: Peligro sísmico en Latinoamérica y el Caribe, Instituto Panamericano de Geografía e Historia, proyecto 89-0190, 1997, 84 P.

[25] Esteva, L. and Villaverde, R., Seismic risk, design spectra and structural reliability, Proceedings of 5th World Conf. Earthquake Eng, 1973, pp. 2586-2597.

[26] Bufaliza, M., Atenuación de intensidades sísmicas con la distancia en sismos mexicanos, MSc Thesis, Facultad de Ingeniería UNAM, 1984.

[27] Ordaz, M., Jara, J.M. y Singh, S.K., Riesgo sísmico y espectros de diseño en el Estado de Guerrero, Procedente del VIII Congreso Nacional de Ingeniería Sísmica, Acapulco, México, 1989, pp. D40-D56.

[28] García, J.D., Estimación de parámetros del movimiento fuerte del suelo para terremotos interplaca e intraslab en México Central, PhD Thesis, Facultad de Ciencias Físicas, Universidad Complutense de Madrid, Madrid, España, 2006.

[29] Boore, D.M. and Atkinson, G.M., Boore-Atkinson NGA ground motion relations for geometric mean horizontal component of peak and spectral ground motion parameters, PEER 2007/01, Pacific Earthquake Engineering Research Center, Berkeley, California, 2007.

[30] Bojórquez, E., Iervolino, I. and Manfredi, G., Evaluating a new proxy for spectral shape to be used as an intensity measure. Seismic Engineering Conference, AIP Conference Proceedings 1020, 2008, pp. 1599-1606. DOI: 10.1063/1.2963788

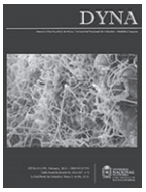
[31] Cordova, P.P., Dierlein, G.G., Mehanny, S.S.F. and Cornell, C.A. Development of a two parameter seismic intensity measure and probabilistic assessment procedure. The second U.S.-Japan Workshop on Performance-Based Earthquake Engineering

Methodology for Reinforce Concrete Biulding Structures, Sapporo, Hokkaido, 2001, pp. 187-206.
[32] Housner, G.W., Measures of severity of ground shaking. U.S. Conference on Earthquake Engineering. Earthquake Engineering Research Institute, 1975.
[33] Arias, A., A measure of earthquake intensity. In Hansen, R.J., Ed. Seismic Design for Nuclear Power Plants, Massachusetts Institute of Technology Press, 1970, pp. 438-483.
[34] FEMA., NEHRP Recommended provisions for the development of seismic regulations for new buildings, 1994.
[35] Mehanny, S.S. and Deierlein, G.G., Modelling of assessment of seismic performance of composite frames with reinforced concrete columns and steel beams. Report 135, Department of Civil and Enviromental Engineering, Stanford University, Sanford, USA, 2000.
[36] Riddell, R. and Garcia, J.E., Hysteretic energy spectrum and damage control. Earthquake Engineering & Structural Dynamics, 30, pp. 1791-1816, 2001. DOI: 10.1002/eqe.93
[37] Mehanny, S.S.F., A broad-range power-law form scalar-based seismic intensity measure. Engineering Structures, 31, pp. 1354-1368, 2009. DOI: 10.1016/j.engstruct.2009.02.003
[38] Baker, J.W. and Jayaram, N., Correlation of spectral acceleration values from NGA ground motion models. Earthquake Spectra, 24(1), pp. 299-317, 2008. DOI: 10.1193/1.2857544

R. Chávez-López, obtuvo el grado de Ingeniero Civil en 2010, en 2012 obtuvo su grado de MSc. en Ciencias de la Ingeniería, ambos por la Universidad Autónoma de Sinaloa, México. Actualmente se encuentra realizando sus estudios de Doctorado en Ciencias de la Ingeniería en la Facultad de Ingeniería, Culiacán, México en la misma universidad. Sus líneas de interés actual son ingeniería sísmica e ingeniería estructural.
ORCID: 0000-0002-9043-9070

E. Bojórquez-Mora, es profesor investigador de tiempo completo en la Facultad de Ingeniería de la Universidad Autónoma de Sinaloa, México. Terminó estudios de MSc y Dr. en la Universidad Nacional Autónoma de México en 2007 y su PhD en Ingeniería Sísmica en la Universidad de Nápoles, Italia en 2008. Coordinador de la Maestría y Doctorado en Ciencias de la Ingeniería (Estructuras) en la UAS, sede Culiacán, México. Actualmente es Editor de la Revista de Ingeniería Sísmica. Sus áreas de interés son: Ingeniería estructural, sísmica, eólica, sismología, inteligencia artificial, optimización multi-objetivo, entre otros.
ORCID: 0000-0001-6402-1693

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS



DYNA
http://dyna.medellin.unal.edu.co/

Mechanical and thermal performance of a geopolymeric and hybrid material based on fly ash

Diego Vásquez-Molina [a], Johanna M. Mejía-Arcila [a] & Ruby Mejía-de Gutiérrez [a]

[a] *Grupo de Materiales Compuestos (GMC), Universidad del Valle, Cali, Colombia*
johanna.mejia@correounivalle.edu.co, ruby.mejia@correounivalle.edu.co

Received: May 25th, de 2015. Received in revised form: October 5th, 2015. Accepted: October 22th, 2015

Abstract
This article presents an evaluation of the thermal behavior of a hybrid (FA/OPC) and geopolymer (FA100) material. The FA100 system is based on fly ash (FA), which has an elevated content of unburned material (14.8%). The FA/OPC system is comprised of a mixture of fly ash and ordinary Portland cement (OPC) at a proportion of 80/20. The thermal performance was evaluated by several tests, such as exposure to high temperatures (up to1000°C) and direct flame resistance. In addition, the effect of cyclic exposure was studied at 700°C for 10 cycles. FA/OPC hybrid material retains 92% of its initial strength and FA100 retains 113.3% of its initial strength at 700°C. Both materials can withstand 10 exposure cycles with a strength loss of less than 45%. In direct flame exposure, a temperature gradient of about 500°C was observed. These results indicate that these types of materials could possibly be used as fire-resistant materials in civil structures.

Keywords: fly ash; geopolymer; hybrid material; fire resistance

Comportamiento mecánico y térmico de un geopolímero y un material híbrido basado en ceniza volante

Resumen
Este artículo presenta la evaluación del comportamiento térmico de un material hibrido (FA/OPC) y un geopolimero (FA100). El sistema FA100 está basado en ceniza volante (FA) con un alto contenido de material orgánico (14,8%). El sistema FA/OPC se compone de la mezcla de FA y cemento portland (OPC) en proporción de 80/20. El desempeño térmico se evaluó mediante diferentes ensayos, exposición a altas temperaturas (hasta 1000°C) y resistencia a la llama directa. Adicionalmente se estudió el efecto de la exposición cíclica del material a temperatura de 700°C durante 10 ciclos. FA/OPC y FA100 retienen el 92%y el 113,3% respectivamente de su resistencia inicial a 700°C. Ambos materiales soportaron 10 ciclos y reportan una pérdida de resistencia menor del 45%. Respecto a la llama directa, se identificó un gradiente del orden de 500°C. Estos resultados indican que este tipo de materiales pueden ser utilizados en estructuras civiles resistentes al fuego.

Palabras claves: ceniza volante; geopolimero; material hibrido; resistencia al fuego

1. Introduction

Fly ash (FA) is a byproduct obtained from electricity generation via carbon in thermoelectric plants and boilers in several industries. This material has become one of several solid wastes that are generated in large volumes worldwide (on the order of 600 million tons reported), primarily due to the growing global energy demand. The large volume generated, its effect on the environment, and the excessive costs for handling and disposal have led numerous researchers to focus their attention on looking for solutions to this problem. However, although the potential markets are known, the decision to use the material in specific applications implies strict quality control, and the resulting material must demonstrate its competitive power against the natural resource or raw material that is to be replaced [1].

The primary variability of fly ash according to the source of origin is attributed to the reduced control over system

variables during coal burning processes; this lack of control is more critical in industrial boilers, where the main FA impurity is unburned carbon (abbreviated as "unburned"), which can reach values of up to 35% [2]. Temuujin et al. [3] investigated the effect of a thermal treatment (500-800°C) on ash to reduce the unburned levels and reported that it affects the vitreous phase at the surface level by 10% due to the crystallization of the mullite and hematite phases. Thus, its reactivity decreases, and the geopolymer mechanical strength decreases by approximately 20%. Lee et al. [4] used the flotation technique to reduce the carbon levels present in the ash and obtained positive results.

Fernández-Jiménez A. et al. [5] studied a group of F-type Spanish fly ash to determine the effect of their physicochemical characteristics on the mechanical strength of alkali-activated materials. The authors reported that the mechanical strength depends on the following factors: proportion of reactive silica and alumina, vitreous phase content, and particle size distribution. Adequate control of these characteristics allows compressive strengths of up to 70 MPa at 1 day of curing to be attained [6]. Additions of other supplementary materials to FA-based geopolymers to form binary mixtures, such as bottom slag, granulated blast furnace slag, and metakaolin, have also been investigated [7-9]. In relation to this, it has been reported that mixtures blended with granulated blast furnace slag (GBFS) are more resistant than those that contain metakaolin (MK) and that the optimum mixture percentages vary with the addition. Recently, hybrid systems have emerged [10], where a portion of the FA is replaced by a clinker or ordinary Portland cement (OPC) in orders of up to 30%); in this case, in addition to the aluminosilicate-type (N-A-S-H) gel characteristic of the alkali-activated FA, a (N-C)-A-S-H-type substitute gel is generated [11]. These types of hybrid materials are part of the currently so-called "*Alkali-activated Portland Cement-based Blended Cements*" [12].

The performance of these materials in aggressive environments and to the phenomenon of leaching has been reported as highly satisfactory [13, 14]. Rostami et al. [15] claim that this new type of binder could revolutionize the concrete industry and ceramics in general because its properties are excellent: a strength development of 90% in the first 24 hours, an ultimate strength of up to 110 MPa, excellent resistance to acid media, such as sulfuric, nitric, hydrochloric, and organic acids in general, and elevated frost resistance, which validates its use in high-performance structural applications. Because these alternative binders have an application in the civil industry, the study of its behavior against temperature, specifically in the event of a fire, has been an object of study. In general, reports have shown a better performance of these materials compared with traditional OPC-based materials, and several authors attribute this to structural transformations at high temperatures that lead to the formation of crystalline phases of greater thermal stability [16,17].

The objective of the present study is to evaluate the fire behavior of a 100% FA geopolymer and an FA/OPC 80/20 hybrid type material produced from FA with a high-unburned level. To accomplish this, the materials were exposed to different temperatures up to 1000°C and to thermal cycles at a temperature of 700°C, and the effect on the compressive strength was evaluated. Additionally, a direct flame exposure test was performed. The study is expanded on with a microstructural evaluation of the materials using the following techniques: XRD, FTIR, and SEM.

2. Methodology and experimental techniques

2.1. Materials

To synthesize the geopolymeric material, FA from the Termopaipa thermoelectric power station was used as the aluminosilicate precursor. The FA particle size was determined by laser granulometry using a Mastersizer 2000 instrument, which indicated a mean size D(4.3) of 63.9 μm; after a milling process (ball mill) for 90 minutes, the mean size decreased to 19.5 μm. To obtain the hybrid binder, OPC was used as a source of calcium. Table 1 shows the chemical composition of the corresponding materials, it was measured by an X-ray fluorescence (XRF) technique using MagixPro PW-2440 X-ray fluorescence spectrometer, equipped with a Rhodium tube with a maximum output of 4.0 KW and 0.02% of sensitivity. The elevated loss on ignition (14.8%) for FA as a consequence of the organic material present (unburned carbon) is noteworthy; likewise, OPC has a reported 9.6% loss on ignition, which is attributed to the addition of limestone in the final stage of the cement production process. Fernández-Jiménez A. et al. [5] determined the factors that affect the quality of the material fabricated with alkali-activated fly ash and reported the following as the main characteristics of fly ash: an unburned percentage below 5%, a Fe_2O_3 content below 10%, a low content of CaO, and a reactive silica content between 40 and 50%. Regarding grain size, it is claimed that at least 80-90% of the grains must be smaller than 45 μm in size; finally, large amorphous-phase content is recommended. When comparing the physicochemical characteristics of the FA used in the present investigation, it is observed that the ultimate criterion for the unburned level is not met. According to the chemical composition data (Table 1), the SiO_2/Al_2O_3 molar ratio of the ash is 4.25, and according to the standard ASTM C618 [18], the ash is of type F.

Fig. 1 shows the morphology of the fly ash grains obtained by scanning electron microscopy (SEM). The images show the presence of spherical grains, some of which are compact and some hollow; the latter contain smaller spheres in their interior,

Table 1.
Chemical composition of the raw materials.

Compound (% wt)	FA	OPC
SiO_2	53.7	20.2
Al_2O_3	21.5	7.0
Fe_2O_3	4.5	4.8
CaO	0.8	58.4
TiO_2	1.0	-
K_2O	1.4	-
S	0.6	-
MgO	0.6	-
P_2O_5	0.5	-
Na_2O	0.3	-
Loss on Ignition (950°C)	14.8	9.6

Source: The authors

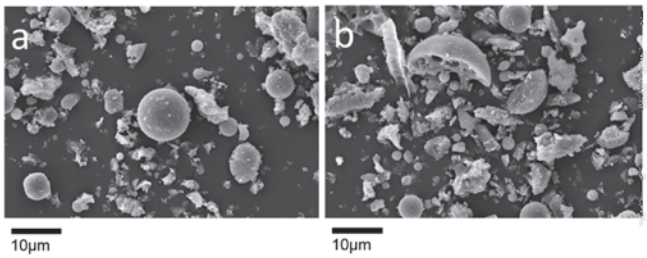


Figure 1. a) Fly ash before milling, b) Fly ash after milling
Source: The authors

which are comprised of the so-called cenospheres and plerospheres [19]. The milling process reduces the fraction of spherical grains, where rupture increases the specific area and, consequently, increases the reactivity of the material [20].

2.2. Specimen preparation

Two cementitious specimens were prepared, one composed of a fly ash-based geopolymer (FA-100) and an FA/OPC 80/20 hybrid system, which contained 20% OPC as a replacement for the fly ash.

2.3. Testing techniques

The microstructure and mechanical strength were evaluated in the paste-type specimens, and the thermal performance of the mortar specimens was analyzed. The mechanical strength was evaluated at early ages of 7 and 28 days of curing using a Universal INSTRON 3369 press at a displacement rate of 1 mm/min. The thermal behavior was

Table 2 shows the composition of the mixtures according to the following factors: molar ratios of SiO2/Al2O3 and Na2O/SiO2 and the FA/FA+OPC ratio; the same table includes the liquid/solid ratio and the applied curing system. The l/s ratio corresponds to the water ratio in the alkali-activator and free water, and s refers to the sum of the proportions of precursors and the anhydrous activator fraction. Sand from Ottawa and a binder:sand ratio of 1:2.75 were used to prepare the mortars. A mixture of sodium hydroxide (Merck analytical grade reactant) and sodium silicate (32.4% SiO2, 13.5% Na2O, 54.1% H2O) was used as the alkali-activator to obtain the molar ratios shown in Table 2.

2.3. Testing techniques

The microstructure and mechanical strength were evaluated in the paste-type specimens, and the thermal performance of the mortar specimens was analyzed. The mechanical strength was evaluated at early ages of 7 and 28 days of curing using a Universal INSTRON 3369 press at a displacement rate of 1 mm/min. The thermal behavior was

Table 2.
Geopolymeric mixtures produced.

System	SiO_2/Al_2O_3	Na_2O/SiO_2	l/s	Curing
FA-100	4.4	0.2	0.4	24 hours at 65°C, ~90% RH
	5.3	0.2		
	6.0	0.15-0.2-0.3		
FA/OPC 80/20	4.4	0.2	0.4	In humid chamber, ~80% RH
	5.3	0.2		
	6.0	0.15-0.2-0.3		

Source: The authors

evaluated to determine the residual strength after exposure of the specimens at 400, 700, and 1000°C at a heating rate of 2°C/min and a holding time of 1 hour at each temperature. Additionally, a direct flame resistance test was performed for 30 minutes; in addition to the flame temperature data, this text identified the temperature transmission of the material during heating. Finally, a study of the effect of cyclic exposure was performed at 700°C with a holding time of 1 hour and cooling in the oven; the procedure was executed for 10 cycles. Microstructural and morphology analysis were conducted by a scanning electron microscopy (SEM) using a JEOL JSM-6490 LV instrument. This device contains an INCAPentaFETx3 detector (brand OXFORD INSTRUMENTS, Model 7573); before examination, the samples were coated with gold by deposition in a Denton Vacuum Desk IV vacuum unit, mineralogical compounds were identified by X-ray diffraction using a X'Pert PANalytical MRD instrument with a CuKα1 tube operated at 45 kV and 40 mA, for which the scan was performed for 30 minutes in the 2θ range of 4-60°.

3. Discussion of results

3.1. Compressive strength and microstructural characterization

Fig. 2 shows the compressive strength results obtained for the pastes of the two systems, FA100 and FA/OPC 80/20, as a function of the molar ratios of SiO_2/Al_2O_3 and Na_2O/SiO_2 and curing age. Here, it should be noted that in both evaluated systems, the optimum molar ratio for SiO_2/Al_2O_3 and Na_2O/SiO_2 correspond to 6 and 0.2, respectively. However, the strengths drastically differed for each system. FA/OPC 80/20 exhibited a compressive strength of up to 2 and 3.8 times that of the FA100 geopolymer at 28 and 360 days of curing, respectively. It should be noted that the presence of OPC in the geopolymeric mixture, in addition to increasing the strength, contributes to eliminate the thermal curing stage (~70°C) primarily for simple fly ash systems, which is in accordance with results from other researchers [21]. It is inferred that the cement hydration heat favors the dissolution of SiO_2 and Al_2O_3 species of the geopolymeric precursor. Palomo A. et al. [22] reported a compressive strength of 38 MPa at 28 days for a fly ash hybrid dosed with 30% OPC, whereas in the present investigation, it was possible to obtain a strength of 55.6 MPa with 20% OPC. Furthermore, an increasing strength development was observed for the hybrid system, whereas in the geopolymeric system (FA100), the strength reached at 7 days is similar to that exhibited by the material at 360 days (approximately 22 MPa); the strength reported for the hybrid

material at the same curing age was 85.5 MPa.

Fig. 3 shows the X-ray diffractograms for FA and the geopolymers, FA-100 and FA/OPC 80/20, after 28 days of curing with a SiO_2/Al_2O_3 molar ratio of 6 and a Na_2O/SiO_2 molar ratio of 0.2. It was observed that the FA presented a high content of crystalline phase, as evidenced by the presence of quartz (Q, ICSD: 062404), mullite (M, ICSD: 100805), and hematite (H, ICSD: 082137). The increase in the base line in the 2θ region of 15-30° is associated with the presence of the vitreous phase. The diffractograms corresponding to FA100 and FA/OPC 80/20 indicate that the crystalline phases characteristic of FA and OPC remain unaltered, i.e., the Q, M, and H of FA and calcite ($CaCO_3$; ICSD: 16710) in the OPC, which confirms that these phases do not participate in the alkali-activation process. In addition, the formation of Trona (ICSD: 062200) ($Na_3H(CO_3)22H_2O$) was identified in the FA100, which is due to the characteristic peaks at 33.84 (100%) and 29.05 (80%) and can be attributed to the action of CO_2 in the environment.

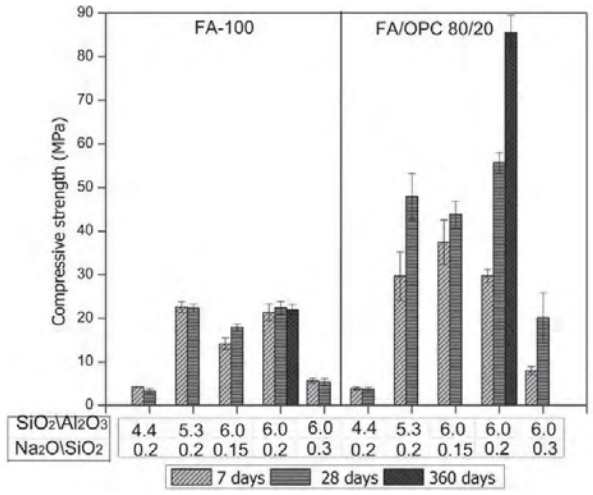


Figure 2. Compressive strength as a function of the molar ratios and curing age.
Source: The authors

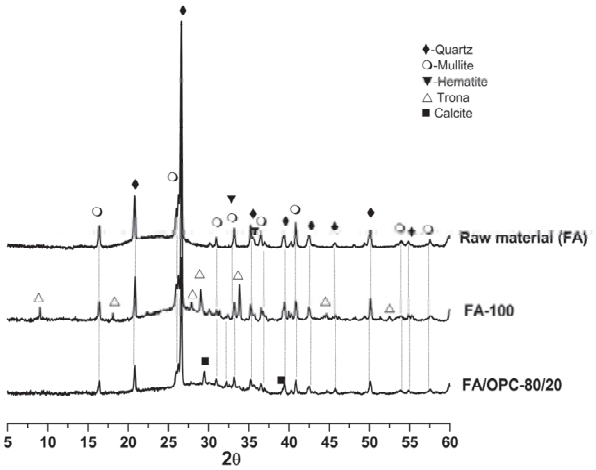


Figure 3. Diffractogram of the alkali-activated systems after 28 days of curing.
Source: The authors

The formation of portlandite was not observed in the hybrid material, which indicates that calcium ions must be part of the amorphous gels formed in the geopolymerization; this result suggests the presence of (N,C)-A-S-H-type gels in this system [23,24]. The amorphous halo between 20-35° 2θ, which is associated with the vitreous phase, is greater for the FA100 geopolymer.

3.2. Thermo-mechanical behavior

3.2.1. Exposure at high temperatures

The thermal behavior of both FA-100 and FA/OPC-80/20 was evaluated by exposing the mortars to different temperatures (400, 700, and 1000°C). Table 3 shows the percentage weight loss and the volumetric change with respect to the initial value (positive values (+) represent expansion, and negative values (-) represent contraction). The weight loss in both materials at 400°C was similar (12.7%), which is attributed to the evaporation of surface water located in the pores, the physical binding of water to the reaction products, and the partial volatilization of organic matter (carbon) provided by FA [24]. At higher temperatures, this loss is attributed to a dehydration process of the structure [26, 27, 17], and to the decomposition of the carbonates present, i.e., the sodium and calcium carbonates identified by XRD. Weight loss at 1000°C for the two evaluated systems was different by approximately 2%, which can be associated with the limestone content present in the cement used (OPC); this result occurs because, based on the OPC chemical composition shown in Table 1, the incorporation of 20% OPC into the system leads to a 1.96% loss on ignition, which is equivalent to the CO_2 of the decomposition of the limestone present in this cement fraction. Regarding the volumetric change in the material, contraction and expansion changes occurred. The physical appearance of the specimens after heating at 1000°C corroborates the transformation of a dense matrix into a porous matrix with a sponge-like appearance and partial surface vitrification (Fig. 4)

The mechanical strength values of the mortars were 18.3 MPa and 31.34 MPa at 28 curing days for the FA100 and FA/OPC-80/20 systems, respectively. After heating at 700°C, a strength increase of 13.3% was observed in the FA geopolymer, whereas a reduction of 7.8% was observed in the FA/OPC hybrid. These results are comparable to the results presented by other researchers [27]. At 1000°C, due to the excessive expansion of the test samples, it was not possible to evaluate mechanical strength. None of the systems contained cracks (Fig. 4). Authors such as *Kong et al.* [27,28] associated this behavior with the presence of hollow particles in the ash, facilitating the escape of water vapor, thus avoiding severe cracking in the material. Heating at the different temperatures modified the color of the test samples; at the end of the test, a grey coloration with small red dots was obtained in FA100. In FA/OPC, the specimens become orange, which is attributed to the loss of carbon and to the oxidation of iron (Fig. 4); other researchers have reported similar results [29].

3.2.2. Exposure to thermal cycles

The exposure to thermal cycles was performed at a temperature of 700°C for 10 cycles. At the end of each cycle,

Table 3.
Weight changes and volume loss with respect to temperature

Mixture	Property	400°C	700°C	1000°C
FA-100	Weight loss, %	12.64	12.92	18.25
	Volumetric Change, %	-29.38	-5.69	73.95
FA/OPC 80/20	Weight loss, %	12.75	14.94	20.01
	Volumetric Change, %	-34.72	5.08	132.12

Source: The authors

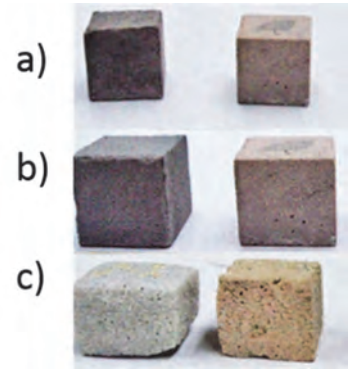


Figure 4. Thermal exposure (from left to right: FA-100, FA/OPC 80/20) - a) 400°C, b) 700°C, and c) 1000°C
Source: The authors

Table 4.
Weight losses and volumetric changes associated with thermal cycles at 700°C ((-) contraction, (+) expansion)

System	Cycle N°	Weight loss, %	Volume change, %
FA-100	2	12.92	(+) 6.59
	6	11.56	(+) 5.40
	10	11.36	(+) 5.17
FA/OPC 80/20	2	14.94	(+) 2.98
	6	12.80	(+) 2.90
	10	12.07	(-) 3.35

Source: The authors

the test samples were removed to determine their weight, volumetric change, and residual strength. Table 4 shows the values associated with weight loss and volumetric change of each system at the end of cycles 2, 6, and 10.

It should be noted that no significant changes were observed in weight loss as the number of 700°C cycles increased. In relation to volumetric change, the FA-100 geopolymer was transformed into a porous, foamed material with an expansion of approximately 5-6% with respect to the initial dimensions of the specimen (before being subjected to 700°C). In contrast, FA/OPC presented a lower expansion during the first two cycles (2.98%); subsequently, FA/OPC contracted by approximately 3%. It should be noted that after the 10 thermal cycles, the materials changed color, and even though they maintained their physical integrity, a 56% decrease in mechanical strength was observed for the FA-100 system, and a 40% decrease was observed for the FA/OPC 80/20 system (Fig. 5). These results can be attributed to the increased in porosity and to microstructural transformations, as mentioned in other studies [30]. The residual strength in the first case was 9.4 MPa, and the second was approximately twice that value (17.65 MPa). These values were determined for specimens exposed to fire after 28 days of curing, whose initial strength values were 18.3 MPa and 31.34 MPa for FA100 and FA/OPC 80/20, respectively.

Fig. 6 shows the SEM images of the two materials at the end of the 10 thermal cycles, where the formation of cavities or pores can be observed; thus, when the densified, continuous structure characteristic of the geopolymeric and hybrid materials disappeared, the irregularities became critical points that weakened the structure [26]. Strength losses of the same order (52%) have also been reported in other studies of fly-ash-based geopolymeric mortars exposed to 800°C [31]. It is claimed that, in general, these materials exhibit a greater thermal stability than that of Portland concretes, which lose their structural integrity at temperatures of around 700 and 800°C due to calcium silicate hydrated (CSH) decomposition. In contrast, in systems that are alkali-activated with sodium silicates, increasing temperature can contribute to generating new crystalline phases, such as Na-feldspars, nepheline ($NaAlSiO_4$), or albite ($NaAlSi_3O_8$), which contribute to a greater thermal stability at high temperatures [29,32]. Fernández-Jiménez et al. (2008) [33] reported the presence of a nepheline phase at temperatures above 600°C and the presence of albite at temperatures of approximately of 1000°C in an FA-based geopolymer [30].

3.2.1. Direct flame resistance

Fig. 7 shows a plot of the values obtained in the direct flame resistance test. The temperature of the flame emitted by the torch incident on the surface of the FA-100 and FA/OPC 80/20 specimens remained stable after the first four minutes at approximately 820°C until the end of the thermal test (30 minutes). The temperature, which was measured with a thermocouple and is associated with the temperature transferred through the material in the FA/OPC 80/20 and FA-100 systems, was significantly lower than the flame temperature. A temperature gradient of approximately 500°C was observed in FA/OPC 80/20, and a slightly lower temperature was observed in the FA-100 geopolymer, which indicates that these materials act as thermal insulators; it is

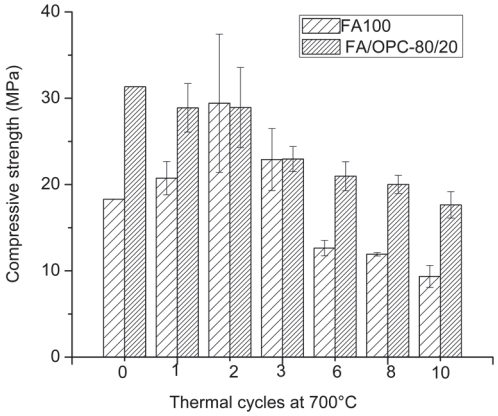


Figure 5. Compressive strength after each exposure cycle
Source: The authors

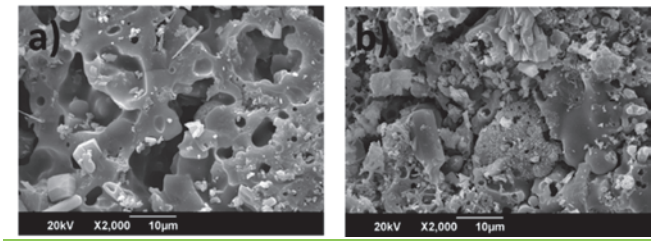


Figure 6. SEM microphotograph. Specimen after 10 thermal cycles. a) FA-100, b) FA/OPC 80/20.
Source: The authors

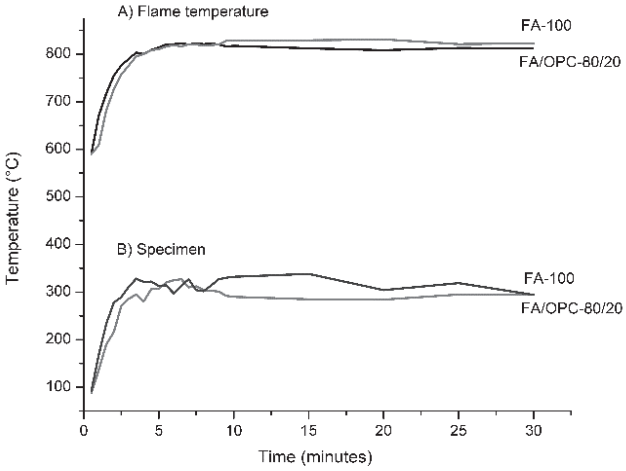


Figure 7 Temperature transfer vs exposure time. A) flame temperature of FA/OPC 80/20 (black) and FA-100 (gray), B) Sepecimen temperature FA/OPC 80/20 (gray) and FA-100 (black)
Source: The authors

inferred that, with the use of these materials as binders in prefabricated elements (panel-type) or as binders in concrete, the structure would be protected in the case of a fire, providing time that could positively contribute to avoiding the catastrophic failure of the structure and the resulting human impact. Cheng et al. [34] produced panels from the alkali-activation of steel slags, which were exposed at 1100°C (flame temperature); after 35 minutes, the temperature on the panel side was approximately 350°C, a reduction in the order of 700°C, which corresponds to the behavior reported here. Several researchers have suggested the use of these materials as thermal insulators on both ceramic and metallic substrates [35,36].

Fig. 8 shows the test samples before and after flame exposure. In the FA/OPC-80/20 system, cracking was observed on the surface that was directly exposed to the flame, and the initial coloration of the material transitioned from black to a lighter color in the direct exposure region, while its surroundings became light pink. In contrast, no cracks were observed in the FA-100 geopolymer, although coloration changes did occur. In samples 3 and 4, which correspond to the opposite and lateral surfaces of the two materials, no effects were observed, which is indicative of the good performance of the materials.

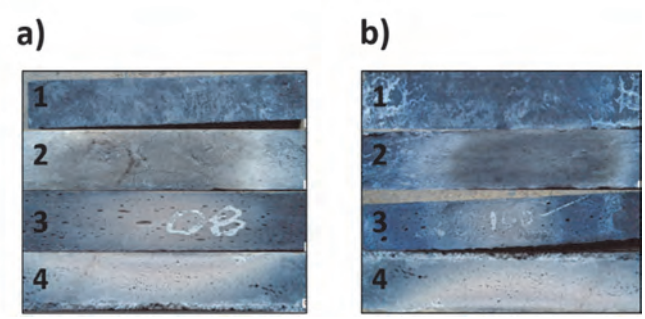


Figure 8. a) FA/OPC 80/20, b) FA-100. 1= before exposure 2= side of direct exposure, 3=opposite side, 4= lateral view of the test sample.
Source: The authors

4. Conclusions

- The results of this study indicate that fly ash containing unburned material of around 15% does not affect the long-term strength (360 days) or thermal exposure performance of the resulting material, which enables the use of lower quality FA to produce geopolymers and hybrid materials.
- A 20% addition of OPC to the FA100 geopolymer increases its mechanical strength by up to 150% and 288% at curing ages of 28 and 360 days, respectively. In addition, FA/OPC 80/20 does not require thermal curing. These are significant advantages of the hybrid material compared with the geopolymer.
- In general, FA-100 and FA/OPC-80/20 exhibited good thermal performance. The residual strength at a high temperature (700°C) was 92% of the initial strength in the hybrid material and 113.3% that of the geopolymer. The two evaluated materials are capable of withstanding 10 heating/cooling cycles at 700°C with less than 45% loss of mechanical strength at the end of 10 cycles.
- The behavior of the materials against a direct flame with a thermal gradient of approximately 500°C from the internal side of the panel was good.
- Based on these results, these materials could represent an alternative that can protect or reinforce structures in the case of fires due to their good mechanical behavior and thermal isolation capacity.

Acknowledgments

The authors, members of Composites Materials Research Group, wish to thank the Colombian Institute for the Development of Science, Technology, and Innovation (Colciencias), Universidad del Valle (CENM) (Cali, Colombia) for providing the financial support needed to conduct the present study, Project Hybricement, Contract N° 0638-2013. The authors also wish to thank GENSA Termopaipa, thermoelectric power station for providing the fly ash used as a raw material to fabricate the geopolymer and hybrid material.

References

[1] Iyer R.S. and Scott. J.A., Power station fly ash - A review of value-added utilization outside of the construction industry. Resources, Conservation

and Recycling, 31(3), pp. 217-228, 2001. DOI: 10.1016/S0921-3449(00)00084-7.
[2] Valderrama, C., Torres, J. and Mejìa-de Gitiérrez, R., A high unburned carbon fly ash concrete's performance characteristics. Ingenieria e Investigación, 31(1), pp. 39-46, 2011.
[3] Temuujin, J. and van Riessen, A., Effect of fly ash preliminary calcination on the properties of geopolymer. Journal of Hazardous Materials, 164(2-3), pp. 634-639, 2009. DOI: 10.1016/j.jhazmat.2008.08.065.
[4] Lee, S., Seo, M.D., Kim, Y.J., Park, H.H., Kim, T.N., Hwang, Y. and Cho, S.B., Unburned carbon removal effect on compressive strength development in a honeycomb briquette ash-based geopolymer. International Journal of Mineral Processing, 97(1-4), pp. 20-25, 2010. DOI: 10.1016/j.minpro.2010.07.007.
[5] Fernandez-Jimenez, A. and Palomo, A., Characterisation of fly ashes. Potential reactivity as alkaline cements. Fuel, 82(18), pp. 2259-2265, 2003. DOI: 10.1016/S0016-2361(03)00194-7.
[6] Duxson, P. and Provis, J.L., Designing precursors for geopolymer cements. Journal of the American Ceramic Society, 91(12), pp. 3864-3869, 2008. DOI: 10.1111/j.1551-2916.2008.02787.
[7] Chindaprasirt, P., Jaturapitakkul, C., Chalee, W. and Rattanasak, U., Comparative study on the characteristics of fly ash and bottom ash geopolymers. Waste Management, 29(2), pp. 539-543, 2009. DOI: 10.1016/j.wasman.2008.06.023.
[8] Puertas, F. and Fernández-Jiménez, A., Mineralogical and microstructural characterisation of alkali-activated fly ash/slag pastes. Cement and Concrete Composites, 25(3), pp. 287-292, 2002.
[9] Fernández-Jiménez, A., Monzó, M., Vicent, M., Barba, A. and Palomo, A., Alkaline activation of metakaolin-fly ash mixtures: Obtain of Zeoceramics and Zeocements, Microporous and Mesoporous Materials, 108(1-3), pp. 41-49, 2008. DOI: 10.1016/j.micromeso.2007.03.024.
[10] Shi, C., Fernández-Jiménez, A. and Palomo, A., New cements for the 21th century: The pursuit of Na alternative to Portland Cement, Cement and Concrete Research, 41, pp. 750-763, pp. 750-763. DOI: 10.1016/j.cemconres.2011.03.016
[11] Garcia-Lodeiro, I., Fernández-Jiménez, A., Palomo, A. and Macphee, D.E., Effect of calcium additions on N-A-S-H cementitious gels. Journal of the American Ceramic Society, 93(7), pp. 1934-1940, 2010. DOI: 10.1111/j.1551-2916.2010.03668.x
[12] Palomo, A., Krivenko, P., Garcia-Lodeiro, I., Kavalerova, E., Maltseva, O. and Fernández-Jiménez, A., A review on alkaline activation: new analytical perspectives, Materiales de Construcción, 64(315), e022, 2014. DOI: 10.3989/mc.2014.00314.
[13] Miranda, J.M., Fernández-Jiménez, A., González, J.A. and Palomo, A., Corrosion resistance in activated fly ash mortars. Cement and Concrete Research 35(6), pp. 1210-1217, 2004. DOI: 10.1016/j.cemconres.2004.07.030.
[14] Martínez-López, C., Mejía-Arcila, J.M, Torres-Agredo, J. and Mejía-de Gutiérrez, R., Evaluation of the toxicity characteristics of two industrial wastes valorized by geopolymerization process, DYNA 82(190), pp. 74-81, 2015. DOI: 10.15446/DYNA.V82N190.43136
[15] Rostami, H. and Brendley, W., Alkali ash material: A Novel fly ash-based cement. Environmental Science Technology, 37(15), pp. 3454-3457, 2003. DOI: 10.1021/es026317b
[16] Fernández-Jiménez, A., Pastor, J., Martín, A. and Palomo, A., High-temperature resistance in alkali-activated cement. Journal of the American Ceramic Society, 93(10), pp. 3411-3417, 2010. DOI: 10.1111/j.1551-2916.2010.03887.x.
[17] Donatello, S., Kuenzel, C., Palomo, A. and Fernandez-Jimenez, A., High temperature resistance of a very high volume fly ash cement paste. Cement and Concrete Composites, 45, pp. 234-242, 2014. DOI: 10.1016/j.cemconcomp.2013.09.010.
[18] ASTM C618-12a., Standard Specification for Coal Fly Ash and Raw or Calcined Natural Pozzolan for Use in Concrete.
[19] Siddique, R., and Iqbal-Khan, M., Supplementary cementing materials, Engineering Materials, Chapter 1 Fly Ash, Springer-Verlag Berlin Heidelberg, 2011, pp. 1-66.
[20] Kumar, S. and Kumar, R., Mechanical activation of a fly ash: Effect on reaction, structure and properties of resulting geopolymer, Ceramics International, 37(2), 2011, pp. 533-541., DOI: 10.1016/j.ceramint.2010.09.038.
[21] García-Lodeiro, I., Palomo, A., Maliseva, O. and Fernández-Jiménez, A., Hybrid alkaline cements. Part I: Fundamentals, Romanian Journal of Materials, 42,(4), pp. 330-335, 2012.
[22] Palomo, A., Fernández-Jiménez, A., Kovalchuk, G., Ordoñez, L. and Naranjo, M., OPC-fly ash cementitious systems: Study of gel binders produced during alkaline hydration. Journal of Materials Science, 42(9), pp. 2958-2966, 2007. DOI: 10.1016/j.cemconcomp.2014.08.008.
[23] García-Lodeiro, I., Fernández-Jiménez, A. and Palomo, A., Variation in hybrid cements over time. Alkaline activation of fly ash–portland cement blends. Cement and Concrete Research, 52, pp. 112-122, 2013. DOI: 10.1016/j.cemconres.2013.03.022.
[24] Macphee, D. and Garcia-Lodeiro, I., Activation of aluminosilicates - some chemical considerations, Proceedings of Slag Valorisation Symposium, 2011. pp. 51-61.
[25] Duxson, P., Lukey, G. and van Deventer, S.J.J., Thermal evolution of metakaolin geopolymers: Part 1 - Physical evolution. Journal of Non-Crystalline Solids, 352(52-54), pp. 5541-5555, 2006. DOI: 10.1016/j.jnoncrysol.2006.09.019.
[26] Buchwald, A., Vicent, M., Kriegel, R., Kaps, C., Monzó, M. and Barba, A., Geopolymeric binders with different fine fillers - Phase transformations at high temperatures. Applied Clay Science, 46(2), pp. 190-195, 2009. DOI: 10.1016/j.clay.2009.08.002.
[27] Kong, D.L.Y. and Sanjayan, J.G., Effect of elevated temperatures on geopolymer paste, mortar and concrete. Cement and Concrete Research, 40(2), pp. 334-339, 2010. DOI: 10.1016/j.cemconres.2009.10.017.
[28] Kong, D.L.Y., Sanjayan, J.G. and Sagoe-Crentsil, K., Comparative performance of geopolymers made with metakaolin and fly ash after exposure to elevated temperatures. Cement and Concrete Research, 37(12), pp. 1583-1589, 2007. DOI: 10.1016/j.cemconres.2007.08.021.
[29] Rashad, A. and Zeedan, S., The effect of activator concentration on the residual strength of alkali-activated fly ash pastes subjected to thermal load. Construction and Building Materials, 25(7), pp. 3098-3107, 2011. DOI: 10.1016/j.conbuildmat.2010.12.044.
[30] Bakharev, T., Thermal behaviour of geopolymers prepared using class F fly ash and elevated temperature curing. Cement and Concrete Research, 36(6), pp. 1134-1147, 2006. DOI: 10.1016/j.cemconres.2006.03.022.
[31] Abdulkareem, O., Al Bakri, M., Kamarudin, H., Nizar, K. and Saif, A., Effects of elevated temperatures on the thermal behavior and mechanical performance of fly ash geopolymer paste, mortar and lightweight concrete. Construction and Building Materials, 50, pp. 377-387, 2014. DOI: 10.1016/j.conbuildmat.2013.09.047.
[32] Junaid, M., Khennane, A., Kayali, O., Sadaoui, A., Picard, D. and Fafard, M., Aspects of the deformational behaviour of alkali activated fly ash concrete at elevated temperatures. Cement and Concrete Research, 60 pp. 24-29, 2014. DOI: 10.1016/J.CEMCONRES.2014.01.026.
[33] Fernandez-Jimenez, A. and Palomo, A., New cementitious materials based on alkali-activated fly ash: Performance at high temperatures. Journal of American Ceramic Society, 91(10), pp. 3308-3314, 2008. DOI: 10.1111/j.1551-2916.2008.02625.x.
[34] Cheng, T.W. and Chiu, J.P., Fire-resistant geopolymer produced by granulated blast furnace slag. Minerals Engineering, 16(3), pp. 205-210, 2003. DOI: 10.1016/S0892-6875(03)00008-6.
[35] Temuujin, J., Rickard, W. and van Riessen, A., Characterization of various fly ashes for preparation of geopolymers with advanced applications. Advanced Powder Technology, 24(2), pp. 495-498, 2013. DOI: 10.1016/j.apt.2013.01.013.
[36] Temuujin, J., Minjigmaa, A., Rickard, W., Lee, M., Williams, I. and van Riessen, A., Fly ash based geopolymer thin coatings on metal substrates and its thermal evaluation. Journal of Hazardous Materials, 180(1-3), pp. 748-752, 2010. DOI: 10.1016/j.jhazmat.2010.04.121.

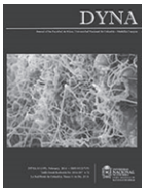
D. Vásquez-Molina, received a BSc. degree in Materials Engineering in 2014 from Universidad del Valle, Cali, Colombia. He worked on alternatives cementitious binders and assessed the mechanical, durability and thermal development on geopolymers and hybrid binders.
ORCID: 0000-0002-6115-8771

J.M. Mejia-Arcila, received a BSc. degree in Materials Engineering in 2010 from Universidad del Valle, Cali, Colombia. She is currently a PhD student at Universidad del Valle. Her research interests include industrial waste

valorization, geopolymers and alkaline activation of by-products and minerals.
ORCID: 0000-0002-4094-0422

R. Mejia-de Gutiérrez, received a BSc. Chemistry degree and MSc. in 1972 and 1986 respectively, from Universidad del Valle, Cali, Colombia, and her PhD. degree was obtained in 1997 from Universidad Complutense, Madrid, España. Currently, she is a Titular professor in the Engineering Materials School, Universidad del Valle, Cali, Colombia. She is coordinator of the Master's and Doctoral programs focus area of Materials Engineering. Senior researcher and director of the research group Composite Materials – GMC (CENM). Her research interests include: industrial waste valorization, geopolymers, alkaline activation, alternative cementitious binders, durability, corrosion of concrete structures and composites materials.
ORCID: 0000-0002-5404-2738

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS



DYNA
http://dyna.medellin.unal.edu.co/

Sol-gel synthesis of zinc oxide nanoparticle at three different temperatures and its characterization via XRD, IR and EPR

Manuel Acosta-Humánez [a], Luis Montes-Vides [b] & Ovidio Almanza-Montero [c*]

[a] *Departamento de Física. Universidad Nacional de Colombia. Bogotá. Colombia. mafacostahu@unal.edu.co*
[b] *Departamento de Geociencias. Universidad Nacional de Colombia. Bogotá. Colombia. lamontesv@unal.edu.co*
[c] *Departamento de Física. Universidad Nacional de Colombia. Bogotá. Colombia. oaalmanzam@unal.edu.co*

Received: May 25th, 2015. Received in revised form: September 11th, 2015. Accepted: September 21th, 2015.

Abstract
In this work, nanoparticles of zinc oxide were synthesized; they were formed using the sol-gel method (citrate route) at calcination temperatures (Tc) of 500, 550 and 600 °C. For all samples studied, IR spectroscopy showed the presence of the bands associated with water molecules present in the zinc oxide and carbon dioxide adsorbed on its surface. The formation of zinc oxide phase was confirmed by XRD, which showed that from 500 ºC it had this type of Wurtzite structure. However, samples calcinated at 600 ºC have higher crystallinity. Crystallite size was calculated using the Scherrer equation. The Rietveld method was used to obtain lattice parameters a and c for Wurtzite cell types as well as cell volume and the ratio c/a of each sample. These parameters do not show significant changes when they are compared with values obtained from samples with different calcination temperatures. Electron paramagnetic resonance showed the presence of defects in the zinc oxide. Three signals with g values of 1.96, 2.00 and 2.04 were associated with oxygen and zinc vacancies. Defects in the structure disappear when the calcination temperature is increased. The sample that was highlighted with the highest concentration of vacancies has a mean crystallite size greater than 30 nm, and this may also be responsible for this feature.

Keywords: X-Ray diffraction, IR spectroscopy, sol-gel method, electron paramagnetic resonance, zinc oxide.

Síntesis sol-gel de nanopartículas de óxido de zinc a tres termperaturas diferentes y su caracterización vía XRD, IR y EPR

Resumen
Se sinterizaron nanopartículas de óxido de zinc (ZnO) mediante el método sol-gel (ruta citrato) a las temperaturas de calcinación (Tc) de 500, 550 y 600 °C.Se mostró por espectroscopía IR la presencia de las bandas asociadas a moléculas de agua presentes en el óxido de zinc así como de dióxido de carbono adsorbido en su superficie, para todas las muestras estudiadas. La tenencia de una estructura tipo wurtzita, propia del ZnO, incluso para las muestras calcinadas a 500 ºC, fue confirmada por XRD, Las muestras preparadas a Tc de 600 ºC tienen mayor cristalinidad que las muestras calcinadas a 500 ºC. El tamaño de cristal fue evaluado por la ecuación de Scherrer. El método Rietveld fue usado para obtener los parámetros de red a y c para la celda tipo wurtzita, así como el volumen de celda y la relación c/a de cada muestra. Estos parámetros no muestran desviaciones significativas cuando ellos son comparados entre valores obtenidos para muestras con distintas temperaturas de calcinación. Por resonancia paramagnética electrónica se mostró la presencia de defectos en el óxido de zinc. Tres señales, con valores de g = 1.96, 2.00 y 2.04 fueron asociados a vacancia de oxígenos y vacancias de Zinc respectivamente. Los defectos en la estructura desaparecen a medida que se incrementa la temperatura de calcinación. La muestra con la mayor concentración de vacancias tiene un tamaño medio de cristalito mayor de 30 nm y esto último puede ser el responsable de esta característica.

Palabras clave: Difracción de rayos X, espectroscopía IR, método sol-gel, resonancia paramagnética electrónica, ZnO.

1. Introduction

Zinc oxide (ZnO) is a type II-VI semiconductor with a direct band gap of 3.37 eV and stable Wurtzite type structure with lattice parameters a = 3.25 Å and c = 5.21 Å [1]. It is an important semiconductor material due to its applications, which include transparent conductive oxides (TCO) [2,3], ultraviolet (UV) blockers, and photocatalysts, among others. As photocatalysts, the

reduction of organic pollutant compounds and the remediation of organic contaminants, mainly azo type (compounds that interact with visible light), can be mentioned as being the most important in terms of usefulness. The presence of free radicals or vacancies in them are responsible for these applications and some authors have established that the amount of them is dependent on the crystal size, given that the ratio surface / volume increases as the size of the material decreases to a nanometer range [4,5]. ZnO is a relatively open structure with a hexagonal-close-packed lattice where Zn atoms occupy half of the tetrahedral sites, while all octahedral sites are empty [6]. The open structure also influences the nature of defects and the diffusion mechanism. The identity, quantity, and stability of these radicals or vacancies are features to consider when this material is required in a particular application. Applications for these types of materials depend on their electrical and magnetic properties and these depend on the method by which they were prepared [1-5].

Many methods for the production of ZnO nanostructures have been described in the literature such as laser ablation [7], hydrothermal methods [8], electrochemical depositions [9], chemical vapor deposition [10], thermal decomposition [11], the combustion method [12] and the co-precipitation method [13,14], resulting in zinc oxide nanoparticles with a nanometric size. The sol-gel method allows the mixture of the initial reagent on an atomic level, which, while there is control of chemical composition and there are quite homogeneous materials in its composition, reduces the possibility of having impurities that are difficult to detect and has good reproducibility. Materials of high surface area can be prepared at a low cost and the experimental procedure can be easily undertaken [15,16]. In this work, pure zinc oxide was synthesized using the sol-gel method (citrate route) at various calcination temperatures. Structural properties and the mean crystal size were determined by XRD, and its purity was assessed by infrared spectroscopy. The presence of free radicals, vacancies or defects was observed by electron paramagnetic resonance, which could be associated to potential uses of these materials as reductors of organic pollutant compounds.

2. Materials and methods

Nanoparticles of Zinc oxide were prepared by the sol-gel (citrate route) method using Zinc Nitrate, $Zn(NO_3)_2 \bullet 6H_2O$ (Panreac) as the starting materials. Citric acid $C_6H_8O_7 \bullet H_2O$ (Panreac) was used as a complexing agent. A stoichiometric weight of metallic cation was homogeneously mixed in deionized water using magnetic stirring at 70 °C. Then, Citric acid was also dissolved in deionized water for 30 minutes. The Citrate/Nitrate ratio used was 1. The solution of Zinc Nitrate in the solution of Citric acid was added slowly. The reaction mixture was heated at 70 °C with continuous stirring until gel formation. The product obtained was dried at 130 °C for 12 hours and pulverized for 30 minutes using an agate mortar. The precursor materials were calcined at the calcination temperature (Tc) of 500, 550 and 600 °C for 12 hours [4]. These samples were denominated in ZnOTc form. The reaction of zinc oxide formation can be written as described in equation (1):

$$Zn(NO_3)_2 \cdot 6H_2O + C_6H_8O_7 \cdot H_2O + 4O_2 \rightarrow ZnO + 2NO_2 + 6CO_2 + 11H_2O \quad (1)$$

Infrared spectra were carried out in a Fourier transform spectrophotometer Shimadzu IRAffinity-1 in the range of wavenumber 4000-600 cm^{-1}. X-ray diffraction measurements were performed on a diffractometer Panalytical Pro MPD using polycrystalline silicon as a calibration standard in 2θ range of 10-90, and using CuKα1 radiation and a step size (2θ) of 0.0260. Sweep time was 18.6860 s. The EPR spectra were carried out on a Bruker X-band spectrometer ESP-300 at different temperatures. A frequency of 9.45 GHz and a modulation amplitude of 10.496 G were also used in this procedure.

3. Results and discussion

The purity and molecular structure of the samples were determined by infrared spectroscopy. Fig. 1 shows the FTIR spectra of the nanocrystalline powders of zinc oxide. The two peaks at ~3428 and 1590 cm^{-1} are attributed to the stretching vibrations of the -OH group and the deformation of H-O-H bond, which are associated with small amounts of water existing in the zinc oxide [1]. The band corresponding to ~ 2400 cm^{-1} is associated with CO_2 molecules that are adsorbed on the surface of ZnO [17,18]. The bands that are near to 1000 cm^{-1} are associated with metal-Oxygen tension and bending, in this case, Zn-O. An increase of the temperature in this band is associated with high Zn-O bonding due to thermal evolution of samples [18]. Two small bands are observed at 2920 and 2850 cm^{-1}, possibly due to -CH_2 groups belonging to the metal-organic chain formed during synthesis [18]; these are small residues and the bands are not intense. All these results were independent of calcination temperature (Tc).

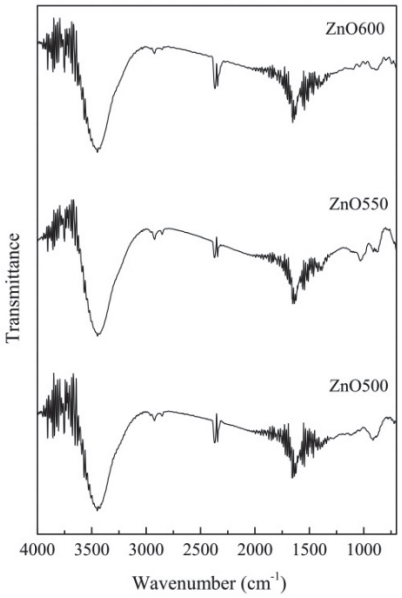


Figure 1. Infrared spectra for Zinc oxide samples at different Tc values. Source: The authors.

Fig. 2 shows diffractograms for ZnO samples as well as the precursor material studied at different calcinations temperatures. At this point, the evolution of the precursor material is displayed, which is an organometallic compound (zinc citrate), and its evolution towards zinc oxide can be appreciated when calcination temperature is higher than 500 ºC. For all samples studied, all diffraction peaks were associated with Zinc oxide according to the PDF-Card 36-1451 (2004 PDF2 database) and the structure was a wurtzite type. There were no observed peaks corresponding to some of the reagent precursors or other secondary phases, and all samples have good crystallinity, which increases at the same time as the calcination temperature Tc. The Wurtzite structure for ZnO, is therefore observed from Tc = 500 ºC and this calcination temperature could be enough to obtain samples that are going to be used as a photocatalyst, that in turn decreasing production costs.

It is worthwhile mentioning that the diffraction peak associated with the plane (101) is the most intense one, and this may be due to the synthesis method used since it has been reported that for thin films the most intense level corresponds to the (002) reflection [19-21]. The preferred orientation of growth, determined using a texture coefficient TC (hkl) was calculated using equation (2) [19] and showed that for all samples calcinated, the preferred growth plane was (100) except ZnO calcinated at 600 °C, which grew preferentially in the plane direction (101).

$$TC(hkl) = \frac{I(hkl)/I_0(hkl)}{N^{-1}\sum_n I(hkl)/I_0(hkl)} \quad (2)$$

In this equation, *I (hkl)* is the measured relative intensity of the plane *(hkl)*, I_0 *(hkl)* is the relative intensity of the standard plane *(hkl)* taken from the JCPDS database, N is the number of reflections in the diffractogram, 7 being the most intense (Fig. 2), and *n* is the number of used diffraction peaks, which in our case was 3 associated with the planes (100), (002) and (101). For a sample that has randomly oriented crystallites, the value of *TC(hkl)* is equal to 1, whereas if this value increases, it means that the greater abundance of crystals is oriented in one particular direction *(hkl)* [20] . *TC(hkl)* data for the three main diffraction peaks (100), (002) and (101) are shown in Table 1.

Mean crystallite size was estimated from the Scherrer equation, with a gaussian fit,

$$D_s = \frac{K\,\lambda}{\beta\cos\theta} \quad (3)$$

where β is integral breadth (in radians), λ is the wavelength of X rays (CuK_α radiation, 0.1540598 *nm*), *K* is Scherrer's constant that depends on the directions of the crystal lattice and crystal morphology [21]; its value is 0.94 and θ is Bragg's angle. Integral breadths of diffraction peaks for all samples were calculated using the following equation:

$$\beta = [(\beta_m)^2 - (\beta_{instr})^2]^{1/2} \quad (4)$$

Where β_{instr} is the integral breadth for a calibration standard, and β_m the integral breadth for diffraction peaks of synthesized samples. The results for D_s are shown in Table 2, and D_s values are nanometric. The lowest D_s is obtained when calcination temperature is 550 °C; ZnO calcinated at 500 °C has an average crystallite size above 30 nm. Differences in crystal sizes (Ds) are determined by the Scherrer equation, and this is true for all samples as they have different integral breadth values as consequence of the fact they were calcinated at different temperature Tc.

The lattice parameters a and c as well as cell volume (V) were calculated by the Rietveld method, using the Fullprof program [22]. Table 2 shows the calculated values. a and c do not change significantly with calcination temperature. The ratio c/a shows the deviation that crystalline structure of the samples presented with respect to the theoretical network hcp (hexagonal compact packing, the value of which is 1.633). The deviations regarding the theoretical value may be products of microstrains and the concentration of defects in the crystal.

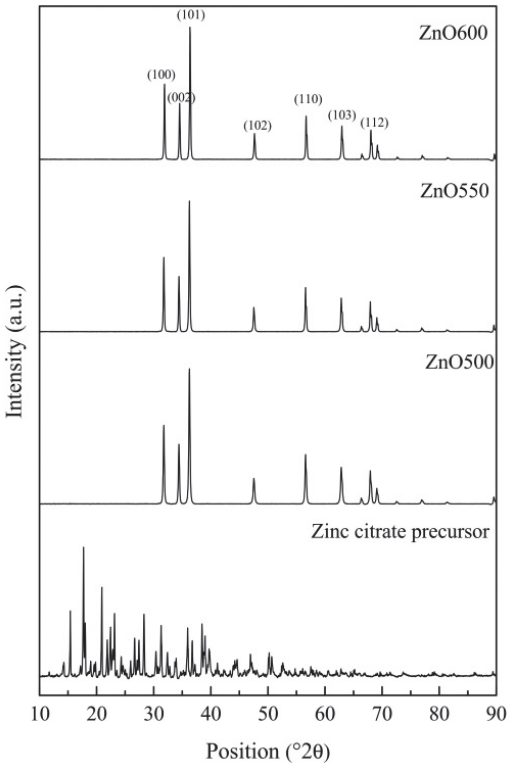


Figure 2. Diffractograms of ZnO, including precursor material, at different Tc values.
Source: The authors.

Table 1.
TC (hkl) values calculated from equation (2).

	Planes		
	(100)	**(002)**	**(101)**
Samples		***TC (hkl)***	
ZnO500	2.370	2.341	2.289
ZnO550	2.376	2.335	2.289
ZnO600	2.314	2.289	2.398

Source: The authors.

Table 2.
Mean crystallite size is given by the Scherrer equation (D_s). Cell parameters a and c for a hexagonal cell, cell volume V and c/a axial ratio are also shown in the table.

Samples	D_S (nm)	a (Å)	c (Å)	V (Å^3)	c/a
ZnO500	30.9	3.2509	5.2090	47.675	1.6023
ZnO550	18.0	3.2507	5.2081	47.661	1.6022
ZnO600	21.9	3.2516	5.2093	47.699	1.6021

Source: The autors.

Fig. 3 shows the EPR spectra for the Zinc oxide samples that are calcinated at three different temperatures. At T_c = 500 ° C three signals are observed having g values of 1.96, 2.00 and 2.04. Some authors have associated these lines to defects in Zinc oxide [4]. The signal localized at the g = 1.96 line was attributed to Oxygen vacancies [4,23,24]. The other two lines are associated with Zinc vacancies [25]. Stehr *et al.* attributed these lines to shallow donors, which show a slight anisotropy because of the Wurtzite type crystal structure of zinc oxide [26]. The sample calcinated at 550 °C showed a small concentration of defects associated with Oxygen vacancies, while the sample calcinated at 600 ° C showed no vacancies, at least in the order of determination of these defects by EPR. It means that pure ZnO calcinated at 550 ºC could be used as a photocatalyst if the oxidation/reduction activity were caused by the presence of defects or vacancies. The sample is highlighted with the highest concentration of vacancies that has a mean crystallite size greater than 30 nm, and this may also be responsible for this feature.

4. Conclusion

Zinc oxide nanoparticles were synthesized at calcination temperatures of 500, 550 and 600 °C. For all samples, the formation of wurtzite-type structures of zinc oxide was observed. The crystal size indicates that samples contain nanosized crystals, and that there are also small microstrains associated with small variations in the hexagonal lattice of zinc oxide. Lattice parameters showed no significant variations in the unit cell due to variation of the calcination temperature. The presence of water and Carbon dioxide molecules on the surface of zinc oxide was evidenced by infrared spectroscopy. Zinc vacancies and oxygen were shown by electron paramagnetic resonance, a process in which their disappearance is shown as long as the calcination temperature increases. The presence of vacancies in samples calcined at Tc = 500 °C, which reacts with a mean crystallite size above 30 nm should be the best catalyst performance or oxidizing/reducing agent.

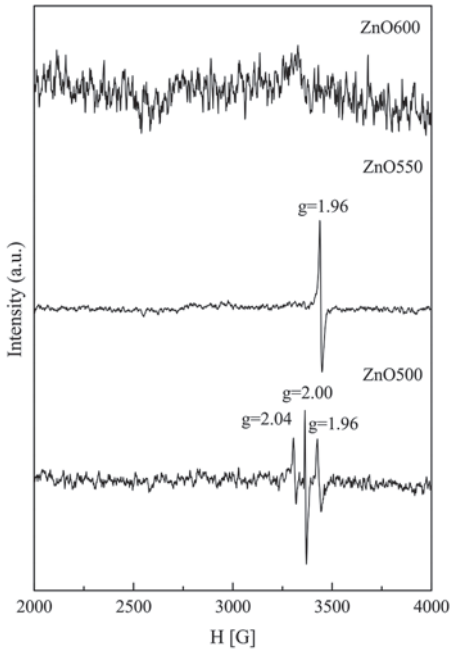


Figure 3. EPR spectra for zinc oxide samples at different Tc values.
Source: The authors.

Acknowledgment

We would like to give thanks to the DIB of the National University of Colombia for their financial support through project 15848. Additionally, we thank professor Mario Barrera for facilitating our use of the equipment and laboratories for the material synthesis.

References

[1] Jagannatha-Reddy, A., Kokila, M.K, Nagabhushana, H., Chakradhar, R.P.S, Shivakumara, C., Rao J.L. and Nagabhushana, B.M., Structural, optical and EPR studies on ZnO:Cu nanopowders prepared via low temperature solution combustion synthesis. J. Alloys. Compd. 509(17), pp. 5349-5355, 2001. DOI: 10.1016/j.jallcom.2011.02.043

[2] Hisono, H., Recent progress in transparent oxide semiconductors: Materials and device application. Thin Solid Films. 515(5), pp. 6000-6014, 2007. DOI: 10.1016/j.tsf.2006.12.125

[3] Becheri, A., Durr, M., Lo Nostro, P. and Baglioni, P., Synthesis and characterization of zinc oxide nanoparticles: Application to textiles as UV absorbers. J. Nanopart. Res. 10, pp. 679-689, 2008. DOI: 10.1007/s11051-007-9318-3

[4] Acosta-Humánez, F., Cogollo-Pitalúa, R. and Almanza, O., Electron paramagnetic resonance in $Zn_{1-x}Co_xO$. J. Magn. Magn. Mater. 329, pp. 39-42, 2013. DOI: 10.1016/j.jmmm.2012.10.026

[5] Wu, C., Shen, L., Shang, Y.C. and Huang, Q., Solvothermal synthesis of Cr-doped ZnO nanowires with visible light-driven photocatalytic activity. Mat. Lett. 65(12), pp. 1794-1796, 2011. DOI: 10.1016/j.matlet.2011.03.070

[6] Cao, B., Gong, H., Zeng, H. and Cai, W.P. Photoluminiscence / Fluorescence spectroscopic technique for nanomaterials characterization, in Nanomaterials: Processing and characterization with lasers, Weinheim, Wiley, 2012, pp. 597-616.

[7] Scarisoreanu, N., Metai, D.G., Dinescu, G, Epurescu, G, Ghica, C, Nistor, L.C. and Dinescu, M., Properties of ZnO thin films prepared by radio-frequency plasma beam assisted laser ablation. Appl. Surf. Sci. 247, pp.518-525, 2005. DOI: 10.1016/j.apsusc.2005.01.140

[8] Ni, Y.H., Wei, X.W., Hong, J.M. and Ye, Y., Hydrothermal synthesis and optical properties of ZnO nanorods. Mater. Sci. Eng., B, 121, pp. 42-47, 2005. DOI: 10.1016/j.mseb.2005.02.065

[9] Chang, S., Yoon, S.O., Park, H.J. and Sakai, A., Luminescence properties of Zn nanowires prepared by electrochemical etching. Mater. Lett. 53, pp. 432-436, 2002. DOI: 10.1016/S0167-577X(01)00521-3

[10] Wu, J.J. and Liu, S.C., Low-temperature growth of well-aligned ZnO nanorods by chemical vapor deposition. Adv. Mater. 14, pp. 215-218, 2002. DOI: 10.1002/1521-4095(20020205)14:3<215::AID-ADMA215>3.0.CO;2-J

[11] Wang, R.C. and Tsai, C.C., Efficient synthesis of ZnO nanoparticles, nanowalls, and nanowires by thermal decomposition of zinc acetate at a

low temperature. Appl. Phys. A. 94, pp. 241-245, 2009. DOI 10.1007/s00339-008-4755-0
[12] Lamas, D.G., Lascalea, G.E. and Walsoc, N.E., Synthesis and characterization of nanocrystalline powders for partially stabilized zirconia ceramics. J. Eur. Ceram. Soc. 18, pp. 1217-1221, 1998. DOI: 10.1016/S0955-2219(98)00045-4
[13] Kumar, S.S., Venkateswarlu, P., Rao, V.R. and Rao, V.R., Synthesis, characterization and optical properties of zinc oxide nanoparticles. Int. Nano Lett [Online]. 30(3), pp. 2-6, 2013. [Date of reference 11th October of 2015]. Available at: http://www.inl-journal.com/content/pdf/2228-5326-3-30.pdf
[14] Ruiz, C.V. and Rodríguez-Páez, J.E., Aluminatos de sodio obtenidos del sistema $Al(NO_3)_3{\cdot}9H_2O$ a través del método de precipitación controlada. Ingeniería e Investigación [Online]. 30(2), pp. 16-24, 2010. [Date of reference 13th October of 2015]. Available at: http://www.revistas.unal.edu.co/index.php/ingeinv/article/view/15726/34068
[15] Jokela, S.J. and McCluskey, M.D., Structure and stability of O-H donors in ZnO from high-pressure and infrared spectroscopy. Phys. Rev. B. 72(11), pp. 113201(1-4), 2005. DOI: 10.1103/PhysRevB.72.113201
[16] Montenegro-Hernández, A. y Rodríguez-Páez, J.E., Síntesis óxido de estaño altamente reactivo utilizando como precursor etilhexanoato de estaño. Ingeniería e Investigación [Online]. 29(1), pp. 47-52, 2009 [Date of reference 13th October of 2015] Available at: http://www.revistas.unal.edu.co/index.php/ingeinv/article/view/15142/34413.
[17] Nejati, K., Rezvani, Z. and Pakizevand, R., Synthesis of ZnO nanoparticles and investigation of the ionic template effect on their size and shape. Int. Nano Lett. [Online]. 1, pp. 75-81, 2011 [date of reference December 15th of 2014]. Available at: http://inljournal.com/?_action=article&vol=1&issue=4&_is=Volume+1%2C+Number+2+%28July+2011%29
[18] Nakamoto, K., Infrared and raman spectra of inorganic and coordination compounds, Applications in coordination, organometallic, and bioinorganic chemistry. Wiley, New York, 2009.
[19] Caglar, Y., Sol-gel derived nanostructure undoped and cobalt doped ZnO: Structural, optical and electrical studies J. Alloys Compd. 560, pp. 181-188, 2013. DOI: 10.1016/j.jallcom.2013.01.080
[20] Xu, C., Cao, L., Su, G., Liu, W., Qu, X. and Yu, Y., Preparation, characterization and photocatalytic activity of Co-doped ZnO powders. J. Alloys Compd. 497(1-2), pp. 373-376, 2010. DOI: 10.1016/j.jallcom.2010.03.076
[21] Yang, J.H, Zhao, L.Y., Ding, X., Yang, L.L., Zhang, Y.J., Wang, Y. X. and Liu, H.L., Magnetic properties of Co-doped ZnO prepared by sol-gel method. Materials Science and Engineering B., 162(3), pp. 143-146, 2009. DOI: 10.1016/j.mseb.2009.03.020
[22] Rodríguez-Carvajal, J., Fullprof suite software: Crystallographic tools for rietveld, profile matching and integrated intensity refinements for X-ray and/or neutron data. Version 5.30. Laboratorie Leon, Brillounin, Gifsuryvette, France, 2012.
[23] Vlasenko, L.S., Magnetic resonance studies of intrinsic defects in ZnO: Oxygen vacancy. Appl Magn Reson. 39(1), pp. 103-111, 2010. DOI: 10.1007/s00723-010-0140-1
[24] Janotti, A. and Van de Walle, C.G., Native point defects in ZnO. Phys. Rev. B. 76(16), pp. 165202(22), 2007. DOI: 10.1103/PhysRevB.76.165202
[25] Vlasenko, L.S. and Watkins, G.D., Optical detection of electron paramagnetic resonance for intrinsic defects produced in ZnO by 2.5-MeV electron irradiation in situ at 4.2 K. Phys. Rev. B. 72(3), pp. 035203(12), 2005. DOI: 10.1103/PhysRevB.72.035203
[26] Stehr, J.E., Meyer, B.K. and Hofmann, D.M., Magnetic resonance of impurities, intrinsic defects and dopants in ZnO. Appl Magn Reson. 39(1), pp. 137-150, 2010. DOI: 10.1007/s00723-010-0142-z

M. Acosta-Humánez, received his BSc. in Chemistry in 2009 from the Universidad de Córdoba, Montería, Colombia. In 2015 he received and MSc. Eng. Degree in Materials and Processes from the Universidad Nacional de Colombia, Bogotá, Colombia. His research interests include: Chemical synthesis process of inorganic materials (oxides), characterization techniques (XRD, EPR, IR, UV-vis) as well as electrochemical methods for materials synthesis and characterization.
ORCID: 0000-0003-0610-4831.

L. Montes-Vides, received his BSc. in Physicist - graduated from the Universidad Nacional de Colombia, MSc. in Engineering in 1987, from the Universidad Nacional de Colombia, and Dr. in Science – Geophysics in 1998. from the Universidade Federal do Pará, Brazil. He has worked at the Universidad Nacional de Colombia in Bogotá since 1990 as a lecturer on the following programs: undergraduate program in Geology, graduate programs in MSc. Geophysics and Doctorate program in Geosciences, all from the Faculty of Sciences. His research areas focus on Seismic prospecting, seismic inversion, seismic modeling and applied computing mathematics. He is a reviewer of the following scientific journals: ESRJ, Revista da SBGf, J.Appl. Geophysics and CT&F.
ORCID: 0000-0002-7470-9202

O. Almanza-Montero, received his PhD. in Physics in 2000, from the Universidad de Valladolid, Spain, and his MSc. in Physics from the Universidad Nacional de Colombia. He is currently a professor in this University and has worked there since 1996. His research is focused on Materials science such as semiconductor and antioxidant activity of Colombian fruits. He is author of more than 50 papers published in different journals. He regularly uses the following characterization techniques: (XRD, EPR, IR, UV-vis).
ORCID: 0000-0002-5141-6079

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS

DYNA
http://dyna.medellin.unal.edu.co/

Evaluation of internal and external stresses on the SPT sampler

Juliana Zapata-Galvis [a] & Edmundo Rogério-Esquivel [b]

[a] *Departamento de Geotecnia, Escola de Engenharia de São Carlos- Universidade de São Paulo, São Carlos, Brasil. julyszapata44@hotmail.com*
[b] *Departamento de Geotecnia, Escola de Engenharia de São Carlos- Universidade de São Paulo, São Carlos, Brasil. esquiveleesc@gmail.com*

Received: June 9th, 2015. Received in revised form: September 11th, 2015. Accepted: September 21th, 2015.

Abstract
The N_{SPT} index is usually used in empirical correlations to estimate parameters of the soil, carrying capacity, foundation settlement, etc. Because these correlations have no scientific basis, researchers have been developing rational methods of analysis, based on energy concepts. The amounts of energy involved in the SPT test are evaluated by the EFV method. With force and acceleration records, amounts of energy, equipment efficiency, experimental dynamic reaction force of the soil and stresses acting on the sampler were assessed. In this study, a sample extractor system was designed, which consists of a base, a hydraulic cylinder and a load cell. The objective of this equipment is to experimentally quantify the internal friction force, allowing the stresses acting on the sampler to be evaluated. Also, the Aoki´s *a* parameter, which is the ratio of internal friction and external friction between the ground and the sampler, could be calculated.

Keywords: SPT test; energy; efficiency; stress on the sampler; dynamic instrumentation; Aoki's *a* parameter.

Evaluación de las tensiones internas y externas que actúan sobre el muestreador del ensayo SPT

Resumen
El índice N_{SPT} es muy utilizado en correlaciones empíricas para estimar parámetros del suelo, capacidad de carga, asentamiento de fundaciones, etc. Debido a que estas correlaciones no tienen ningún fundamento científico, investigadores han desarrollado métodos racionales de análisis, basados en conceptos de energía. La cantidad de energía envuelta en el ensayo SPT es evaluada por el método EFV. Con los registros de fuerza y aceleración, pueden ser obtenidas las cantidades de energía presentes en el ensayo SPT, la eficiencia del equipo, la fuerza de reacción dinámica experimental del suelo y las tensiones que actúan sobre la cuchara partida. En este estudio, se diseñó un sistema de extracción de muestras, el cual se compone de una base, un cilindro hidráulico y una célula de carga. El objetivo de este equipo es cuantificar experimentalmente la fuerza de fricción interna, lo que permite determinar las demás fuerzas y tensiones que actúan sobre la cuchara partida. Además, se puede calcular el parámetro *a* de Aoki que es la relación de la fricción interna y externa entre el suelo y el muestreador.

Palabras clave: Ensayo SPT; energía; eficiencia; tensiones que actúan en el muestrador; instrumentación dinámica; parámetro *a* de Aoki.

1. Introduction

The standard penetration test (SPT) does not directly measure the shear strength of soils. Because of this, different authors have suggested empirical correlations to estimate relative density, friction angle and other soil parameters [1]. However, as these correlations have no scientific basis, researchers have been developing rational methods of analysis that are based on energy concepts.

The SPT (Standard Penetration Test), standardized in Brazil by ABNT-NBR 6484 (2001) [2], is usually used to estimate soil resistance through the dynamic resistance index (N_{SPT}). This index refers to the number of hammer blows required to penetrate the SPT sampler 0.30 m into the soil after an initial penetration of 0.15 m. The blows are produced by lifting the 65-kg hammer at a height of 0.75 m and dropping it in free fall, transmitting its potential energy (478.2 J) to the string of rods.

However, as this N_{SPT} index does not directly represent the real value of soil resistance, researchers have developed energy-based methods to estimate the theoretical static reaction force (R_S) and the theoretical dynamic reaction force (F_d) of the soil [3-5]. Both theoretical reactions forces and the experimental dynamic resistance (R_D), mobilized during the penetration of the sampler, are obtained from the energy transferred during the propagation of waves in the string of rods and the sampler.

This paper presents the results of SPT tests conducted in the Experimental Field of the University of Sao Paulo at Sao Carlos. The SPT tests were performed with an instrumented subassembly, which was placed at the base of the string of rods. The amount of energy transmitted to the sampler could be evaluated using measured accelerations and forces, which allowed the equipment efficiency to be calculated.

The accelerations and forces were measured through an instrumented subassembly, which is composed of two accelerometers and a load cell, and placed at the top of the sampler.

The internal friction force between the soil and the internal sampler wall was evaluated using a sample extracting system in order to subsequently assess the other stresses acting on the sampler. Finally, knowing these stresses, the Aoki's *a* parameter, which is the ratio of internal friction and external friction between the soil and the sampler wall, respectively, can be calculated.

2. The sample energy and reactions.

2.1. Energy and efficiency

The amount of energy transferred to the sampler (E_{base}), can be either theoretically or experimentally estimated. Experimentally, the amount of energy can be evaluated through the EFV method, which consists of the integration of the product of the normal force (F) and velocity (v) with respect to time (eq. 1). The velocities (v) were calculated by integrating acceleration with respect to time.

$$E_{base} = \int_{t_i}^{t_f} F \times v \, dt \tag{1}$$

where the t_i is the initial instant corresponding to the beginning of the event, that is, when the force becomes different from zero, and t_f is the final instant, when the force and velocity signals becomes zero and no subsequent impacts occur.

Theoretically, according to Odebrecht [6], the system's potential energy (EP_{system}) is the theoretical potential energy (EP_i) plus the potential energy of the hammer and rods due to the sampler penetration into the soil (eq. 2).

$$EP_{system} = EP_i + (\Delta\rho \times M_h \times g) + (\Delta\rho \times M_r \times g) \tag{2}$$

Where,
H= height of fall (m).
$\Delta\rho$= permanent sampler penetration (m).
M_r= rod mass (kg).
M_h= hammer mass (kg).
g= acceleration of gravity (m/s^2).

However, the amount of energy that reaches the sampler ($E_{sampler}$) is lower than EP_{system} due to energy losses that occur in the process of penetrating the sampler into the soil [7]. These energy losses depend on the type of equipment (type of hammer, the length of the string of rods, the geometry of the sampler, test procedures, i.e., mechanical or automatic, etc.), loose rod couplings and reflections due to cross section changes.

According to Odebrecht et al. [8], these losses are represented by three distinct coefficients: η_1 (hammer efficiency), η_2 (efficiency of the string of rods) and η_3 (system efficiency). The energy actually reaching the sampler is shown in eq. (3).

$$E_{sampler} = \eta_3[\eta_1(EP_i + (\Delta\rho \times M_h \times g)) + \eta_2(\Delta\rho \times M_r \times g)] \tag{3}$$

Where,

$$\eta_1 = \frac{\int_{t_i}^{t_f} E_{Base} \times v \, dt}{[(0{,}75 + \Delta\rho) \times M_h \times g]} \tag{4}$$

$$\eta_2 = 1 \tag{5}$$

$$\eta_3 = 1 - 0.0042 \, L_r \tag{6}$$

L_r = total length of the string of rods.

The equipment efficiency (η) is the ratio of the energy that reaches the sampler (E_{base}) and the system's potential energy (EP_{system}).

$$\eta = \frac{E_{base}}{EP_{system}} \tag{7}$$

2.2. Experimental dynamic reaction (R_D)

The experimental dynamic reaction for one blow (R_{di}) is the average of the resultant forces caused by the impacts during the time interval when the SPT sampler is penetrated into the soil (Fig. 1).

The experimental dynamic resistance (R_D) at a certain depth is defined as the average of the values of the experimental dynamic resistance for each blow, divided by the number of blows (N) for the last 0.3 m of the sampler's penetration into the soil (eq. 8).

$$R_D = \frac{\sum R_{di}}{N} \tag{8}$$

2.3. Reaction forces and stresses developed during the sampler penetration.

At the instant that the stress wave reaches the sampler, the soil reacts, producing an experimental dynamic force (R_D), which is constituted by the stresses and forces shown in Fig. 2. [4]

The balance of non-conservative vertical forces acting on the sampler (Fig. 2) is represented by:

$$R_D + W_h = R_1 + R_2 + R_3 + R_4 \quad (9)$$

Where:

$$R_1 = \pi \times D_{ext} \times (L_{ext} - L_p) \times r_l \quad (10)$$

$$R_2 = \pi \times D_{shoe} \times r_{li} \times L_{shoe} \quad (11)$$

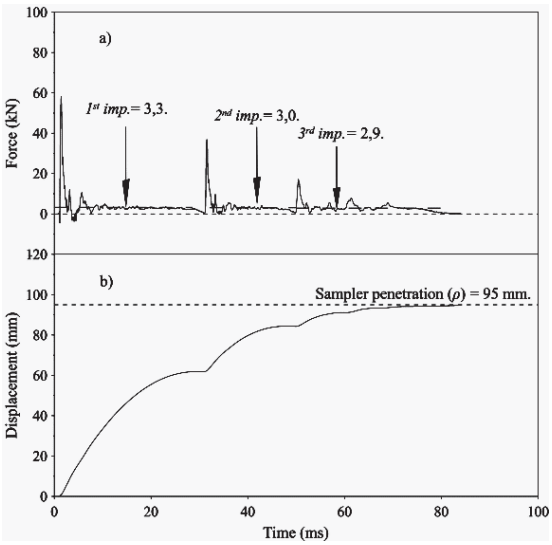


Figure 1. Force and displacement versus time for a typical record corresponding to a section just above the sampler (4th blow at the depth of 4 m).
Source: The authors.

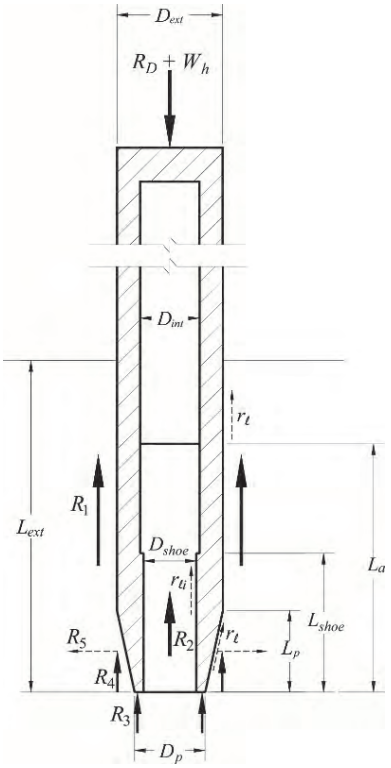


Figure 2 Reaction forces and stresses when the sampler is being penetrated into the soil.
Source: Adapted from [4]

Where,
D_{ext}= 50.8 mm
D_{int}= 34.7 mm
D_{shoe}= 32.0 mm
D_p= 34.7 mm
L_p= 21.5 mm
L_{shoe}= 50.36 mm
D_{ext}= sampler's external diameter.
D_{int}= sampler's internal diameter.
D_{shoe}= sampler's shoe internal diameter.
D_p= sampler's tip diameter.
L_p= sampler's tampered section length.
L_{ext}= sampler's total penetration.
L_a= sample's length.
L_{shoe}= sampler's shoe length.
r_l= friction on the sampler's outer surface and chamfer.
r_{li}= friction on the sampler's inner surface.
R_D= dynamic experimental reaction.
W_h= weight of the rods and anvil.
R_1= frictional force at the sampler's outer surface.
R_2= frictional force at the sampler's inner surface.
R_3= vertical reaction force on the annular section of the sampler's tip (may be considered negligible).
R_4 = vertical component of frictional force along the sampler's tampered surface.
R_5= horizontal component of frictional force along the sampler's tampered surface.

The balance of non-conservative vertical forces acting on the sampler (Fig. 2) is represented by:

$$R_D + W_h = R_1 + R_2 + R_3 + R_4 \quad (9)$$

Where

$$R_1 = \pi \times D_{ext} \times (L_{ext} - L_p) \times r_l \quad (10)$$

$$R_2 = \pi \times D_{shoe} \times r_{li} \times L_{shoe} \quad (11)$$

Following Aoki [3], the a parameter can be defined as the ratio of the internal and external friction:

$$a = \frac{r_{li}}{r_l} \quad (12)$$

Eq. (11) can be rewritten:

$$R_2 = \pi \times D_{shoe} \times a \times r_l \times L_{shoe} \quad (13)$$

Fig. 3 shows the forces acting on the sampler's open tip of. It can be seen from this figure that R_2 is equal to the resistance reaction acting in the open tip of the standard sampler (eq.14):

$$R_2 = \pi \times \frac{D_{int}^2}{4} \times r_p \quad (14)$$

Similarly to the CPT test, the friction ratio (R_f) is defined as the ratio of the external lateral friction and the tip resistance of the soil (r_p).

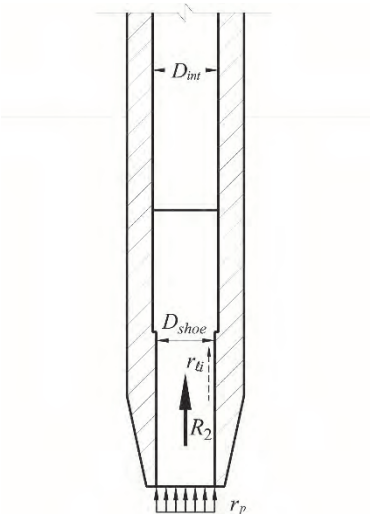


Figure 3 Forces acting on the open tip of the sampler.
Source: Adapted from [4]

$$R_f = \frac{r_l}{r_p} = \frac{r_{li}}{(a \times r_p)} \tag{15}$$

Combining eq. (11), (14) and (15), the following equation is obtained:

$$R_f = \frac{D_{int}^2}{4 \times a \times D_{shoe} \times L_{shoe}} \tag{16}$$

The R_3 and R_4 forces can be assessed by eq. (17) and (18).

$$R_3 = \frac{\pi}{4} \times \left(D_p^2 - D_{int}^2\right) \times \frac{r_l}{R_f} \tag{17}$$

$$R_4 = r_l \times \left(S_L \times \frac{L_p}{L}\right) \tag{18}$$

Where,

$$L = \left\{L_p^2 + \left[\frac{\left(D_{ext} - D_p\right)}{2}\right]^2\right\}^{0,5} \tag{19}$$

$$S_L = \frac{\pi \times L \times \left(D_{ext} + D_p\right)}{2} \tag{20}$$

Therefore,

$$r_l = \frac{R_D + W_h}{G + I + J + U} \tag{21}$$

Where,

$$G = \pi \times D_{ext} \times \left(L_{ext} - L_p\right) \tag{22}$$

$$I = \pi \times D_{shoe} \times a \times L_{shoe} \tag{23}$$

$$J = \frac{\pi}{4} \times \left(D_p^2 - D_{int}^2\right) \times \frac{1}{R_f} \tag{24}$$

$$U = S_L \times \frac{L_p}{L} \tag{25}$$

3. SPT equipment

The following is a list of the specifications of the SPT equipment that was used in this research:

- Tripod
- String of rods (24.3-mm inner diameter, 33.4-mm outer diameter. lengths of 1 to 2 m and unit weight of with 32.3-N/m)
- 65-kg pinweight hammer with a wood cushion.
- Anvil (90-mm height, 87.5-mm diameter and 3.3-kg weight)
- Rope (to lift the hammer)
- Conventional couplings (to connect the rods)
- Helical auger
- Raymond type split sampler

4. Instrumentation

4.1. Instrumentation for measuring forces and accelerations

The instrumented subassembly, which was developed by Lukiantchuki [9], consists of one segment of rod in which one load cell and two accelerometers are mounted. For more information about the instrumentation, data logger and software used in this research refer to Zapata [10].

4.2. Equipment for measuring the friction force between the soil and the sampler

An equipment to extract samples from inside the sampler and to evaluate the friction between the soil and the sampler was developed by Zapata [10]. The friction between the soil and the sampler can be evaluated by measuring the force R_2 needed to move the sample inside the sampler. The force R_2 is applied by a hydraulic cylinder and measured by a load cell. This procedure is performed for each meter of the SPT test.

This equipment consists of the following (Fig. 5):

- Naval aluminum base and supports for the sampler and the hydraulic cylinder
- Hydraulic cylinder to apply the sample extracting load
- Load cell (HBM model RSCBC3 with a maximum capacity of 500 kg) to measure the force applied to move the sample inside the sampler (Fig. 6)

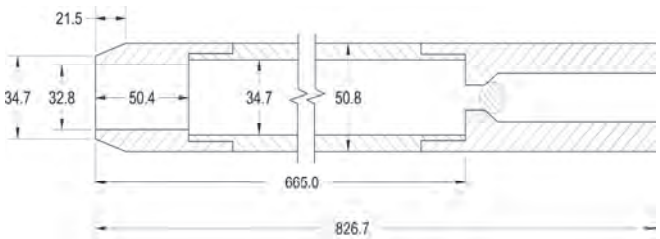


Figure 4 Dimensions (mm) of the sampler used in this research.
Source: The authors.

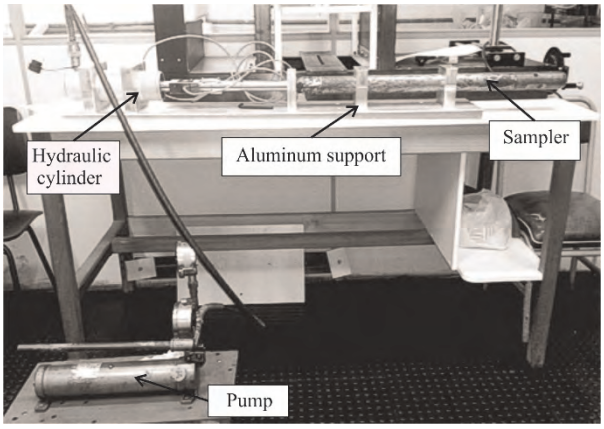


Figure 5 Equipment for measuring the internal friction force between the soil and the sampler.
Source: The authors.

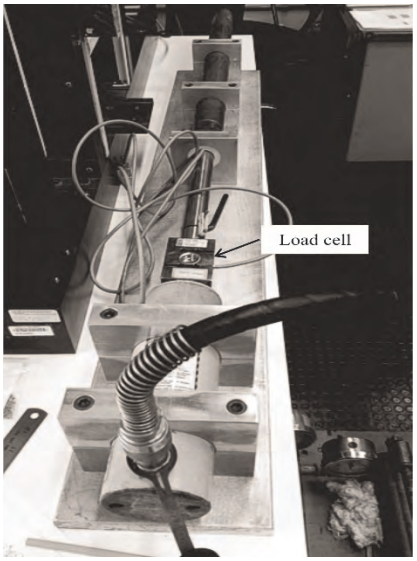


Figure 6 Load cell to measure the internal frictional force (R_2).
Source: The authors

The load cell force signals (R_2) are sent to and recorded by a portable data acquisition system (HBM model MX-410), which is controlled by the CatmanEasy software.

From the measured internal shear force R_2 and the experimental dynamic reaction (R_D) corresponding to a specific depth, the remaining reactions (tip, internal and external friction) that take place when the sampler is penetrated into the soil can be assessed. Furthermore, the Aoki´s *a* parameter and the friction ratio (R_f) are evaluated by using the equations described in section 2.3.

5. Results and discussions

5.1. Energy and efficiency

The amounts of energy and the efficiency values at the top of the sampler were determined for the blows that corresponded to the last 0.3 m of the total sampler penetration into the soil.

Table 1 shows the penetration of the sampler, the energy estimated at the bottom of the string of rods (top of the sampler, (E_{base}) and the system's potential energy (EP_{system}). The differences between the nominal potential energy (EP_i) values and the system's potential energy (EP_{system}) values (Table 1), show the importance of correcting the energy according to Odebrecht [6].

The system's potential energy that corresponds to higher resistance soil layers (N_{SPT} > 5 blows) is at least 8.5% higher than the nominal potential energy (478.2 J), whereas in the lower resistance soil layers (N_{SPT} < 3 blows); this percentage may be as high as 30.6%.

The estimated energy at the base of the string of rods and the system's potential energy curves show a parallel trend (Fig. 7). The distance between the two curves is the energy loss, which can be represented by efficiency (η).

Table 1
Energy and penetration results of the SPT test.

Depth (m)	Blow	L_r (m)	N_{SPT}	ρ (m)	E_{base} (J)	EP_{system} (J)
	2			0.114	393.50	555.55
1	3	1.28	2.69	0.106	429.24	550.13
	4			0.114	432.53	555.55
				Average	418.42	553.74
	2			0.148	411.94	583.30
2	3	2.28	2.01	0.151	432.08	585.43
				Average	422.01	584.37
	2			0.207	476.47	631.74
3	3	3.28	1.52	0.187	460.04	616.91
				Average	468.26	624.33
	2			0.116	417.01	567.94
4	3	4.28	2.99	0.09	403.09	547.83
	4			0.095	435.75	551.70
				Average	418.62	555.82
	2			0.091	336.44	551.49
5	3	5.28	2.97	0.109	389.68	565.98
	4			0.103	375.23	561.15
				Average	367.12	559.54
	2			0.135	438.69	591.18
6	3	6.28	3.03	0.075	350.26	540.99
	4			0.087	420.94	551.03
				Average	403.30	561.06
	3			0.062	363.47	532.07
	4			0.064	380.14	533.81
7	5	7.28	5.36	0.049	361.81	520.79
	6			0.048	359.87	519.92
	7			0.057	383.97	527.73
				Average	369.85	526.86
	2			0.081	364.66	551.14
	3			0.061	353.92	533.14
8	4	8.28	5.19	0.051	341.13	524.14
	5			0.046	345.76	519.64
	6			0.05	360.07	523.24
				Average	353.11	530.26
	4			0.045	339.98	520.16
	5			0.041	327.68	516.44
	6			0.044	329.81	519.23
9	7	9.28	6.89	0.05	362.33	524.82
	8			0.04	343.76	515.51
	9			0.037	330.33	512.71
	10			0.048	370.77	522.96
				Average	343.52	518.83
	3			0.047	320.68	523.52
	4			0.053	367.01	529.30
	5			0.048	364.58	524.48
10	6	10.28	6.82	0.046	362.67	522.55
	7			0.039	353.27	515.81
	8			0.036	335.27	512.92
	9			0.039	354.11	515.81
				Average	351.08	520.63

Source: The authors

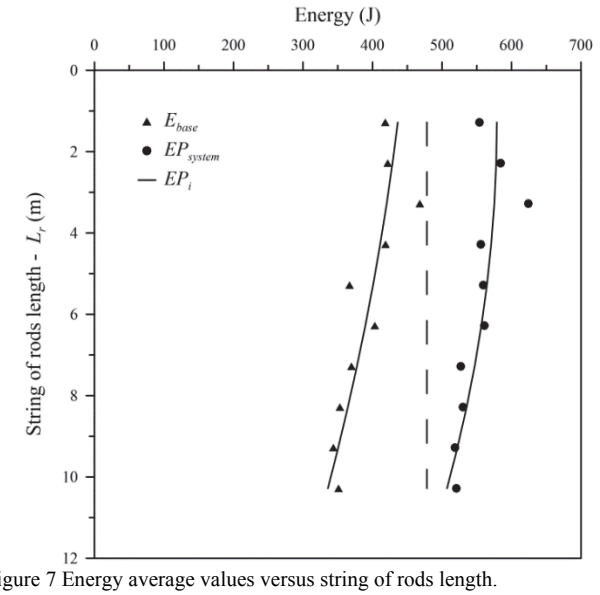


Figure 7 Energy average values versus string of rods length.
Source: The authors.

Table 2
Average efficiency values.

η (%)	σd (%)	cv (%)
70.59	2.46	3.51

Source: The authors.

According to Fig. 7, the shorter the length of the string of rods, the greater the system's potential energy and the greater the energy that reaches the sampler.

Table 2 shows the average values of efficiency for the tests conducted in this research as well as the standard deviation and the variation coefficient.

Table 3 displays the SPT test efficiency (η) values, the energy values that reach the sampler (according to Odebrecht et al. [8]), the hammer´s efficiencies (η_1), the string of rods' efficiencies (η_2) and the system's efficiencies (η_3). These parameters can be calculated by eq. (4), (5) and (6), respectively.

Table 3 shows the estimated energy at the base of the string of rods (E_{base}) is lower than the energy that reaches the sampler ($E_{sampler}$), calculated by eq. (3) for shorter string of rods. However, if the string of rods' length is increased, the values of E_{base} will be closer and closer to the sampler's energy values ($E_{sampler}$), possibly even exceeding them.

This is the case as the hammer efficiency (η_1) is always higher than the efficiency (η). For longer strings of rods, the influence of the system efficiency (η_3) is higher as it drops from almost 100% to 96%; as a consequence, the energy that reaches the sampler is decreased ($E_{sampler}$).

5.2. *Reactions and stresses developed during sampler penetration.*

The values of the force R_2 and other reactions acting on the sampler, in addition to Aoki's a parameter [4] and friction ratio (R_f), are shown in Table 4.

Table 3.
Results of the SPT test efficiency and hammer and system efficiencies according to Odebrecht et al. [8].

Prof. (m)	Blow #	L_r (m)	N_{SPT}	E_{base} (J)	η base (%)	$E_{sampler}$ (J)	η_1 (%)	η_3 (%)
	2			393.50	70.83	395.96	71.42	99.46
1	3	1,28	2.69	429.24	78.03	431.18	78.64	99.46
	4			432.53	77.86	434.78	78.51	99.46
			Average	418.42	75.57	420.64	76.19	99.46
2	2	2.28	2.01	411.94	70.62	418.56	71.94	99.04
	3			432.08	73.81	438.72	75.21	99.04
			Average	422.01	72.21	428.64	73.57	99.04
3	2	3.28	1.52	476.47	75.42	491.09	78.08	98.62
	3			460.04	74.57	472.85	77.00	98.62
			Average	468.26	75.00	481.97	77.54	98.62
	2			417.01	73.43	424.94	75.52	98.20
4	3	4.28	2.99	403.09	73.58	407.81	75.26	98.20
	4			435.75	78.98	440.53	80.87	98.20
			Average	418.62	75.33	424.43	77.22	98.20
	2			336.44	61.01	343.85	62.74	97.78
5	3	5.28	2.97	389.68	68.85	398.84	71.14	97.78
	4			375.23	66.87	383.74	68.99	97.78
			Average	367.12	65.58	375.48	67.62	97.78
	2			438.69	74.21	453.25	77.74	97.36
6	3	6.28	3.03	350.26	64.74	355.52	66.58	97.36
	4			420.94	76.39	426.67	78.87	97.36
			Average	403.30	71.78	411.81	74.40	97.36
	3			363.47	68.31	366.20	70.20	96.94
	4			380.14	71.21	382.80	73.24	96.94
7	5	7.28	5.36	361.81	69.47	361.67	71.01	96.94
	6			359.87	69.22	359.57	70.72	96.94
	7			383.97	72.76	384.95	74.62	96.94
			Average	369.85	70.19	371.04	71.96	96.94
	2			364.66	66.16	372.47	68.82	96.52
	3			353.92	66.39	357.04	68.44	96.52
8	4	8.28	5.19	341.13	65.08	342.16	66.79	96.52
	5			345.76	66.54	345.36	68.12	96.52
	6			360.07	68.82	360.19	70.59	96.52
			Average	353.11	66.60	355.44	68.55	96.52
	4			339.98	65.36	339.42	67.07	96.10
	5			327.68	63.45	326.47	64.97	96.10
	6			329.81	63.52	329.37	65.14	96.10
9	7	9.28	6.89	362.33	69.04	362.31	71.03	96.10
	8			343.76	66.68	341.64	68.24	96.10
	9			330.33	64.43	327.89	65.82	96.10
	10			370.77	70.90	369.86	72.87	96.10
			Average	343.52	66.20	342.42	67.88	96.10
	3			320.68	61.26	321.46	63.10	95.68
	4			367.01	69.34	367.66	71.68	95.68
	5			364.58	69.51	363.77	71.65	95.68
10	6	10.28	6.82	362.67	69.40	361.32	71.45	95.68
	7			353.27	68.49	350.15	70.22	95.68
	8			335.27	65.36	331.99	66.89	95.68
	9			354.11	68.65	350.95	70.39	95.68
			Average	351.08	67.43	349.62	69.34	95.68

Source: The authors

Table 4
Values of the reactions acting on the sampler during its penetration into the soil, the friction ratio Rf , and the Aoki's a parameter.

Prof. (m)	RD+Wh (kN)	a	Rf	R1 (kN)	R2 (kN)	R4 (kN)	r_t (kPa)	r_L (kPa)	rp (kPa)
1	3.0	0.2	0.8	2.8	0.0	0.1	37.1	8.8	48.6
2	2.6	0.2	0.9	2.4	0.0	0.1	34.0	7.1	39.3
3	2.3	2.9	0.1	1.9	0.3	0.1	20.2	58.3	321.0
4	3.4	1.0	0.2	3.1	0.2	0.1	42.3	41.3	227.1
5	3.7	0.7	0.2	3.4	0.2	0.1	49.8	37.3	205.2
6	3.8	0.9	0.2	3.5	0.2	0.1	47.0	43.8	241.3
7	5.2	1.2	0.1	4.6	0.4	0.2	61.4	74.5	409.9
8	5.5	0.5	0.3	5.1	0.2	0.2	68.7	36.1	198.9
9	6.3	1.0	0.2	5.7	0.4	0.2	76.9	78.3	430.7
10	6.2	0.6	0.3	5.8	0.3	0.2	80.8	48.3	265.8

Source: The authors

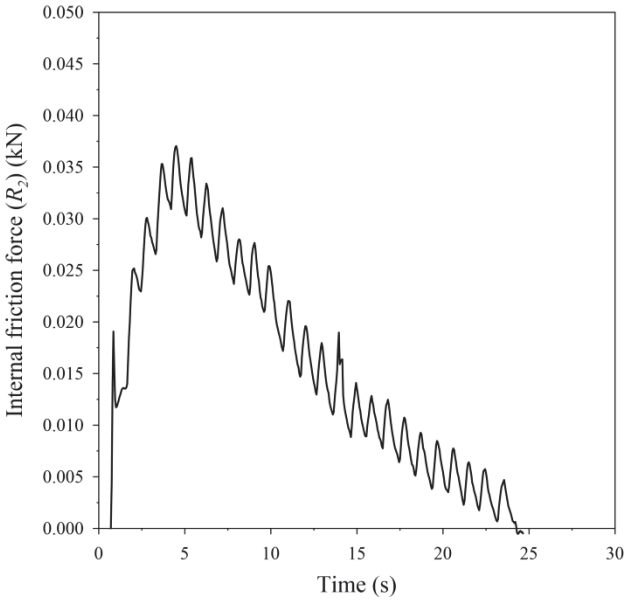


Figure 8 Internal friction force (R2) versus time to 2m deep.
Source: The authors

Fig. 8 shows a representative curve of the internal friction force versus elapsed time.

Table 4 shows that Aoki's *a* parameter values are approximately equal to one for any bore depth, with the exception of 1 and 2 m bore depths. This makes sense as the specific weight of the soil inside the sampler (γ_i) and the specific weight of the natural soil (γ_e), found in [11] are approximately equal (Table 5).

6. Conclusions

The differences between the nominal potential energy (EP_i) values and the system's potential energy (EP_{system}) values, shown in Table 1, clearly show the importance of correcting the energy according to Odebrecht [6].

Table 5.
Values of the specific weight of the soil inside the sampler and the natural ground.

Prof. (m)	γ_i (kN/m3)	γ_e (kN/m3)	γ_i / γ_e
1	18.7	15.6	1.2
2	14.4	15.5	0.9
3	16.3	15.8	1.0
4	15.2	16.9	0.9
5	16.4	17.2	1.0
6	17.5	17	1.0
7	17.6	18.3	1.0
8	16.8	19	0.9
9	17.3	18.4	0.9
10	19.4	18.9	1.0

Source: The authors.

The system's potential energy corresponding to soil layers of higher resistance ($N_{SPT} > 5$ blows) is at least 8.5% higher than the nominal potential energy (478.2 J), whereas in the lower resistance soil layers ($N_{SPT} < 3$ blows) this percentage increases up to 30.6%.

The estimated energy at the base of the string of rods and the system's potential energy curves show a parallel trend. The distance between the two curves is the energy loss, which may be represented by efficiency (η).

The average efficiency at the base of the string of rods is 70.59%, with a standard deviation of 2.46% and coefficient of variation of 3.51.

The estimated energy at the base of the string of rods (E_{base}) is lower than the energy that reaches the sampler ($E_{sampler}$) for shorter string of rods. However, if the string of rods' length is increased, the values of E_{base} will be closer and closer to the energy values of the sampler ($E_{sampler}$), possibly even exceeding them. This is the case as the hammer efficiency (η_1) is always higher than the efficiency (η). For longer strings of rods, the influence of the system efficiency (η_3) is higher as it drops from almost 100% to 96%; as a consequence, the energy that reaches the sampler is decreased ($E_{sampler}$).

The development of equipment for estimating the internal friction force (R_2) allowed the stresses acting on the SPT sampler to be evaluated.

The Aoki's *a* parameter values are approximately equal to one for any bore depth, with 1 and 2 m bore depths being the exception. These results show the efficacy of the new equipment that is designed to estimate the internal friction force (R_2).

Acknowledgements

The authors are very thankful to FAPESP (Grant No. 2008/08268-4) and CNPq (Grant No. 479001/2009-0) for their financial support. Furthermore, the first author is very thankful to CNPq for granting her a scholarship that made her Master's thesis possible.

References

[1] Diaz-Segura, E.G., Método simplificado para la estimación de la carga última de pilotes sometidos a carga vertical axial en arenas, DYNA, [Online]. 80(179), pp. 109-115, 2013. Available at: http://www.revistas.unal.edu.co/index.php/dyna/article/view/30280/40655.

[2] ABNT. Associação Brasileira de Normas Técnicas. Solo. Sondagens de simples reconhecimento com SPT. Método de ensaio. In: ABNT (NBR 6484). Associação Brasileira de Normas Técnicas. Rio de Janeiro, Brasil. 2001 pp. 3-12.

[3] Aoki, N. and Cintra, J.C.A., The application of energy conservation Hamilton's principle to the determination of energy efficiency in SPT tests. Proceedings of International Conference on the Application of Stress Wave Theory to Piles, 6., São Paulo, 1, pp.457-460, 2000.

[4] Aoki, N., Inovação no SPT, In: Cintra, J.C.A., Aoki, N., Tsuha, C. and Giacheti, H.L., Eds. Fundações, Ensaios estáticos e dinâmicos. São Paulo: Oficina de Textos, 2013. pp. 25-38.

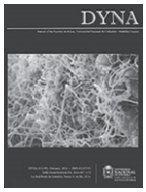
[5] Schnaid, F., Odebrecht, E., Rocha, M.M. ans Bernardes, G.P., Prediction of soil properties from the concepts of energy transfer in dynamic penetration tests. Journal of Geotechnical and Geoenvironmental Engineering - ASCE, 135(8), pp. 1092-1100, 2009. DOI: 10.1061/(ASCE)GT.1943-5606.0000059

[6] Odebrecht, E., Medidas de energia no ensaio SPT, PhD Thesis, Departamento de Engenharia Civil, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil, 2003.
[7] Odebrecht, E., Schnaid, F., Rocha, M.M. and Bernardes, G.P., Energy measurements for standard penetration tests and the effects of the length of rods. Geotechnical and Geophysical Site Characterization, pp. 351-358, 2004. ISBN 90 5966 009 9.
[8] Odebrecht, E., Schnaid, F., Rocha, M.M. and Bernardes, G.P., Energy efficiency for standard penetration tests. Journal of Geotechnical and Geoenvironmental engineering, ASCE, pp. 1252-1263, 2005.
[9] Lukiantchuki, J.A., Interpretação de resultados do ensaio SPT com base em instrumentação dinâmica. PhD. Thesis, Department of Geotechnics Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, SP, Brasil, 2012.
[10] Zapata, J.G., Estimativa das tensões internas e externas atuantes no amostrador SPT durante sua cravação. MSc. Thesis, Department of Geotechnical Engineering, Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, SP, Brasil, 2015.
[11] Peixoto, A.S.P., Estudo do ensaio SPT-T e sua aplicação na prática de engenharia de fundações. PhD. Thesis, Faculdade de Engenharia Agrícola, Universidade Estadual de Campinas (UNICAMP/FEAGRI). Campinas- SP, Brasil. 2001.

J. Zapata-Galvis, received her BSc. Eng. in Civil Engineering in 2011 from the Escuela de Ingeniería de Antioquia, and her MSc. in Geotechnical Engineering in 2015 from the University of Sao Paulo, Sao Carlos, Brazil. From 2011 to 2013 she worked in the field of geotechnical design in the area of foundations, slopes and retaining walls, soil profile analysis and testing with geotechnical instrumentation (SPT, inclinometers, pressuremeters and CPT).
ORCID: 0000-0001-9932-3325.

E.R. Esquivel, received his BSc. Eng. in Civil Engineering in 1971 and his MSc. in Structural Engineering, both from the University of São Paulo, Sao Paulo, Brazil. He earned his PhD. in Geotechnical Engineering from the University of Colorado, Boulder, USA. From 1972 to 1988, he worked for consulting companies that specialized in subway projects. From 1989 to 1995 he worked as research assistant at the University of Colorado USA. Since 2003 he has been a professor at the School of Engineering in Sao Carlos at the University of Sao Paulo, Sao Carlos, Brazil. His research interests include: sub-surface investigation, geotechnical instrumentation, soil-structure interaction, numerical and physical modeling of geotechnical structures, foundation engineering, geotechnical properties of natural and compacted soils and soil improvement.
ORCID: 0000-0002-2542-7081

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS



DYNA
http://dyna.medellin.unal.edu.co/

Environmental study of gold mining tailings in the Ponce Enriquez mining area (Ecuador)

Elizabeth Peña-Carpio [a] & Juan M. Menéndez-Aguado [a,b]

[a] *Escuela Superior Politécnica del Litoral (ESPOL), Guayaquil Ecuador, gepena@espol.edu.ec*
[b] *Escuela Politécnica de Mieres, Universidad de Oviedo, Oviedo, España.maguado@uniovi.es*

Received: July 7th, 2015. Received in revised form: December 15th, 2015. Accepted: January 25th, 2016.

Abstract
As a result of the artisanal mining activities in Ecuador (specially in Ponce Enriquez mining area, province of Azuay), great quantities of mining wastes have been disposed, usually without adequate management standards. Consequent acid rock drainage has been responsible of surface and underground water and surface water bodies. The aim of this study is to show a methodology which can characterize mining wastes to assist the taking decision process about remediation and/or ultimate disposition possibilities.

Keywords: mining tailings, mining environmental impact, acid drainage, remediation

Estudio de las colas de tratamiento de oro de la explotación minera en Ponce Enríquez (Ecuador) desde una perspectiva ambiental

Resumen
En todas las regiones mineras del Ecuador, especialmente en la zona minera de Ponce Enríquez (Azuay), se han depositado a lo largo del tiempo importantes volúmenes de residuos provenientes de las operaciones mineras artesanales, sí que por regla general se hayan considerado los estándares ambientales deseables en su disposición. Como consecuencia, la producción de Drenaje Ácido de Roca (DAR) ha sido una consecuencia habitual, produciendo la contaminación de cuerpos acuosos superficiales y/o subterráneos. El objetivo de esta investigación es contribuir a mejorar la disposición de las colas de los procesos de recuperación de oro mediante una metodología de caracterización geoquímico-ambiental de dichos residuos que sirva de base para la toma de decisiones sobre las alternativas de disposición final y/o remediación.

Palabras clave: colas mineras, impacto ambiental minero, drenaje acido, remediación.

1. Introducción

En todas las regiones mineras del Ecuador, se han depositado a lo largo del tiempo, grandes volúmenes de residuos provenientes de las operaciones mineras, con importantes contenidos en sulfuros (pirita, pirrotina). Dado que en raras ocasiones dicho residuo es gestionado conforme a los estándares deseables desde el punto de vista ambiental en su disposición, la acción de las aguas superficiales o de lluvia y el oxígeno de la atmósfera, van generando Drenaje Ácido de Roca (DAR) que produce la contaminación de cuerpos acuosos superficiales y/o subterráneos [1-4].

Este es el caso más habitual en el distrito minero del cantón Camilo Ponce Enríquez (provincia Azuay), en el que de forma casi general el manejo y disposición de relaves se realiza sin medidas ambientales acertadas en los diferentes procesos de beneficio mineral [5, 6]. En muchas ocasiones, debido a la baja eficiencia del proceso de tratamiento, incluso puede evaluarse la posibilidad de reprocesamiento del depósito de colas [7]

También es importante considerar que las operaciones mineras generalmente no concluyen con el cierre de operaciones de explotación, como debería ser una Responsabilidad Ambiental enmarcada en un Plan de Cierre y una Restauración de los sitios afectados, utilizando técnicas de remediación que tengan un carácter local [8], lo cual permitiría eliminar los riesgos, limitando la producción y propagación de contaminantes

a fin de generar un estado visual aceptable y compatible con un posible uso futuro.

Las pruebas de evaluación del Drenaje Ácido de Roca (DAR) [9-11] consisten en pruebas geoquímicas estáticas a partir de la caracterización química de los residuos y pruebas geoquímicas dinámicas en celdas húmedas, utilizadas para predecir el potencial neto de neutralización y la tasa de generación de ácido de residuos mineros, además de la calidad de los lixiviados generados en función del tiempo. Esta información es clave para definir el manejo ambiental de dichos residuos, evaluar el efecto de impacto ambiental que éstos están generando y establecer propuestas de remediación ambiental de los sitios mineros en los que han sido dispuestos.

Todas las pruebas geoquímicas a realizar están basadas en la Norma ASTM D5744-9; en el caso de las pruebas geoquímicas estáticas, la finalidad es determinar el poder de generación de ácido que podrían generar los residuos, basándose en el balance entre los minerales que pueden producir ácido (potencial de acidificación PA) y los minerales neutralizantes (potencial de neutralización PN) de la muestra. Las pruebas geoquímicas dinámicas, permiten evaluar la cinética de generación de drenajes ácidos y las cargas de metales pesados que pudieran aportar [12-14]

Así, el objetivo de esta investigación es contribuir a mejorar la disposición de las colas de los procesos de recuperación de oro en Ponce Enríquez mediante una metodología de caracterización geoquímico-ambiental de dichos residuos que sirva de base para la toma de decisiones sobre las alternativas de disposición final y/o remediación.

2. Metodología

2.1. Toma de muestras

En coordinación con el personal técnico de una de las sociedades mineras que operan en el Cantón de Ponce Enríquez, se tomaron muestras representativas de un dique de colas, el mismo que se encuentra cerca de la planta de procesamiento mineral.

Tomando en cuenta el volumen de este dique de relaves se consideró el criterio de "recolección de muestras de mayor sospecha por presencia de sulfuros"; lo que permitirá "predecir el escenario ambiental" más negativo; y por tanto, las "medidas a ser recomendadas" serán en el ámbito de los casos extremos.

2.2. Preparación de las muestras

Las muestras de los relaves fueron procesadas de acuerdo al flujograma presentado en la Fig. 1.

2.3. Caracterización de las muestras

La caracterización de la muestra de relaves mineros fue la siguiente:

2.3.1. Caracterización química

El Análisis Químico de la muestra preparada, pasó por la determinación de los siguientes parámetros: S_{TOTAL} – SO4=, As, Cd, Pb, Zn, Cu, Fe y otros (Ca, Mg). Los análisis se realizaron en el Laboratorio Certificado SPECTROLAB.

2.3.2. Caracterización mineralógica

A la muestra preparada se le realizó un proceso de separación magnética y luego por un proceso de flotación de sulfuros, cuyo concentrado sulfuroso "espumas" fue analizado por Difracción de Rayos X.

2.3.3. Caracterización Bacteriológica

Las muestras preparadas fueron sometidas a una determinación de pH en pasta a objeto de establecer el pH de hábitat de microorganismos.

2.4. Metodología de las Pruebas Geoquímicas Estáticas

La prueba geoquímica estática se basa en el balance del potencial de generación de acidez (AP) y el potencial de neutralización (NP), y es utilizado para determinar el ácido producido y ácido consumido por los componentes minerales que se encuentran en la muestra del residuo minero.

La diferencia entre estos dos valores, el potencial neto de neutralización (NNP) permite la clasificación de la muestra como un residuo minero generador o consumidor potencial de ácido.

Para facilitar la comparación de los valores del NP, AP y el PNN; estos expresados en unidades de toneladas de CaCO3 por tonelada de residuo minero.

El potencial de ácido es determinado por análisis del total de sulfuros y el cálculo del potencial ácido (AP) asumiendo una total conversión de los sulfuros a sulfatos, y que 2 moles de H+ son producidos por cada mol de pirita que se oxida.

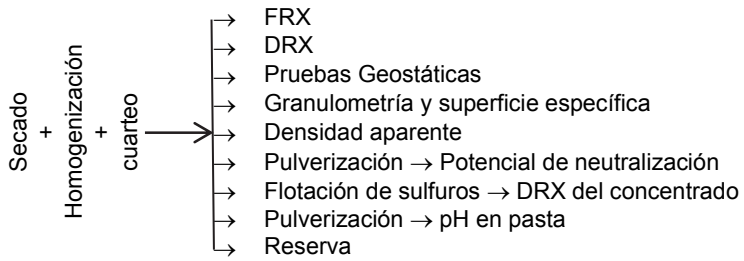


Figura 1. Flujograma del Tratamiento de la Muestra de Relave Minero Ponce Enríquez-Ecuador.
Fuente: Los autores.

El potencial de neutralización es determinado utilizando un procedimiento de titulación con un exceso de ácido clorhídrico, calentando para asegurarse que la reacción sea completa. La prueba FIZZ es empleada para conseguir que la cantidad de ácido adicionado sea suficiente para hacer reaccionar todos los minerales presentes en la muestra que son consumidores de ácido. El exceso de ácido es titulado por una base hasta el pH inicial de pasta del residuo; lo que permitirá calcular el carbonato de calcio que es equivalente al ácido consumido.

Finalmente, el potencial neto de neutralización (NNP), en toneladas CaCO3 equivalente por toneladas de material está dado por:

$$\mathbf{NNP = NP - AP}$$

La interpretación de los resultados obtenidos en las pruebas estáticas se desarrolla a partir de los dos criterios:

El **primer criterio**, determina el valor de NNP a partir de la diferencia entre el valor NP menos AP. De acuerdo al valor del NNP, el residuo podrá ser considerado como generador (cuando el valor de NNP es menor a -20 Kg de $CaCO_3$/tonelada) o no generador de DAR (cuando el valor de NNP es mayor a +20 Kg de $CaCO_3$/tonelada), o encontrarse en la zona de incertidumbre, hecho que precisa llevar a cabo una prueba geoquímica dinámica.

Mientras que, en el **segundo criterio**, se debe determinar la relación entre NP/AP. Si la relación es menor a 1, el residuo es generador de DAR; por el contrario, si la relación es mayor a 1, el residuo es estable químicamente.

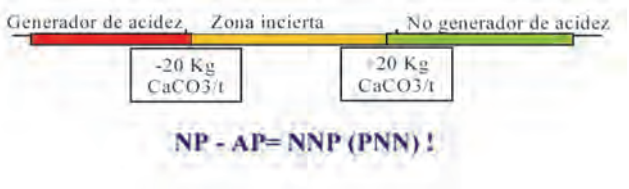


Figura 2. Primer criterio ABA.
Fuente: los autores.

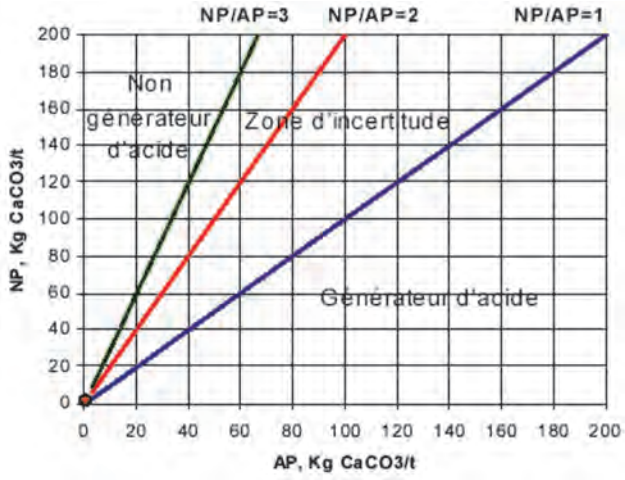


Figura 3. Segundo Criterio ABA.
Fuente: los autores

2.5. Metodología de las Pruebas Geoquímicas Dinámicas

Las pruebas geoquímicas dinámicas permiten someter a efectos de intemperismo extremo, muestras de residuos mineros, bajo condiciones controladas del laboratorio. Dichas pruebas se realizan en los casos en los que el valor del NNP, determinado mediante pruebas geoquímicas estáticas, se encuentra en la zona de incertidumbre; o en aquellos casos, en los que se desea determinar la tasa de generación de DAR y/o la carga de metales pesados que aportan los residuos al ecosistema circundante.

Es decir, mientras que las pruebas estáticas proporcionan información sobre las características de los desmontes y el potencial total para la generación de ácido, independientemente del tiempo; las pruebas cinéticas definen explícitamente las velocidades de reacción tanto para la disolución de metales, como para la generación de ácido y lixiviación, a través del tiempo y bajo condiciones específicas.

En consecuencia, las pruebas geoquímicas dinámicas son más caras y tardan más. Por tal razón, los objetivos de su realización, deben ser bien definidos con el fin de desarrollar un programa de pruebas eficiente y efectivo. Si bien se sabe que las pruebas dinámicas muestran el comportamiento geoquímico de la muestra para períodos cíclicos semanales, la predicción, a partir de los resultados obtenidos de la calidad de las soluciones de percolación, permite conocer el comportamiento del residuo en el futuro.

Existen diferentes tipos de pruebas que llamamos cinéticas; las diferencias entre ellas, están dadas por el tamaño de la prueba y la velocidad y cantidad de agua utilizada. Algunas, pueden ser realizadas en el terreno mismo (piezómetros, celdas de terreno); mientras que otras, se realizan en laboratorio, bajo condiciones que se acercan lo más posible a las del terreno. Asimismo, cada una tiende a ser utilizada para un tipo diferente de predicción.

Las pruebas cinéticas permiten además determinar el potencial de contaminación de un material cualquiera y de analizar escenarios de remediación como el efecto de un aditivo, la puesta en lugar de las coberturas, el efecto del nivel freático, etc.

En este estudio, se realizaron pruebas Geoquímicas Dinámicas en Celdas Húmedas [10,12]

Una celda de humidificación es una cámara climática diseñada para facilitar el control del ingreso del aire, temperatura y humedad; en ella, se simulan condiciones de intemperismo que asemejan las condiciones a las que el residuo minero se encuentra sometido en el lugar en el que se encuentra dispuesto.

Es decir, es un reactor que permite la oxidación de la muestra de residuo minero bajo condiciones extremas; generando así, información necesaria de predicción del comportamiento del mismo. El análisis físico-químico de los lixiviados recolectados permite predecir la tasa de generación de acidez del residuo minero estudiado; además de la calidad del drenaje generado en función del tiempo; y con ello, se puede entonces predecir el impacto ambiental que pueden generar los residuos sino se consideran medidas ambientales.

De las muestras preparadas, se determinó el potencial de generación de acidez (AP) y el potencial de neutralización (NP).

El potencial ácido AP de las muestras, en toneladas de

$CaCO_3$ por tonelada de residuo minero, fue determinado a partir de la siguiente relación: **AP** = Porcentaje de sulfuro x 31.25

El porcentaje de sulfuros se determinó a partir de la relación:

% Sulfuro = % STotal - % $SO_4^=$/2.996

El potencial de neutralización fue determinado utilizando un procedimiento de titulación con un exceso de ácido clorhídrico, calentando para asegurarse que la reacción sea completa. La prueba FIZZ es empleada para conseguir que la cantidad de ácido adicionado sea suficiente para hacer reaccionar todos los minerales presentes en la muestra que son consumidores de ácido. El exceso de ácido es titulado por una base hasta el pH inicial de pasta del residuo; lo que permitirá calcular el carbonato de calcio que es equivalente al ácido consumido.

Es decir, el potencial de neutralización NP de la muestra está dado por:

$$NP = \frac{50a\left[x - \left(b/a\right)y\right]}{c}$$

NP = potencial de neutralización en toneladas de $CaCO_3$ por tonelada.

A = normalidad del HCl.

B = normalidad de NaOH

c = peso de muestra en gramos

x = volumen del HCl añadido en ml.

Y = Volumen de NaOH añadido hasta pH 7.0 en ml.

Finalmente, el potencial neto de neutralización (NNP), en toneladas $CaCO_3$ equivalente por toneladas de material está dado por:

NNP = NP – AP (kg $CaCO_3$ / ton Residuo Minero)

La Fig. 4, muestra la batería de una celda húmeda montada para la realización de pruebas.

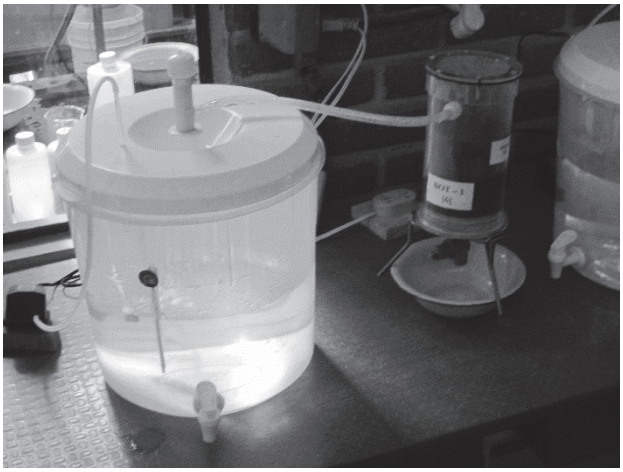


Figura 4. Celda de humidificación.
Fuente: los autores

Para la carga de cada una de las pruebas, se pesaron 500 gramos de muestra, preparada según el flujograma presentado en la Fig. 1.

El período de duración de cada ciclo fue de 7 días. Las pruebas se realizaron de acuerdo a la Norma ASTM (D 5744-96).

Después de la circulación de aire seco a través de la muestra durante 3 días, se hizo circular aire húmedo durante otros 3 días; para finalmente, efectuar el enjuague de la muestra al séptimo día con 500 ml de agua desionizada; el volumen de la solución de dicho proceso de enjuague fue determinado y, previa preservación, se enviaron las muestras a SPECTROLAB para ser sometidas al análisis físico-químico respectivo; para luego, iniciar un nuevo ciclo.

3. Resultados y discusión

3.1. Resultados de la caracterización química

En la Tabla 1, se presentan los resultados de la caracterización química de la muestra de relave objeto de estudio.

3.2. Resultados de la caracterización mineralógica

Con el fin de identificar las especies sulfuradas presentes en la muestra de relave, se realizó la flotación bulk de sulfuros y el concentrado de sulfuros para determinar con mayor precisión los componentes "acidogénicos" presentes en la muestra. Mediante análisis DRX sobre los productos obtenidos se obtuvieron los resultados mostrados en la Tabla 2.

También se determinó la densidad específica de la muestra, siendo esta de 2,5571 gr/cm3. En la Fig. 5, se muestra los resultados del análisis granulométrico de la muestra estudiada.

Tabla 1.
Análisis Químico de la Muestra de Relave Minero (Ponce Enríquez-Ecuador). Fuente: los autores

Parámetro	Valor (%)
Antimonio (Sb)	< 0,03
Arsénico (As)	0,12
Calcio (Ca)	1,22
Cadmio (Cd)	<0,01
Cobre (Cu)	0,13
Hierro (Fe)	11,51
Magnesio (Mg)	3,07
Plomo (Pb)	<0,03
Zinc (Zn)	0,06
Sulfatos (SO4=)	0,03
Azufre (S)	5,84

Fuente: los autores

Tabla 2.
Resultados del análisis DRX. (* minerales acidogénicos).

Mineral	Unidad	Valor
Esfalerita *	%	8
Pirita *	%	20
Calcopirita *	%	3
Arsenopirita *	%	35
Tetraedrita *	%	2
Pirrotina *	%	18
Oro (III) Teluro	%	3
Krennerita	%	3
GANGA		
Sílice	%	8

Fuente: los autores

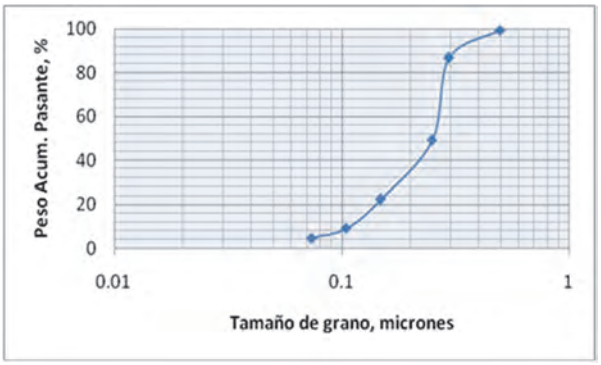


Figura 5. Análisis Granulométrico de la Muestra de Relave Minero.
Fuente: los autores

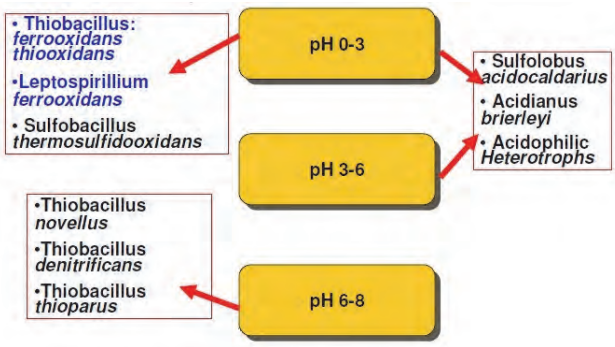


Figura 6.- Relación del valor de pH pasta y la presencia de microorganismos.
Fuente: los autores

Tabla 3.
Resultados del Potencial Neto y de Neutralización de la Prueba Geoquímica Estática.

Parámetro	Unidad	Valor
S total	%	5,84
Sulfato	%	0,03
Sulfuro	%	5,83
NP	kg $CaCO_3$/t	0,99
AP	kg $CaCO_3$/t	182,2
NNP	kg $CaCO_3$/t	-181,2
NP/AP	kg $CaCO_3$/t	0,0054

Fuente: los autores

3.3. Resultados de la caracterización bacteriológica

Se determinó el pH en pasta de la muestra de relave minero, que resultó ser de valor pH 5,5. La presencia de bacterias implicadas en la formación del drenaje ácido está directamente relacionada con el pH en pasta de acuerdo a la Fig. 6.

3.4. Resultados de las pruebas geoquímicas estáticas

Se determinó el Potencial Neutro (NP) y el Potencial Ácido (AP) de la muestra de relave minero con el fin de calcular el Potencial Neto de Neutralización (NNP). La Tabla 3, presenta los resultados de dichos cálculos.

Conforme al primer criterio mencionado, la muestra sería claramente generadora de drenaje ácido por ser el valor de NNP menor a -20 kg $CaCO_3$/t.

Aunque con este criterio ya sería suficiente, se llevó a cabo la aplicación del segundo criterio, según el cual si el valor de NP/AP es menor que la unidad, entonces el residuos sería considerado como generador de drenaje ácido, hecho que en este caso también se cumple con claridad.

3.5 Resultados de las pruebas geoquímicas dinámicas

Para el desarrollo de las pruebas geoquímicas dinámicas, se utilizó la celda húmeda construida y se realizaron ciclos de humidificación con una duración de 7 días por ciclo. Los resultados obtenidos de la muestra, se presentan a continuación.

3.5.1. Comportamiento Eh-pH-conductividad

Las gráficas siguientes muestran el comportamiento del potencial REDOX, el pH y la conductividad para cada ciclo de enjuague.

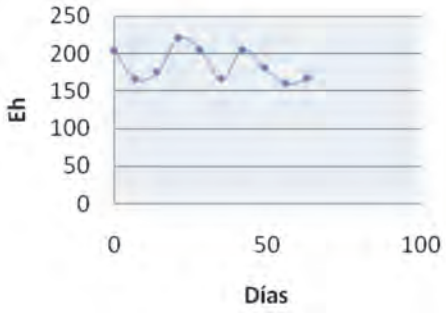


Figura 7.- Variación del Eh de la muestra.
Fuente: los autores

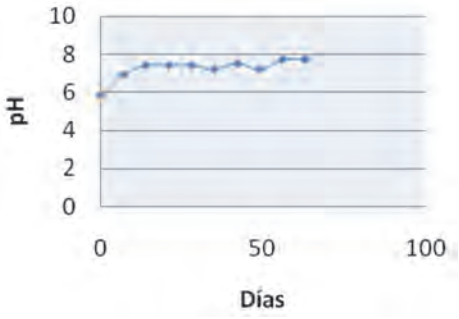


Figura 8.- Variación del pH de la muestra.
Fuente: los autores

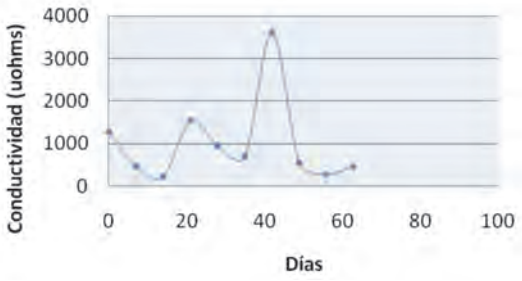


Figura 9.- Variación de la conductividad de la muestra.
Fuente: los autores

Tabla 4
Tasa de Generación de DAR de la Muestra.

Año	gr H_2SO_4/t
1	64,1841
2	58,4133
3	56,3716
4	56,3716
5	56,3716
6	57,1540
7	55,9960
8	57,1540
9	55,2741
10	55,2741
Total	572,5647

Fuente: los autores

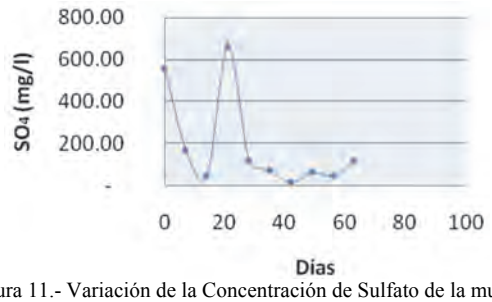


Figura 11.- Variación de la Concentración de Sulfato de la muestra.
Fuente: los autores

De las figuras anteriores es posible diferenciar tres etapas marcadas por el comportamiento del pH, Eh y la conductividad.

Teniendo en cuenta que los potenciales REDOX de las soluciones de enjuague de la muestra de relave minero objeto de estudio, se encuentran en el rango de 160 a 205 mV y los pHs entre 5.8 a 7.7. Por lo tanto se puede evidenciar en el diagrama Eh-pH del sistema Fe-S-H2O presentado en la Fig. 10, que el FeOOH es la fase estable, aunque no se puede descartar la formación de jarositas.

La Tasa de Generación de DAR, a partir de los valores del pH, es presentada en la Tabla 4. Estos valores han sido referidos a gramos de H_2SO_4 por tonelada de residuo minero.

3.5.2. Comportamiento de iones

La variación de la concentración de SO4= en función del tiempo, es presentada en la Fig. 11.

Como se puede observar en la Fig. 11, se tiene una tendencia descendente de las concentraciones de sulfato en las diferentes soluciones de enjuague en función del tiempo; aspecto que, podría hacer pensar que no se estaría generando drenaje ácido a partir del residuo minero estudiado. No hay que olvidar que la oxidación de los sulfuros, genera sulfato como producto de la reacción química.

Sin embargo, si representamos la eliminación del Sulfato vs Tiempo (Fig. 12); es decir, si le vamos restando la cantidad de sulfato disuelto a la cantidad inicial que existía en el mineral, veremos que se genera sulfato procedente de la oxidación de los sulfuros presentes.

Es decir, los valores negativos, expresan que existe un incremento con respecto al valor inicial; o sea que, se demuestra que el SO4= generado proviene de la oxidación de los sulfuros presentes, especialmente de la pirrotina y pirita.

La tendencia descendente de la gráfica, muestra que la oxidación de los sulfuros no ha sido agotada.

El poder neutralizante del mineral, puede confirmarse si se grafica el comportamiento del Calcio en función del tiempo. La representación de dicho comportamiento es presentado en la Fig. 13. En ésta, se puede corroborar que la concentración de iones calcio en las soluciones de enjuague disminuye de manera general en función del tiempo.

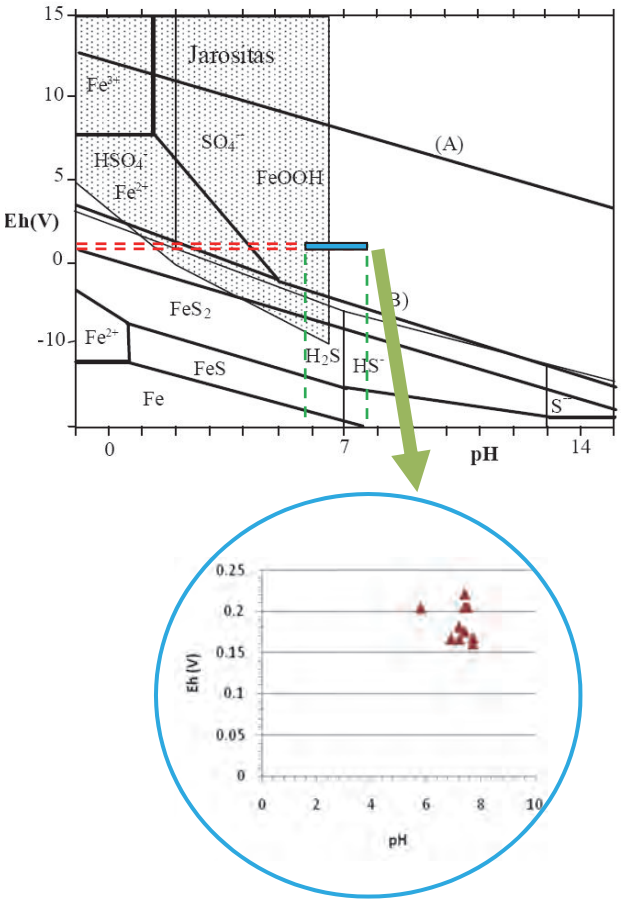


Figura 10.- Diagrama Eh – pH del Sistema Fe-S-H2O a 25º C.
Fuente: los autores

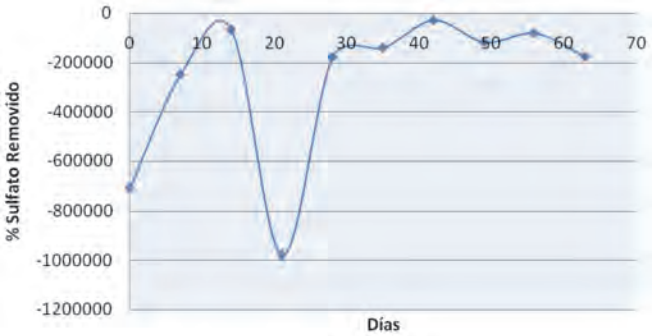


Figura 12.- Variación de la concentración de Sulfato de la muestra generadora de DAR.
Fuente: los autores

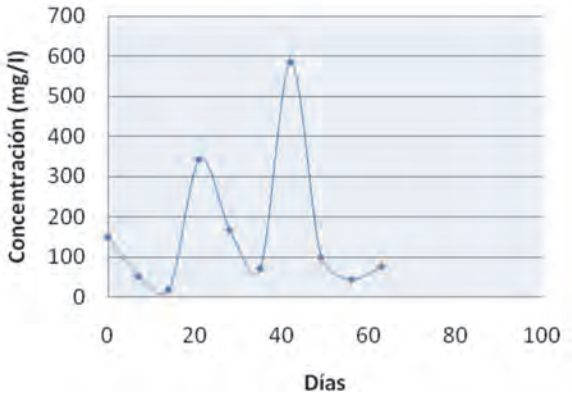


Figura 13.- Variación de la concentración de Calcio de la muestra.
Fuente: los autores

Las dos reacciones, en función del valor del pH, que explican la presencia de Ca^{+2} en las soluciones de enjuague son:

$$CaCO_3 + H \Leftrightarrow HCO_3^- + Ca^{++} \quad \text{(pH entre 6.4 y 10.3)}$$

$$CaCO_3 + 2H \Leftrightarrow H_2CO_3 + Ca^{++} \quad \text{(pH inferiores a 6.4)}$$

El $CaCO_3$ descrito en las dos reacciones anteriores, puede representar a los minerales acidívoros presentes en el residuo minero (roca de caja, aditivos de cal, u otros).

Es decir que, la acidez generada por las reacciones de oxidación de los sulfuros, principalmente pirrotina y pirita, inmediatamente es neutralizada por los minerales de la roca de caja o aditivos de cal, presentes en el residuo minero. Producto de dicha neutralización, los iones Ca+2 se encuentran en la solución de enjuague.

Por lo expuesto, deberá existir una correlación lineal entre la cantidad de Calcio Acumulado versus la cantidad de Sulfato Acumulado.

Finalmente, para completar con el análisis geoquímico, se muestra en la Fig. 14, la disolución de los metales pesados As, Pb, Cd, Cu, Fe y Zn, además en la Fig. 15, la disolución de Calcio.

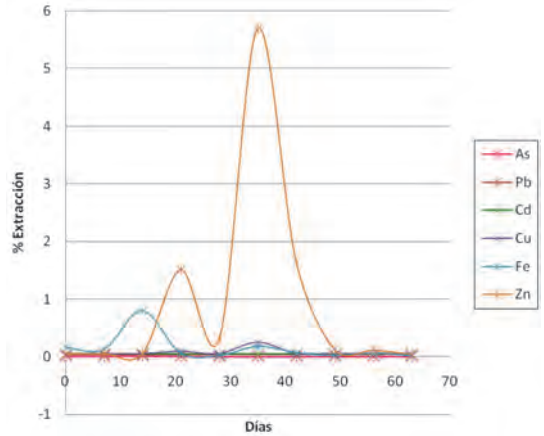


Figura 14.- Disolución de los metales pesados en las pruebas geoquímicas dinámicas.
Fuente: los autores

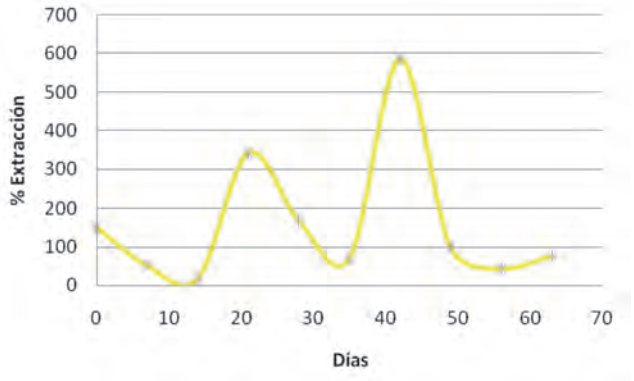


Figura 15.- Disolución Calcio en las pruebas geoquímicas dinámicas.
Fuente: los autores

La presencia de metales pesados en las soluciones de enjuague se debe probablemente a las posibles siguientes reacciones químicas:

$$CuFeS_2 + 4O_2 \Leftrightarrow Cu^{2+} + Fe^{2+} + 2SO_4^{2-}$$
$$CuFeS_2 + 16\,Fe^{3+} + 8H_2O \Leftrightarrow Cu^{2+} + 17Fe^{2+} + 2SO_4^{2-} + 16H^+$$

$$ZnS + 2O_2 \Leftrightarrow Zn^{2+} + SO_4^{2-}$$
$$ZnS + 8Fe^{3+} + 4H_2O \Leftrightarrow Zn^{2+} + 8Fe^{2+} + SO_4^{2-} + 8H^+$$

$$PbS + 2O_2 \Leftrightarrow Pb^{2+} + SO_4^{2-}$$
$$PbS + 8Fe^{3+} + 4H_2O \Leftrightarrow Zn^{2+} + 8Fe^{2+} + SO_4^{2-} + 8H^+$$

$$2\,FeAsS + 13/2\,O_2 + 3H_2O \Leftrightarrow 2(SO_4)^{2-} + 2(AsO_4)^{3-} + 6H^+ + 2Fe^{2+}$$
$$FeAsS + 11Fe^{3+} + 7H_2O \Leftrightarrow H_3AsO_3 + SO_4^{2-} + 11H^+ + 12Fe^{2+}$$

$$Fe_{1-x}S + (2-x/2)O_2 + xH_2O \Leftrightarrow (1-x)Fe_2 + SO_4^{2-} + 2xH^+$$
$$Fe_{1-x}S + (8-2x)Fe^{3+} + 4H_2O \Leftrightarrow (9-3x)Fe^{2+} + SO_2 + 8H^+$$

De las Figs. 14 y 15, es posible establecer lo siguiente:

- El As, Pb y Cd pueden pasar a la solución pero en baja proporción. Por lo que, su presencia en las soluciones de enjuague puede ser que sea en forma de complejos.
- El Fe, Cu y Zn son considerados como las mayores fuentes de generación de metales pesados. Dicha aseveración, corrobora los datos de las tasas de generación de acidez presentados en la Tabla 4.
- La tasa de disolución de Ca procedente del aditivo presente en el relave y de los minerales presentes en la roca de caja es limitado.
- Finalmente, las tasas de generación acumuladas por año de metales pesados por tonelada de residuo minero, se presentan en la Tabla 5.

3.5.3. Relación entre los elementos neutralizantes (Ca) y los productores de acidez (SO_4)

Para poder establecer la correlación entre los minerales acidívoros y los acidogénicos, es necesario representar el comportamiento de los valores acumulados de Ca frente a

Tabla 5
Tasa de generación acumulada de metales pesados de la muestra.

AÑO	Zn (g/t)	As (g/t)	Pb (g/t)	Cd (g/t)	Cu (g/t)	Fe (g/t)
1	0,2631	0,0315	0,1578	0,2631	0,2631	0,8421
2	0,4545	0,0590	0,2727	0,4545	0,4545	1,3181
3	0,7142	0,1095	0,4285	0,7142	0,7142	5,190
4	7,5454	0,1318	0,5454	0,9090	1,0909	5,2272
5	8,9545	0,1409	0,6818	1,1363	1,3181	5,3181
6	27,3928	0,1214	0,6428	1,0714	1,9285	4,8214
7	38,0000	0,1469	0,8571	1,4285	2,4081	5,7959
8	34,9259	0,1407	0,8888	1,4814	2,3703	5,3333
9	38,9387	0,1632	1,1020	1,8367	2,8163	6,1224
10	42,5777	0,1866	1,3333	2,2222	3,2888	6,7555

Fuente: los autores

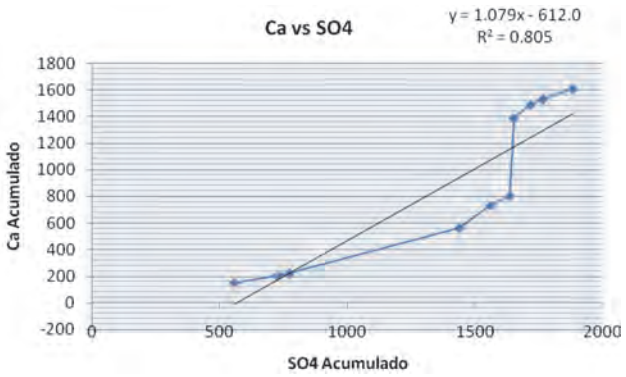


Figura 16.- Correlación entre Ca y SO_4 acumulados de las pruebas geoquímicas dinámicas.
Fuente: los autores

SO4=. La Fig. 16 muestra los resultados obtenidos, en la que también se muestra un ajuste lineal de los datos.

3.5.4 Interpretación final del potencial de acidificación de los residuos

Aplicando los criterios de clasificación de los residuos mineros descritos, dado que el relave minero objeto de estudio tiene un valor del Potencial Neto de Neutralización (NNP) menor a – 20 Kg $CaCO_3$/ton y una relación de NP/NA menor a la unidad, se puede considerar que los relaves en cuestión son generadores de drenaje ácido de roca, y por tanto químicamente inestable. Esto indicaría que en la etapa de operaciones mineras, se deben tomar las medidas de gestión de control del drenaje ácido y que en la fase de cierre de la operación minera, serán requeridas medidas de remediación.

4. Conclusiones

De la caracterización química y mineralógica del relave minero estudiado, se determinó que los minerales acidogénicos o generadores de drenaje ácido mayormente presentes en el residuo minero estudiado son pirrotina, pirita, esfalerita, arsenopirita y tetraedrita; también se detectó la presencia de oro (III), teluro y krennerita. Por otra parte, los minerales acidívoros o consumidores de drenaje ácido son minerales presentes en la roca encajante como los feldespatos o CaO. La ganga en su mayoría es sílice.

De la caracterización bacteriológica, se evidencia la presencia de poblaciones bacteriales de los géneros Sulfolobus acidocaldarius, Acidianus brierleyi, Acidophilic Heterotrophs; su valor de pH en pasta conduce a pensar que el relave se encontraría en la Fase I de formación de drenaje ácido.

Se realizó una estimación de la carga de metales pesados acumulada a 10 años, que aportaría el relave minero generador de DAR. Igualmente, mediante las pruebas geoquímicas se determinó el potencial ácido y el potencial neutro, así como el potencial neto de neutralización, lo que permitió aplicar los dos criterios y concluir por ambos que el residuo minero tiene un alto potencial de generación de drenaje ácido, por lo que se deben tomar las medidas oportunas tanto en las fases de operación como en las fases de cierre de la mina.

Agradecimientos

Los autores agradecen a la Secretaría Nacional de Educación Superior, Ciencia, Tecnología e Innovación (SENESCYT) del Gobierno de Ecuador por el apoyo recibido dentro del programa de becas Prometeo.

Referencias

[1] Nordstrom, D.K. and Alpers, C.N., Geochemistry of acid mine waters. The Environmental Geochemistry of Mineral Deposits, 6 A, pp. 133-160, 1999.

[2] Blowes, D.W., Ptacek, C.J. and Jurjovec, J., Mill tailings: Hydrogeology and geochemistry. In: Buckby, R.A., Otton, J.K. and Johnson, C.A., Eds., Environmental Aspects of Mine Wastes, Mineralogical Association of Canada, 31, 2003.

[3] Evangelou, V.P. and Zhang, Y.L., A review: Pyrite oxidation mechanisms and acid mine drainage prevention. Critical Reviews in Environmental Science and Technology, 25(2), pp. 141-199, 1995. DOI: 10.1080/10643389509388477

[4] Lindsay, M.B.J., Condon, P.D., Jambor, J.L., Lear, K.G., Blowes, D.W. and Ptacek, C.J. Mineralogical, geochemical, and microbial investigation of a sulfide-rich tailings deposit characterized by neutral drainage. Applied Geochemistry, 24(12), pp. 2212-2221, 2009. DOI: 10.1016/j.apgeochem.2009.09.012

[5] Carling, G.T., Diaz, X., Ponce, M., Perez, L., Nasimba, L., Pazmino, E., Rudd, A., Merugu, S., Fenandez, D.P., Gale, B.K. and Johnson, W.P., Particulate and dissolved trace element concentrations in three southern Ecuador rivers impacted by artisanal gold mining, Water, Air, and Soil Pollution 224(2), pp. 1415, 2013. DOI: 10.1007/s11270-012-1415-y

[6] Appleton, J.D., Williams, T.M., Orbea, H. and Carrasco, M., Fluvial contamination associated with artisanal gold mining in the Ponce Enríquez, Portovelo-Zaruma and Nambija Areas, Ecuador, Water, Air, and Soil Pollution 131(1), pp 19-39, 2001. DOI: 10.1023/A:1011965430757

[7] Tremolada, J., Dzioba, R., Bernardo-Sánchez, A. and Menéndez-Aguado, J.M., The preg-robbing of gold and silver by clays during cyanidation under agitation and heap leaching. International Journal of Mineral Processing, 94(1-2), pp. 67-71, 2010. DOI: 10.1016/j.minpro.2009.12.001

[8] Martínez, E.A., Tobón, J.I. and Morales, J.G., Coal acid mine drainage treatment using cement kiln dust. DYNA, 81(186), pp. 87-93, 2014. DOI: 10.15446/dyna.v81n186.38834

[9] Parbhakar-Fox, A.K., Edraki, M., Walters, S. and Bradshaw, D., Development of a textural index for the prediction of acid rock drainage. Minerals Engineering, 24(12), pp. 1277-1287, 2011. DOI: 10.1016/j.mineng.2011.04.019

[10] Sapsford, D., Bowell, R., Dey, M., Williams, C. and Williams, K., A comparison of kinetic nag tests with static and humidity cell tests for the prediction of ARD. Mine Water and the Environment, pp. 325-328. 2008.

[11] Verburg, R., Bezuidenhout, N., Chatwin, T. and Ferguson, K., The global acid rock drainage guide (GARD Guide). Mine Water and the Environment, 28(4), pp. 305-310, 2009. DOI: 10.1007/s10230-009-0078-4
[12] Standard Test Method for Accelerated Weathering of Solid Material Using A Modified Humidity Cell, Designation D 5744-96, 2001.
[13] Barazzoul, L., Sexsmith, K., Buckham, C. and Lopex, D., Application of an advanced mineralogical technique: Sulphide mineral availability and humidity cell interpretations based on MLA analysis. 9th International Conference on Acid Rock Drainage, 2012.
[14] Paktunc, A.D., Mineralogical constraints on the determination of neutralization potential and prediction of acid mine drainage. Environmental Geology, 39(2), pp. 103-112, 2000. DOI: 10.1007/s002540050440

E. Peña Carpio, es graduada en Ing. Geológica por la Universidad de Guayaquil, Ecuador y Dra. por la Universidad Politécnica de Madrid, España desde 1995, y Profesora Titular de la Facultad de Ingeniería en Ciencias de la Tierra de la Escuela Superior Politécnica del Litoral (ESPOL), Ecuador. Su principal ámbito de trabajo es el impacto ambiental de las operaciones de Mineralurgia y Metalurgia Extractiva.
ORCID: 0000-0001-9362-4506

J.M. Menéndez-Aguado, es Ing. de Minas y Dr. por la Universidad de Oviedo, España, donde es profesor desde el año 1999. Es profesor visitante de varias universidades en Argentina, Colombia, Cuba, Ecuador y Venezuela. Su principal ámbito de investigación es la Mineralurgia, con énfasis en la eficiencia energética en molienda, procesos de separación
ORCID: 0000-0002-6216-6984

DYNA
83 (195), February, 2016
is an edition consisting of 250 printed issues
which was finished printing in the month of January 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

- A multiobjective approach for non-discretionary variables in data envelopment analysis
- Kinetic aspects of a dried thin layer carrot in a heat pump dryer
- Factors influencing the occurrence of traffic accidents in urban roads: A combined GIS-Empirical Bayesian approach
- Crack growth in pyrographite under the conditions of radiation
- In vitro behavior of the dentin and enamel calcium hydroxyapatite in human premolars subjected to high temperatures
- Sparse representations of dynamic scenes for compressive spectral video sensing
- Influence of demand, control and social support on job stress. Analysis by employment status from the V European working conditions survey
- Saraz method adjustment for the quantification of ammonia emissions generated in opened or hybrid animal production facilities
- Implementation of a non-conventional method to characterize voltage sags and swells
- Influence of strain rate and heat treatments on tensile and creep properties of Zn-0.15Cu-0.07Ti alloys
- Non-destructive in situ analysis of garnet by combining scanning electron microscopy and X-ray diffraction techniques
- Simultaneous dual true random numbers generator
- Modeling waste management in a bioethanol supply chain: A system dynamics approach
- Fine material effect on kaolin suspensions rheology
- Numerical simulation to assess the elastic-strain energy distribution in a silicon rubber disk subjected to a punch shear test (PST)
- Optimal estimating the project completion time and diagnosing the fault in the project
- Fingerprint verification using computational geometry
- Managing relationships and communications in higher education efficiently through digital social networks: The importance of the relational coordination model
- Use of residual powder obtained from organic waste to partially replace cement in concrete
- Detection and localization of potholes in roadways using smartphones
- Forecasting of short-term flow freight congestion: A study case of Algeciras Bay Port (Spain)
- A mixed-integer linear programming model for harvesting, loading and transporting sugarcane. A case study in Peru
- Voltage sag assessment using an extended fault positions method and Monte Carlo simulation
- Conversion of an existing electrostatic precipitator casing to Pulse Jet Fabric filter in fossil power plants
- Design of an open source-based control platform for an underwater remotely operated vehicle
- Probabilistic seismic hazard analysis using a new ground motion intensity measure
- Mechanical and thermal performance of a geopolymeric and hybrid material based on fly ash
- Sol-gel synthesis of zinc oxide nanoparticle at three different temperatures and its characterization via XRD, IR and EPR
- Evaluation of internal and external stresses on the SPT sampler
- Environmental study of gold mining tailings in the Ponce Enriquez mining area (Ecuador)

- Un enfoque multiobjetivo para variables no discrecionales en análisis envolvente de datos
- Aspectos cinéticos del secado de capa delgada de zanahoria en un secador de bomba de calor
- Factores que influyen en la ocurrencia de accidentes de tránsito en vías urbanas: Un enfoque combinado GIS-Bayesiano empírico
- Crack growth in pyrographite under the conditions of radiation
- Comportamiento in vitro de la hidroxiapatita de calcio de la dentina y del esmalte en premolares humanos sometidos a altas temperaturas
- Representaciones dispersas de escenas dinámicas y reconstrucciones a partir de muestreo compresivo
- Influencia de la demanda, control y apoyo social sobre el estrés laboral. Análisis por situación laboral de la V Encuesta Europea sobre condiciones de trabajo
- Método Saraz ajustado para cuantificar emisiones de amoníaco generadas en instalaciones de producción animal abiertas o híbridas
- Implementación de un método no convencional para caracterizar hundimientos y elevaciones de tensión
- Influencia de la velocidad de deformación y tratamientos térmicos en las propiedades de tensión y fluencia del Zn-0.15Cu-0.07Ti
- Análisis no destructivos in situ de granate, combinando las técnicas de microscopía electrónica de barrido y difracción de rayos-X
- Generador dual simultáneo de números verdaderamente aleatorios
- Modelamiento de la gestión de residuos en la cadena de suministro de bioetanol, enfoque con dinámica de sistemas
- Efecto del material fino en la reología de suspensiones de caolín
- Simulación numérica para evaluar la distribución de energía de deformación en un disco de caucho siliconado sometido a un ensayo de corte por punzonado (PST)
- Estimarción optima de terminación el tiempo del proyecto y diagnóstico de la falla en el proyecto
- Verificación de huella dactilar utilizando geometría computacional
- Cómo gestionar las relaciones y comunicaciones de manera eficiente a través de las redes sociales digitales en instituciones de educación superior: Una propuesta desde el modelo de coordinación relacional
- Utilización de polvo de residuos orgánicos como sustitución parcial del cemento en el hormigón
- Detección y localización de imperfecciones viales utilizando smartphones
- Predicción a corto plazo de la congestión del flujo de mercancías: El caso de estudio del Puerto Bahía de Algeciras (España)
- Modelo de programación lineal entera mixta para el corte, carga y transporte de caña de azúcar. Un caso de estudio en Perú
- Evaluación de hundimientos de tensión mediante un método extendido de posiciones de falla y simulación de Monte Carlo
- Transformación de un precipitador electrostático a un filtro de mangas Pulse Jet en centrales térmicas convencionales
- Diseño de una plataforma de control basada en fuente abierta para un vehículo subacuático operado remotamente
- Análisis de peligro sísmico probabilístico usando una nueva medida de intensidad sísmica
- Comportamiento mecánico y térmico de un geopolímero y un material híbrido basado en ceniza volante
- Síntesis sol-gel de nanopartículas de óxido de zinc a tres termperaturas diferentes y su caracterización vía XRD, IR y EPR
- Evaluación de las tensiones internas y externas que actúan sobre el muestreador del ensayo SPT
- Estudio de las colas de tratamiento de oro de la explotación minera en Ponce Enríquez (Ecuador) desde una perspectiva ambiental

DYNA
Publication admitted to the
Sistema Nacional de Indexación y Homologación de
Revistas Especializadas CT+I - PUBLINDEX, Category A1

Red Colombiana de Revistas de Ingeniería

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN
FACULTAD DE MINAS