Ingeniería e Investigación
Journal
Abbreviated Journal Title: **Ing. Investig.**

Frequency
Quarterly, 3 issues per year
April, August and December

Cover Layout
Carlos Andrés Ortiz Valle

Proofreader
Janeth Alejandra García

Layout Artist
Patricia Chávez R.

Photography
Courtesy of the Dept. of Electrical and Electronic Engineering (Anonymous Author)

Printing
Corcas Editores S.A.S.

For additional information contact
revii_bog@unal.edu.co
Bogotá - Colombia
December - 2020

Table of Contents

Papers published in *Ingeniería e Investigación journal* are abstracted/indexed in

- Science Citation Index Expanded (SciSearch®), Clarivate Analytics
- Scopus - Elsevier
- Scientific Electronic Library Online - SciELO, Colombia
- Chemical Abstract
- Índice de Revistas Latinoamericanas en Ciencias Periódica
- Redalyc-Red de Revistas Científicas de América Latina y el Caribe, España y Portugal
- Dialnet
- Sistema Regional de Información en Línea para Revistas Científicas de América Latina, El Caribe, España y Portugal - Latindex
- Ebsco Publishing
- DOAJ - Directory of Open Access Journals
- Redib - Red Iberoamericana de Innovación y Conocimiento Científico

Ingeniería e Investigación journal was created in 1981. This is an entity in charge of spreading the teaching, scientific and technical research developed in the Universidad Nacional de Colombia's Engineering Faculty and other national and international institutions. *Ingeniería e Investigación journal* deals with original, unedited scientific research and technological developments in the various disciplines related to engineering. *Ingeniería e Investigación journal* contributes towards the development of knowledge, generating a global impact on academia, industry and society at large, through an exchange of knowledge and ideas maintaining a set of serious and recognized quality standards.

Ingeniería e Investigación journal publications are developed for the academic community who is interested in research and engineering knowledge development. We invite readers to be part of this Journal and participate either as authors, peer reviewers or subscribers.

For additional information contact:
www.revistas.unal.edu.co/index.php/ingeinv
E-mail: revii_bog@unal.edu.co
Tel: 57(1) 3 16 5000 Ext. 13674

Tabla de Contenido

Research Article / Civil - Sanitary Engineering http://dx.doi.org/10.15446/ing.investig.v40n3.83170

Bearing Capacity and Settlement Prediction of Multi-Edge Skirted Footings Resting on Sand

Capacidad de Carga y Predicción de Asentamiento de Zapatas Bordeadas de Bordes Múltiples que Descansan Sobre Arena

Tammineni Gnananandarao[1], Vishwas Nandkishor Khatri[2], and Rakesh Kumar Dutta[3]

ABSTRACT

This paper presents the application of artificial neural networks (ANN) and multivariable regression analysis (MRA) to predict the bearing capacity and the settlement of multi-edge skirted footings on sand. Respectively, these parameters are defined in terms of the bearing capacity ratio (BCR) of skirted to unskirted footing and the settlement reduction factor (SRF), the ratio of the difference in settlement of unskirted and skirted footing to the settlement of unskirted footing at a given pressure. The model equations for the prediction of the BCR and the SRF of the regular shaped footing were first developed using the available data collected from the literature. These equations were later modified to predict the BCR and the SRF of the multi-edge skirted footing, for which the data were generated by conducting a small scale laboratory test. The input parameters chosen to develop ANN models were the angle of internal friction (ϕ) and skirt depth (D_s) to the width of the footing (B) ratio for the prediction of the BCR; as for the SRF one additional input parameter was considered: normal stress (σ). The architecture for the developed ANN models was 2-2-1 and 3-2-1 for the BCR and the SRF, respectively. The R^2 for the multi-edge skirted footings was in the range of 0,940-0,977 for the ANN model and 0,827-0,934 for the regression analysis. Similarly, the R^2 for the SRF prediction might have been 0,913-0,985 for the ANN model and 0,739-0,932 for the regression analysis. It was revealed that the predicted BCR and SRF for the multi-edge skirted footings with the use of ANN is superior to MRA. Furthermore, the results of the sensitivity analysis indicate that both the BCR and the SRF of the multi-edge skirted footings are mostly affected by skirt depth, followed by the friction angle of the sand.

Keywords: square/circular skirted footings, multi-edged skirted footings, bearing capacity ratio, settlement reduction factor, artificial neural networks, multivariable regression analysis

RESUMEN

Este documento presenta la aplicación de redes neuronales artificiales (ANN) y el análisis de regresión multivariable (MRA) para predecir la capacidad de carga y el asentamiento de las zapatas bordeadas de bordes múltiples en arena. Estos parámetros se definen, respectivamente, en términos de la relación de capacidad de carga (BCR) de carga de la zapata con zócalo y sin zócalo y el factor de reducción de asentamiento (SRF), la razón de la diferencia en la solución de zócalo sin zócalo y zapatas bordeadas para el asentamiento de zapatas sin falda a una presión determinada. Las ecuaciones modelo para predecir la BCR y el SRF de la zapata de forma regular se desarrollaron primero utilizando los datos disponibles recopilados de la literatura. Estas ecuaciones se modificaron posteriormente para predecir la BCR y el SRF de la zapata bordeada de bordes multiples, para la cual se generaron los datos mediante la realización de una prueba de laboratorio a pequeña escala. Los parámetros de entrada elegidos para desarrollar modelos ANN fueron el ángulo de fricción interna (ϕ), la profundidad del faldón (D_s) al ancho de la relación de zapata (B) para la predicción del BCR; en cuanto al SRF, se consideró un parámetro de entrada adicional: la tensión normal (σ). La arquitectura para los modelos ANN desarrollados fue 2-2-1 y 3-2-1 para la BCR y el SRF, respectivamente. El R^2 para las zapatas bordeadas de bordes múltiples estuvo en el rango de 0,940-0,977 para el modelo ANN y 0,827-0,934 para el análisis de regresión. De manera similar, el R^2 para la predicción del SRF pudo haber sido de 0,913-0,985 para el modelo ANN y 0,739-0,932 para el análisis de regresión. Se reveló que la BCR predicha y el SRF para las zapatas con borde de múltiples bordes con el uso de ANN es superior al MRA. Además, los resultados del análisis de sensibilidad indican que tanto el BCR como el SRF de las zapatas bordeadas de bordes múltiples se ven más afectados por la profundidad de la falda, seguida del ángulo de fricción de la arena.

Palabras clave: zapatas de zócalo cuadradas/circulares, zapatas bordeadas de bordes múltiples, relación de capacidad de carga, factor de reducción de liquidación, redes neuronales artificiales, análisis de regresión multivariable

Received: October 26th, 2019
Accepted: July 30th, 2020

[1]Civil Engineering Andhra University, M.E. Andhra University, Ph.D. National Institute of Technology Hamirpur, India. Affiliation: Research Scholar, Civil Engineering Department, NIT Hamirpur, India. E-mail: anandrcwing@gmail.com
[2]Civil Engineering Shivaji University, M.Tech. Indian Institute of Science, Bangaluru, Ph.D. Indian Institute of Science, Bengaluru, India. Affiliation: Assistant professor, Department of Civil Engineering, Indian Institute of Technology, Dhanbad, India. Email: vishuiisc@gmail.com
[3]Civil Engineering K.N.I.T. Sultanpur, M.Tech. University of Rookee, Ph.D. Indian Institute of Technology Delhi, India. Affiliation: Professor, Department of Civil Engineering, NIT Hamirpur, India. E-mail: rakeshkdutta@gmail.com

How to cite: Gnananandarao, T., Khatri, V. N., and Dutta, R. K. (2020). Bearing capacity and settlement prediction of multi-edge skirted footings resting on sand. *Ingeniería e Investigación*, *40*(3), 9-21. 10.15446/ing.investig.v40n3.83170

Introduction

The prediction of the bearing capacity and the footing settlement with a reasonable accuracy is required in the field of foundation design and affects the overall economy of a project. For this purpose, the bearing capacity equations for conventional footings such as strip, square, circular, and rectangular were well established and are available in the literature. However, different, unconventional geometries for shallow footings were sometimes required for the economy and architectural matters. Numerical analysis using FLAC 3D software was carried out by Ghazavi and Mokhtari (2008) to observe the failure behavior of the sand beneath multi-edge shallow footings. The laboratory tests were performed by Davarci, B., Ornek, M., and Turedi, Y. (2014) on multi-edge footings and reported that the performance of such footings was better than that of square ones of the same width. Furthermore, for these types of footings, no bearing capacity and settlement equations were available in the literature. Hence, the researchers had to resort to experimental research, but generating this kind of data is very expensive. An alternative approach could be building a mathematical model to understand the relationships between the various parameters by calibrating and fitting the generated experimental data. To this effect, the power of ANN to store, learn and capture the complicated relationships between multiple parameters without any prior assumptions is the best choice in determining the bearing capacity ratio and the settlement reduction factor. Hence, the paper presents the application of ANN to model such parameters in unconventional multi-edge skirted footings on the sand.

Background

There have been several studies with ANN in the geotechnical engineering, such as those related to bearing capacity and the settlement of regular shaped footings in different mediums (Kalinli, Acar, and Gunduz, 2011; Marto, Hajihassani, and Momeni, 2014; Ziaee, Sadrossadat, Alavi, and Shadmehri, 2015; Nazir et al., 2015a; Rezaei, Nazir, and Momeni, 2016; Momeni, Armaghani, Fatemi, and Nazir, 2017; Khudier, 2018). The bearing capacity of the footing on sand was predicted by Kalinli et al. (2011) using ANN based on 97 datasets by varying the footing width, embedment depth, geometry, unit weight of sand, and the friction angle of the cohesionless soil. On similar lines with Kalinli et al. (2011), other researchers have explored the application of ANN on the prediction of the bearing capacity of footings resting on sand and rock. The list of these references with the scope of their work, type of material, type of data collected, settlement to width ratio, data set, and input parameter is shown in Table 1.

Note that, for the references listed in Table 1, the dataset in various cases ranged between 75 and 150. Similarly, some papers predicted the ultimate bearing capacity and settlement of the piles using experimental data with help from different soft computing techniques (Nazir, Momeni, Marsono, and Sohaie, 2013; Nazir et al. 2015b; Momeni et al. 2015a; Harandizadeh, Armaghani, and Khari, 2019; Chen et al. 2020; Khari et al. 2020; Yong et al. 2020). Furthermore, the results obtained from all the studies above indicated that ANN-based predictive models could be satisfactorily used in predicting the bearing capacity and the settlement of regular shaped footings. In the present paper, the data for regular shaped (square and circular) skirted footings were collected from the published literature, whereas the data for the multi-edge (T, Plus, Double box) skirted footings were generated through experimentation in the laboratory. The data was used during training, testing, and the predictive phase of ANN models. Input variables for the bearing capacity ratio were skirt depth to width ratio of the footing and friction angle of the sand for ANN modeling. Similarly, to model the settlement reduction factors, the considered input variables were skirt depth to width ratio of the footing, friction angle of the sand, and normal stress. The outputs for these ANN models were bearing capacity ratio and the settlement reduction factor, respectively.

Table 1. Dataset and parameters varied in the literature for the development of ANN model

References	Input parameters	Type of material	Type of data collected	Output parameters	Dataset (No.)
Kalinli et al. (2011)	Width of the footing, embedment depth of the footing, footing geometry, unit weight of sand, friction angle of the cohesionless soil	Sand	Field/laboratory both	Bearing capacity	97
Nazir et al. (2013)	Footing length, footing width, embedded depth of the footing, average vertical effective stress of the soil at B/2 below the footing, friction angle of the soil, where B is the width of the footing	Sand	Laboratory only	Bearing capacity	75
Momeni et al. (2017)	Width of the footing, sand friction angle, unit weight of the sand, and footing thin-wall ratio	Sand	Laboratory only	Bearing capacity	150
Khudier (2018)	Liquid limit, plasticity index, percentage of fines and sand, optimum moisture content, sulfur trioxide, total suspended solids, chlorine, and gypsum	Sand	Laboratory only	Bearing capacity	87

Source: Authors

Experimental Materials and Methods

As mentioned earlier, the data related to the bearing capacity and the settlement of multi-edge skirted footings on the sand was generated by performing experiments in the laboratory. All tests were performed in s tank (700 mm length x 450 mm width x 600 mm depth), which was prepared with a 15 mm thick perspex sheet, stiffened by the mild steel plate as shown in Figure 1.

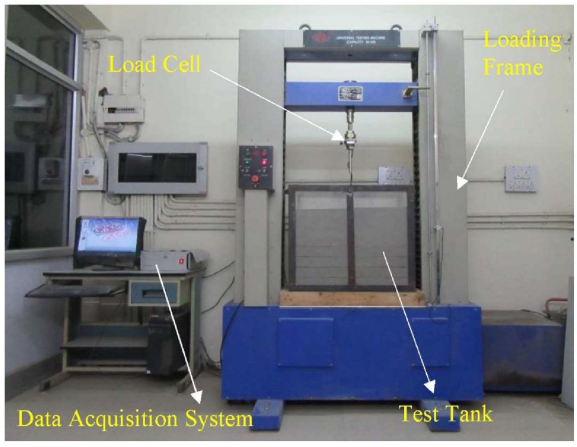


Figure 1. Complete setup of the testing program.
Source: Authors

The 10 mm thick steel footings with plan dimensions 80 mm x 80 mm were prepared in the shape of a square, T, plus, and double box, as shown in Figure 2. The skirts with 5 mm thickness and depths between 0,25 B and 1,5 B were firmly welded to these footings to simulate skirted footings, as shown in Figure 3.

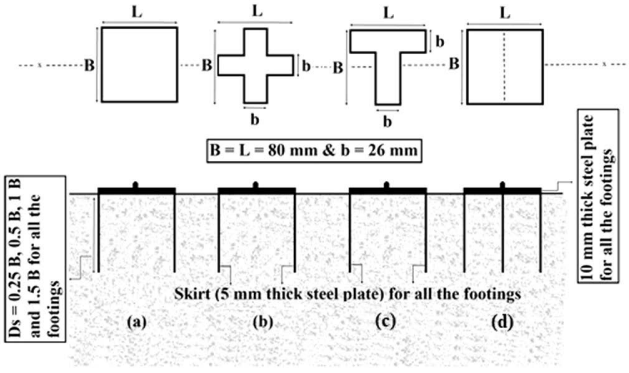


Figure 2. Plan and a sectional view of various shapes of unskirted and skirted footings: (a) square, (b) Plus, (c) T shape, (d) Double box.
Source: Authors

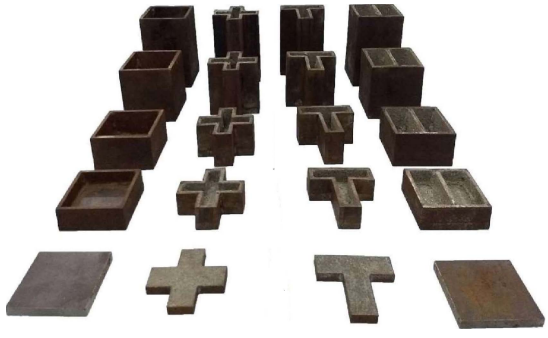


Figure 3. Photograph of model footings.
Source: Authors

In the case of rough footings, the sand was pasted to the base of the footing and also to the inner surfaces of the skirt; while, as for partly rough footings, a machined surface was used as it was. Beas river sand was used in this research and had a specific gravity (2,67), coefficient of uniformity (C_u = 1,46), coefficient of curvature (C_c = 0,98), and minimum and maximum dry unit weights of 13,06 kN/m^3 and 15,97 kN/m^3, respectively. As per IS 1498, the sand was classified as poorly graded (SP). The consolidated drained triaxial friction angle of sand at relative densities of 30%, 40%, 50%, and 60% was measured as 36,06°, 38,64°, 39,86°, and 41,72°, respectively. It is pertinent to note that due to the dilatancy effect, the friction angle of sand depends on the prevalent stress level or confining pressure in the test. Hence, it is likely that the friction angle developed in the small scale laboratory test, such as this one, will be much greater than the mobilized friction angle for the field size footing at failure with similar loading and soil conditions. Therefore, the predicted bearing capacity for the field size footing will be generally higher if extrapolated linearly from the results of the small-scale laboratory test data. This aspect is referred to as the size or scale effect on the bearing capacity (Chakraborty and Kumar 2013; Tang et al. 2014). To study this effect, it is necessary to carry out the load test with different footing sizes, which is a cumbersome task, especially with respect to multi-edge skirted footings. Hence, in the present study, no attempt has been made to study it. This implies that care is paramount when extrapolating the results of the present study for the field size footings. The tank was filled with sand to achieve the targeted relative density, which was varied from 30% to 60% in this investigation. The sand bed was placed in the tank in 8 equal 60 mm thick layers. The weight of the sand in each layer was calculated corresponding to the required relative density by using the unit weight of the sand and the volume of the layer. The weighed sand was then poured from a constant height and was compacted using a wooden 6 N rammer by giving several blows (obtained using the trial and error method) to reach the required relative density. It was ensured that the difference in measured relative densities was within ±1%. This was achieved by placing four steel bowls of known volume in each of the layers and measuring the achieved relative density of sand in them. As shown in Figure 1, the test on the prepared sand bed was performed with a strain-controlled loading frame of 50 kN and the employment of a load cell of 5 kN capacity. Note that, rather than measuring the pressure below the footing, the load applied on the top of the footing was recorded with a data acquisition system. This implies that this measured load, divided by the plan area of the footing, represents the magnitude of the average uniform pressure below the footing. Furthermore, any non-linear variation of normal stress was not considered (Kumar, 2009). This assumption is in linewith the literature, wherein the magnitude of the bearing capacity is defined based on the average uniform pressure. All the tests were performed using a 0,24 mm/min strain rate. It was intended that at this low strain rate, it would be possible to capture the non-linear pressure-settlement behavior. It was assumed that the measured penetration depth of the applied load would be a reasonable

representation of the immediate settlement of the footing. However, further study is required to confirm this, since it is beyond the scope of the present work. After the completion of each test, the pressure-settlement curve was plotted, and, subsequently, the bearing capacity and the settlement at a given pressure were calculated. It was anticipated that the applied load on the skirted footing would be resisted by the shear resistance mobilized along the skirt-soil interface, along with the shear strength of the sand. The bearing capacity of the skirted footing was denoted in terms of the BCR. Further, the reduction in settlement of the footing, due to the provision of the skirts, at a given pressure, was expressed in terms of settlement reduction factor. For the development of the models (ANN and MRA), for the prediction of the BCR and the SRF, the data from the literature on square and circular footings was collected. After the development of the models for regular-shaped skirted footings, the models were extended for the prediction of the BCR and the SRF of T, Plus, and double box shapes by means of simple multiplication factors. Finally, the ANN and MRA model predictions were compared to bring out the superiority of the ANN over the MRA model.

Test Results

Pressure settlement curve and Bearing capacity variation of footing with skirt depth

The pressure-settlement behavior of the footings (skirted and unskirted) with different plan shapes corresponding to a relative density of 30% is shown in Figure 4. It is pertinent to mention here that the ultimate bearing capacity was taken with respect to the minimum of the peak pressure or the pressure corresponding to an s/B ratio of 10% on the pressure settlement curve. The settlement ratio was used for the calculation of the settlement of the footing. Furthermore, if a clear peak in the pressure-settlement ratio curve was not obtained, the ultimate bearing capacity was calculated by using a double tangent method.

The results obtained in the current study related to the bearing capacity of the unskirted square footings (partly rough and completely rough) and were compared with the classical bearing capacity formula reported by Terzaghi (1943). To calculate the bearing capacity of the footing, the friction angle obtained from the triaxial test for different relative densities was used. This comparison is shown in Table 2.

Table 2. Comparison of bearing capacity values of the square footing with literature

R_d (%)	Present work		Terzaghi (1943)
	Partly rough	Rough	
30	65,50	73,30	40,68
40	120,39	153,24	45,09
50	158,53	207,81	53,73
60	228,31	268,93	91,09

Source: Authors

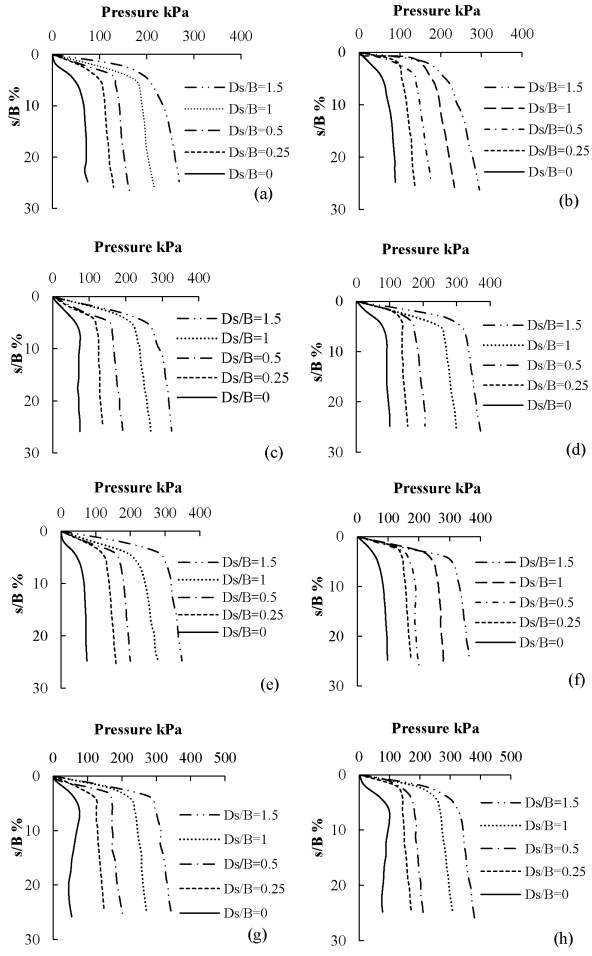


Figure 4. Pressure settlement behavior for footings with partly rough (a, c, e, g) and rough (b, d, f, h) interfaces for square (a, b), plus (c, d), double box (e, f), T (g, h) corresponding to a relative density of 30%.
Source: Authors

From this table, it can be seen that the observed bearing capacity in the present case was higher in comparison to the one obtained by using Terzaghi's formula in all circumstances. This is perhaps due to a higher mobilized friction in the test on account of the dilatancy of the sand, which is more significant at a low-stress level; and due to a slight localized densification of the sand nearby the footing given the applied load. The study of Figure 4 shows that, irrespective of the plan shape of the footing, the bearing capacity increases with the increase in skirt depth. The obtained bearing capacity for unskirted footings is shown in Table 3.

A careful study of this table indicates that, for a given relative density, irrespective of the interface condition of the footing, T-shaped footings is provide the highest bearing capacity, followed by Plus, double box, and square shapes. The difference in the bearing capacity of the footing is quite substantial for the lower relative density, i.e., at R_d = 30%. It should be noted that, without any skirts, double box and square-shaped footings are both the same, hence the identical value of the bearing capacity.

Table 3. Variation of bearing capacity with relative density for the different plan shape of footing

R_d (%)	Interface condition	Ultimate bearing capacity (kPa)			
		Square	Plus	Double box	T
30	Partly rough	65,5	74,29	65,5	77,14
	Rough	73,3	91,43	73,3	100
40	Partly rough	120,39	128,57	120,39	131,43
	Rough	153,24	160	153,24	165,71
50	Partly rough	158,53	165,71	158,53	168,57
	Rough	207,81	210,57	207,81	211,43
60	Partly rough	228,31	231,43	228,31	234,29
	Rough	268,93	270	268,93	271,42

Source: Authors

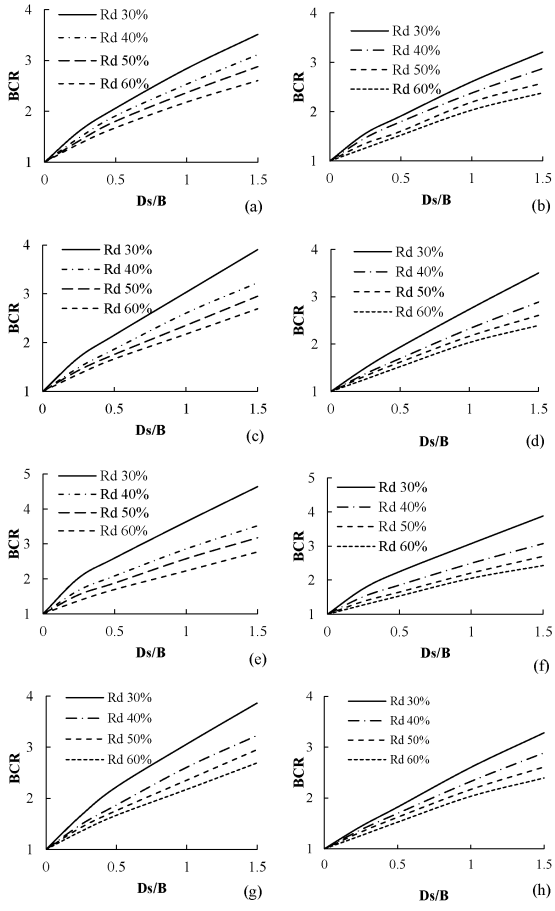


Figure 5. BCR variation for partly (a, c, e, g) and completely (b, d, f, h) rough footing with skirt depth for plan square (a, b), plus (c, d), double box (e, f), and T (g, h) for relative density of 30%.
Source: Authors

The variation of the bearing capacity of the skirted footing (expressed in the form of BCR) with the skirt depth for the different relative density considering various plan shapes is shown in Figure 5. In all cases, the BCR for the square and the multi-edge skirted footings increases with the increase in skirt depth. The BCR at a relative density of 30% for a partly rough footing with D_s/B = 0,25 was observed to be 1,62, 1,67, 1,91, and 2,03 for square, T, plus shape, and double box shapes, respectively. When the D_s/B increased to 1,5, the BCR increased to 3,51, 3,86, 4,43, and 4,64, respectively. For a given skirt depth, the BCR decreased with the increase in relative density.

Variation of SRF with skirt depth

The provision of the skirt along the periphery of the footing increased the bearing capacity and decreased the settlement below the footing. The settlement reduction in skirted footings was defined in a quantitative manner by using a settlement reduction factor:

$$\mathrm{SRF}_{\mathrm{Square/Circle}} = \frac{s - s_{sk}}{s} \quad (1)$$

where, s and s_{sk} are the settlement of the unskirted and skirted footings, respectively, corresponding to a given pressure σ.

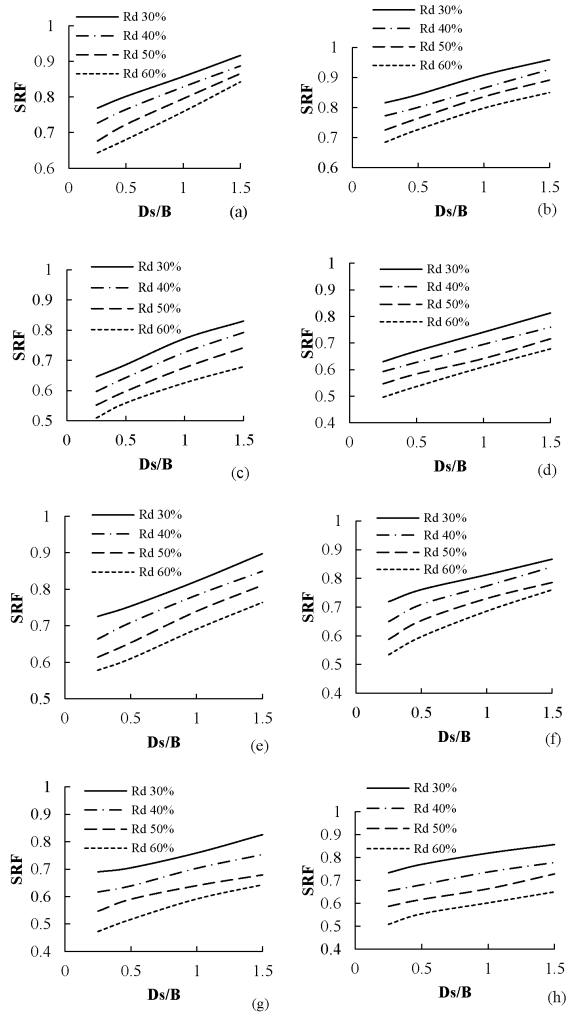


Figure 6. SRF Variation for partly (a, c, e, g) and completely (b, d, f, h) rough footing with skirt depth with different relative densities of sand for σ = 100 kPa with a plan square (a, b), plus (c, d), double box (e, f), and T (g, h) shapes respectively.
Source: Authors

In the present study, the SRF was calculated at pressures of 25 kPa, 50 kPa, 100 kPa, and 200 kPa. Its variation with the skirt depth, for the different relative densities, at a pressure of 100 kPa, and for the various plan shape of the footing, is presented in Figure 6. As anticipated, for a given relative density, the SRF increased with the increase in skirt depth. In contrast, for the constant skirt depth value, the SRF decreased with the increase in relative density for all the plan-shaped footings. The considerable reduction in the settlement was obtained for the footing with a rough interface and at D_s/B = 1,5 and R_d = 30% as the SRF was observed to be 0,87, 0,83, 0,89, and 0,86 for square, Plus, double box, and T plan-shaped footings, respectively. This implies that, at a relative density of 30%, the reduction in the settlement due to the provision of the skirts could range between 80% and 90%, which is quite significant. Thus, skirted footings can be considered as an alternative ground improvement technique in loose sand wherein large settlements are often anticipated.

Artificial Neural Network

This work aims to model the neural network architecture to predict the bearing capacity ratio and settlement reduction factor for multi-edge footings on sand. However, considering the data availability on the bearing capacity and the settlement of the square and circular skirted footings in literature, an ANN model was first developed, which was applicable for these shapes. This model was later modified to suit the aim of this research. Generally, the bearing capacity of non-cohesive soil is dependent on the friction angle (Meyerhof, 1963; Vesic, 1973). In this study, along with the friction angle, skirt depth to width of the footing ratio was also used, since it is considered to affect the bearing capacity. Skirts play a significant role in improving the bearing capacity, as reported by some studies (Khatri, Debbarma, Dutta, and Mohanty, 2017; Khatri and Kumar, 2019). Regarding settlement prediction, normal stress was also plays a role, along with the friction angle and skirt depth (Al-Aghbari and Dutta, 2008; Al-Aghbari and Mohamedzein, 2018; Gnananandarao, Dutta, and Khatri, 2020). Hence, to predict BCR, the non-dimensional skirt depth (D_s/B) and friction angle (ϕ) were considered as input variables. In contrast, the prediction of SRF requires an additional input of pressure, along with the variables above.

The difficult job in the ANN model development is to determine the number of hidden layers and their neurons. The accuracy of the network model is dependent on the initially assigned weights and other associated parameters, as well as its architecture. However, until now, there is no defined technique to achieve the optimal architecture and parameter settings for an ANN model. Therefore, researchers have to follow the time-consuming trial and error approach. However, ANN have a disadvantage, such as being stuck in local minima and slow learning rates (Marto et al. 2014). Still, this technique is popular in Geotechnical applications.

Network Structure Preparation and Data Set Used

The performance of an artificial neural network entirely depends on its structure. The first step in its creation is to fix the number of hidden layers and hidden layer nodes. Generally, a rule of thumb has been adopted for choosing the hidden layer (one in this case) and the neurons in the hidden layer, which is 2/3 of the size of the input variables (Shahin, Maier, and Jaksa, 2002; Dutta et al. 2015b; Rezaei et al. 2016; Dutta, Rani, and Gnananandarao, 2018; Gnananandarao, Dutta, and Khatri, 2018). Considering this, the number of hidden layers and the hidden layer nodes were 1 and 2, respectively. The chosen ANN architecture for the prediction of the BCR and the SRF is shown in Figure 7.

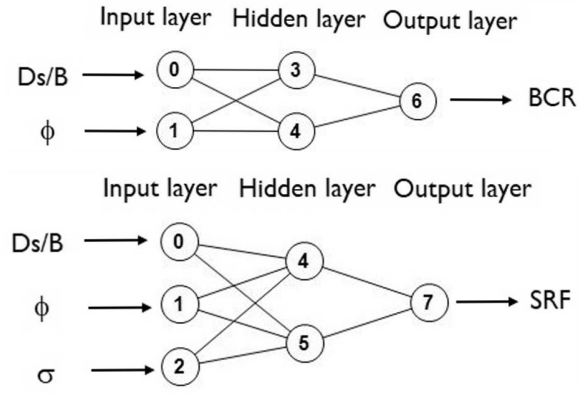


Figure 7. Artificial neural network diagram for (a) BCR and (b) SRF.
Source: Authors

The next difficult task is to fix the number of epochs. An excess number leads to overfitting of the data, whereas a low number leads to poor prediction. The statistical parameter (mean square error) was calculated between the actual and the predicted value corresponding to different epochs. The lowest mean squared error corresponding to an epoch is chosen to develop the neural network model. By following the procedure above, 100 and 450 were selected as the number of epochs for the BCR and the SRF, respectively. Finally, in this study, 2-2-1 and 3-2-1 structures for developing the ANN model were chosen for the BCR and the SRF, respectively. A total of 43 and 131 model experimental data points were collected from the literature (Prasanth and Kumar, 2017; Momeni et al. 2015a; Al-Aghbari, 2002; Al-Aghbari, 2007; Eid, Alansari, Odeh, Nasr, and Sadek, 2009) to predict these parameters in regular shaped skirted footings (square and circular).

Furthermore, an additional 94 and 266 data points were collected from the author's experimental work reported elsewhere (Gnananandarao, Khatri, and Dutta, 2018). Hence the total data points used for modeling were 137 and 397 for the BCR and the SRF, respectively. The data for the other regular shaped footings (rectangular and strip) were excluded due to scarcity in the literature. The range of the various input parameters for which the developed model is applicable is also presented in Table 4.

Table 4. Details of parameters used in ANN and regression models for BCR and SRF

Parameters	Minimum value	Maximum value
The angle of internal friction (ϕ) (Deg.)	32,50	42,9
Skirt depth to width of the footing ratios (D_s/B)	0	2
Bearing capacity of the skirted footing to bearing capacity of unskirted footing ratio (BCR)	1	7,8
Normal stress (σ) (kPa)	25	200
Settlement of skirted footing to settlement of unskirted footing ratio (SRF)	0,06	1

Source: Authors

To check the generalization capability of the model, about 70% and 30% of the data from the total data were respectively selected randomly for training and testing purposes. After its development, the model for these regular-shaped skirted footings was modified to account for the shape of multi-edge skirted footings with the introduction of multiplication factors. These factors were obtained by dividing the BCR/SRF of multi-edge footings with those of the predicted BCR/SRF in regular-shaped footings from the model. The multiplication factor for the BCR and the SRF was termed as F_{mbcr} and F_{msrf}, respectively. A total of 120 and 450 data on the BCR and the SRF of T, Plus, and double-box-shaped skirted footings were taken from the present experimentation. These data points were used to calculate the average multiplication factors to predict the BCR and the SRF of multi-edge skirted footings.

Activation Function Selection, Performance Measure, and Sensitivity Analysis

Artificial neural networks make use of various activation functions to create a relationship between the input and the output variables at each neuron layer. These functions are generally mathematical expressions and are used to produce the outputs. The different activation functions used were linear, threshold, threshold symmetric, sigmoid, sigmoid stepwise, sigmoid symmetric, sigmoid symmetric stepwise, gaussian, gaussian symmetric, gaussian stepwise, elliot, elliot symmetric, linear piece, linear piece symmetric, sin symmetric, cos symmetric, sin, and cos. All of these are available in the open-source Agiel neural network software. To choose the best activation function among these 18 functions, a comparison was made with various performance measures such as the coefficient of determination (R^2), variance account for (VAF), mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE).

The predictive models with an R^2 of 1, VAF of 100%, and MSE, RMSE, and MAE of zero indicate a perfect model (Hajihassani, Abdullah, Asteris, and Armaghani, 2019; Huang, Asteris, Koopialipoor, Armaghani, and Tahir, 2019; Xu, Zhou, Asteris, Armaghani, and Tahir, 2019). The mathematical formulas

Table 5. Mathematical expressions for the performance measures for the BCR

Statistical coefficient	Mathematical expression
Coefficient of determination (R^2)	$R^2 = 1 - \frac{\sum_i (BCR_{hp} - BCR_{ht})^2}{\sum_i (BCR_{hp} - \overline{BCR}_{hp})^2}$
Variance accounts for (VAF)	$\text{VAF} = \left[1 - \frac{var\left(BCR_{ht} - BCR_{hp}\right)}{var\left(BCR_{ht}\right)}\right] \times 100$
Mean square error (MSE)	$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(BCR_{ht} - BCR_{hp})^2$
Root mean square error (RMSE)	$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(BCR_{ht} - BCR_{hp})^2}$
Mean absolute error (MAE)	$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left\|BCR_{ht} - BCR_{hp}\right\|$
Mean absolute percentage error (MAPE)	$\text{MAPE} = \left[\frac{1}{n}\sum_{i=1}^{n}\left\|\frac{BCR_{ht} - BCR_{hp}}{BCR_{ht}}\right\|\right] \times 100$

Note: BCR_{ht}, BCR_{hp} target and predicted BCR; $\overline{BCR}_{ht}$, $\overline{BCR}_{hp}$: mean of the target and predicted BCR, respectively; $S_{BCR_{ht}}$, $S_{BCR_{hp}}$: standard deviation of the target and predicted BCR, respectively; n: number of observations

Source: Authors

for these performance measures (reported by Dutta et al. 2015b; Dutta et al. 2018; Gnanandarao et al. 2019) are provided in Table 5 for the BCR. In the same table, the performance measures for the SRF can be written just by replacing the BCR with the SRF. Based on the best statistical results, sigmoid symmetric and sigmoid were selected as the activation functions for the BCR and the SRF, respectively. The performance measures for these functions in the training and testing phase are shown in Table 6.

Table 6. Performance measures for the training and testing of data for BCR and SRF for the best activation function

Parameter	Training		Testing	
	BCR	SRF	BCR	SRF
Activation function	Sigmoid symmetric	Sigmoid	Sigmoid Symmetric	Sigmoid
R^2	0,96	0,91	0,94	0,93
VAF (%)	91,21	87,62	92,84	89,28
MSE	0,30	0,01	0,36	0,01
RMSE	0,55	0,09	0,60	0,09
MAE	0,36	0,07	0,42	0,06
MAPE	15,97	19,25	17,61	18,90

Source: Authors

The next important step is to choose a learning rate that analyses the performance of the activation function both for the training and the testing datasets after fixing the optimal epochs. The learning rate is vital for understanding any neural network structure, as it explains the influence of the errors on the hidden weights and biases. If the learning rate is low, the values will take a longer time to converge; otherwise, the model may be overfit to the target value. Thus, a default value of 0,7 was considered in this study. After

getting the predicted BCR and the SRF of the regular shaped footings, using the procedure mentioned above, the next step was to check the accuracy of the predicted parameters regarding the target values. It was achieved after minimizing the errors. The comparison of predicted and the target BCR and the SRF values during training and testing step is shown in Figure 8a and b, respectively. This figure indicates that the coefficient of determination ranges between 0,91 and 0,96, which implies a good fit.

Furthermore, the sensitivity analysis was performed to know the direct or the indirect relation between the input and the output parametersaccording to the method used by Erzin and Gul (2014). The connection weights and biases obtained in the neural network were examined to understand the relative importance of the input and output parameters.

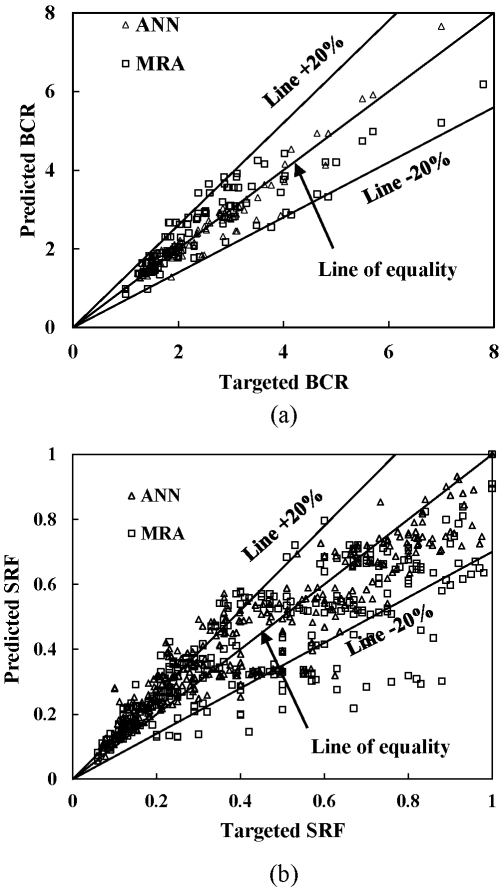


Figure 8. Plot of predicted versus targeted (a) BCR (b) SRF values using ANN for training and testing.
Source: Authors

The result of the sensitivity analysis is presented in Figure 9, which suggests that for the prediction of the BCR as well as SRF, the input variable D_s/B affects the most that is about 64% and 57%, respectively.

Comparison with multivariable regression analysis

In the present study, apart from the development of ANN models, additional multivariable regression analyses were

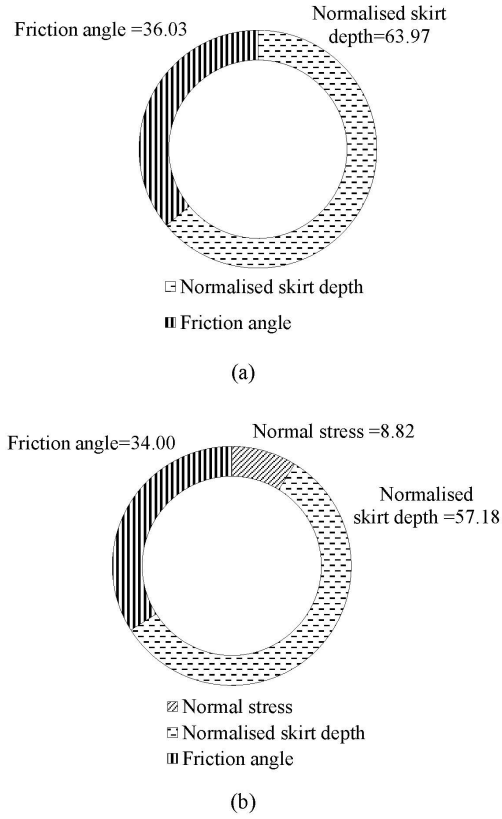


Figure 9. Sensitivity analysis for (a) BCR and (b) SRF.
Source: Authors

carried out to develop expressions for the predictions of the BCR and the SRF. The forecasts from these expressions were compared with ANN predictions. On similar lines with the ANN models, the MRA was first performed on the data related to regular shaped footings, and later, the developed expressions were modified for prediction of the BCR and the SRF of multi-edge footings with the introduction of multiplication factors such as F_{mbcr} and F_{msrf}. To decide the form of the expression, the BCR of the skirted footing was compared with the ratio of the bearing capacity of the embedded footing to the surface footing:

$$\text{BCR}_{\text{Square/Circle}} = \frac{\gamma D_s N_q s_q d_q + 0{,}5 B \gamma N_\gamma s_\gamma d_\gamma}{0{,}5 B \gamma N_\gamma s_\gamma} \qquad (2)$$

therefore,

$$\text{BCR}_{\text{Square/Circle}} = \frac{2 D_s N_q s_q d_q}{B N_\gamma s_\gamma} + d_\gamma \qquad (3)$$

where, $s_q = s_\gamma$ = shape factors, and $d_q = d_\gamma$ = depth factors, according to Meyerhof's bearing capacity theory (1951, 1963).

A careful study of Equation (3) suggests that the form of the expression for the regression analysis can be chosen as

$$BCR_{\text{Square/Circle}} = \left(P \left(\frac{D_s}{B} \right) f(\phi) + Q \right) d_\gamma \qquad (4)$$

In the expression above, P and Q are constants that will be obtained through regression analysis, since the N_q/N_γ ratio is dependent on the friction angle, which is replaced with the function of ϕ i.e. $f(\phi)$. The depth factor d_γ is determined as follows:

$$d_\gamma = 1 + 0{,}1\frac{D_s}{B}\tan\left(45 + \frac{\phi}{2}\right) \quad (5)$$

In the present study, the non-linear regression analysis was carried out by using the Datafit software, version 9,1 (trial version). The final equation after regression analysis for the prediction of the BCR is:

$$BCR_{\text{Square/Circle}} = \left(67{,}2\left(\frac{D_s}{B}\right)\frac{1}{\phi} + 0{,}85\right)d_\gamma \quad (6)$$

Finally, the BCR of the multi-edge skirted footing can be related to the BCR given by Equation (6) as:

$$BCR_{\text{multi-edge}} = BCR_{\text{square/circular}} \times F_{\text{mbcr}} \quad (7)$$

where F_{mbcr} can be calculated by a procedure similar to the one described in the previous sections. The SRF equation for the regular shaped footing was obtained after the regression analysis as:

$$\text{SRF}_{\text{Square/Circle}} = e^{(6{,}07\sigma - 0{,}93\frac{D_s}{B} - 1{,}18\phi + 29{,}78)} \quad (8)$$

Furthermore, the SRF of multi-edge footings is related regular shaped footings as:

$$\text{SRF}_{\text{multi-edge}} = \text{SRF}_{\text{Square/Circle}} \times F_{\text{msrf}} \quad (9)$$

The R^2 for Equations (6) and (8) is about 0,87 and 0,84, respectively, which is acceptable. A comparison between the BCR and the SRF predictions for the square and the multi-edge footings with partly rough and completely rough interfaces obtained from the regression analysis and the ANN is shown in Figures 10 and 11, respectively. These figures indicate that the predicted BCR and SRF are almost within the ±20% of the line of equality. The R^2 ranges between 0,940 and 0,977 for the ANN model and 0,827 and 0,934 for the regression analysis. Furthermore, the study of Figure 11 suggests that the R^2 for the SRF prediction may lie between 0,913 and 0,985 for the ANN model and 0,739 and 0,932 for the regression analysis. A comparison of F_{mbcr} and F_{msrf} obtained from the ANN and the regression analysis for the parameters in question is provided in Table 7. It is quite noteworthy that these factors are comparable, even though they were obtained with different methodologies.

Table 7. Multiplication factors F_{mbcr} and F_{msrf} for prediction of BCR and SRF for multi-edge footings in ANN and regression models

Plan shape of footing	Multiplication factor (F_{mbcr} and F_{msrf})			
	Artificial neural networks		Regression Analysis	
	BCR	SRF	BCR	SRF
Plus	1,03	0,98	1,02	0,99
Double box	1,11	0,89	1,09	0,92
T	1,04	0,97	1,03	0,98

Source: Authors

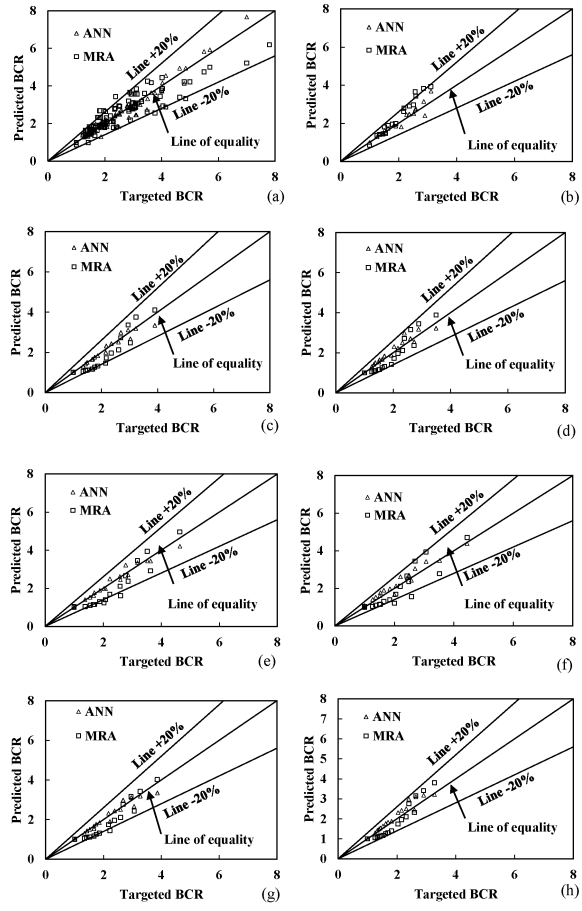


Figure 10. Comparison of ANN with MRA models after the prediction of BCR for partly rough (a, c, e, g) and rough (b, d, f, h) interfaces of square (a, b), plus (c, d), double box (e, f), T (g, h) plan shapes.
Source: Authors

Equation for the BCR and the SRF from ANN

The goal of the present study was to develop an equation for the reliable prediction of the BCR and the SRF of the regular shaped and multi-edge skirted footings. From the comparison of the predictions obtained with the ANN and the regression analyses, it is quite clear that ANN predictions were always superior. Therefore, the equations for predicting the BCR and the SRF are also presented here.

The ANN model was developed with the optimum number of epochs (100) using the open-source Agiel software to obtain the weights and the biases introduced between the hidden layer and the output layer. The generalized function of the ANN model for the output (BCR) of the regular shaped skirted footing is defined by Equation 10:

$$\text{BCR}_{\text{Square/Circle}} = f\left\{b_0 + \sum_{(k=1)}^{h}\left[w_k f\left(b_{hk} + \sum_{(j=1)}^{m} w_{jk}X_j\right)\right]\right\} \quad (10)$$

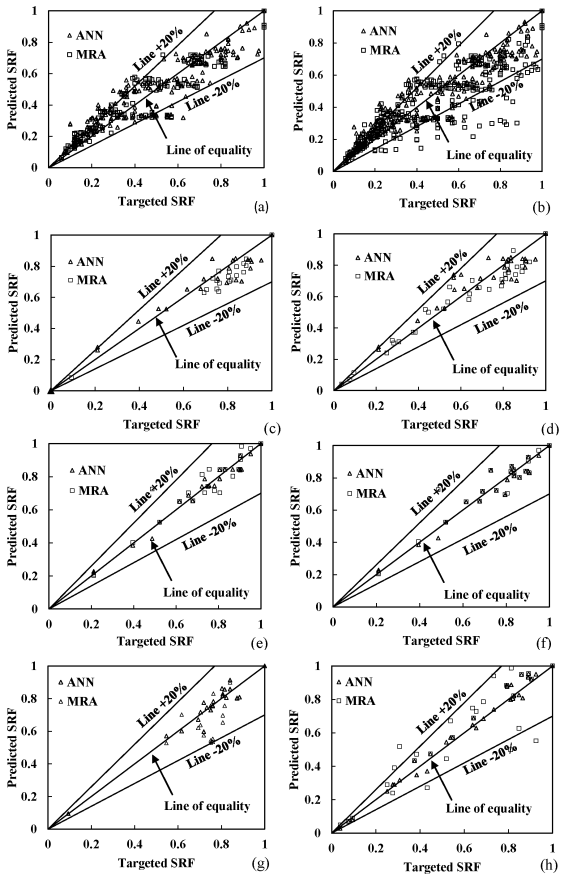


Figure 11. Comparison of ANN model with MRA after the prediction of SRF for partly rough (a, c, e, g) and rough (b, d, f, h) interfaces of square (a, b), plus (c, d), double box (e, f), T (g, h) footings.
Source: Authors

$$A = -1{,}18 + 1{,}43\frac{D_s}{B} - 1{,}05\phi \tag{11}$$

$$B = 0{,}50 - 1{,}29\frac{D_s}{B} + 0{,}10\phi \tag{12}$$

$$E = 1{,}07 + \frac{2{,}66}{(1+e^{-A})-1} - \frac{2{,}18}{(1+e^{-B})-1} \tag{13}$$

$$\mathrm{BCR}_{\mathrm{Square/Circle}} = \frac{2}{(1+e^{-E})-1} \tag{14}$$

The $\mathrm{BCR}_{\mathrm{Square/circle}}$ resulting from Equation 14 is between −1 and 1 for the activation function (sigmoid symmetric). Hence, there is a need for the denormalization of the output to get the actual value. The denormalized equation is as follows:

$$\mathrm{BCR}_{\mathrm{Square/Circle}} = 0{,}5\left(\mathrm{BCR}_{\mathrm{Square/Circle}} + 1\right)\left(\mathrm{BCR}_{\mathrm{Square/Circle(max)}} - \mathrm{BCR}_{\mathrm{Square/Circle(min)}}\right) + \mathrm{BCR}_{\mathrm{square/Circle(min)}} \tag{15}$$

where $\mathrm{BCR}_{\mathrm{Square/Circle(max)}}$ and $\mathrm{BCR}_{\mathrm{Square/Circle(min)}}$ are the maximum and the minimum predicted bearing capacity ratios of regular shaped skirted footings, respectively.

The BCR of multi-edge footings can be obtained by following Equation (7), which was defined earlier. On similar lines with the BCR, the SRF can be predicted by following Equations 16-19 as given below:

$$A = -0{,}26 + 0{,}01s + 0{,}07\frac{D_s}{B} - 0{,}05\phi \tag{16}$$

$$B = -0{,}13 - 0{,}01s + 0{,}06\frac{D_s}{B} - 0{,}03\phi \tag{17}$$

$$E = 0{,}40 + 0{,}16\frac{1}{1+e^{-A}} + 0{,}17\frac{1}{1+e^{-B}} \tag{18}$$

$$\mathrm{SRF}_{\mathrm{Square/Circle}} = \frac{1}{1+e^{-E}} \tag{19}$$

The $\mathrm{SRF}_{\mathrm{Square/Circle}}$ predicted by Equation (19) ranged between −1 and 1 for the activation function (sigmoid symmetric). Hence, there is a need for the denormalization of output (SRF) to get the actual value. The denormalized equation (20) is shown below:

$$\mathrm{SRF}_{\mathrm{Square/Circle}} = 0{,}5\left(\mathrm{SRF}_{\mathrm{Square/Circle}} + 1\right)\left(\mathrm{SRF}_{\mathrm{Square/Circle(max)}} - \mathrm{SRF}_{\mathrm{Square/Circle(min)}}\right) + \mathrm{SRF}_{\mathrm{Square/Circle(min)}} \tag{20}$$

Where $\mathrm{SRF}_{\mathrm{Square/Circle(max)}}$ and $SRF_{\mathrm{Square/Circle(min)}}$ are the maximum and the minimum predicted settlement reduction factors of regular shaped skirted footings. Furthermore, the SRF of multi-edge footings is obtained following Equation (9) and Table 7. The weights and biases between the hidden layer neurons with input and the output nodes for the $\mathrm{BCR}_{\mathrm{Square/Circle}}$ and the $\mathrm{SRF}_{\mathrm{Square/Circle}}$ prediction are provided in Tables 8 and 9, respectively.

Table 8. Weights and biases between hidden layer neurons with input and output nodes for BCR prediction

Neurons	Weights (W_{jk}) Biases				
	D_s/B	ϕ	BCR	b_{hk}	b_0
Hidden neuron 1 (k = 1)	1,43	-1,05	1,33	-1,18	1,07
Hidden neuron 2 (k = 2)	-1,29	0,10	-1,09	0,50	–

Source: Authors

Table 9. Weights and biases between hidden layer neurons with input and output nodes for SRF prediction

Neurons	Weights (W_{jk}) Biases					
	σ	D_s/B	ϕ	SRF	b_{hk}	b_0
Hidden neuron 1 (k = 1)	0,01	0,07	-0,05	0,16	-0,26	0,40
Hidden neuron 2 (k = 2)	-0,01	0,06	-0,03	0,17	-0,13	–

Source: Authors

Conclusions

The study carried out in this work showed the feasibility of using a simple ANN and multivariable regression analysis to predict the bearing capacity ratio and settlement reduction

factor of multi-edge skirted footings resting on sand. Based on the obtained results, the following conclusions can be drawn:

1. For a given relative density, the non-dimensional bearing capacity of square/circular and multi-edge footings was defined in terms of BCR increases with an increase in depth of the skirt. The BCR of multi-edge footings was marginally higher than that of its square/circular counterpart, which is desirable.

2. For a given constant pressure, the reduction in settlement of skirted footings, expressed in the form of SRF, increased with the increase in skirt depth. However, contrary to the observation of BCR, the SRF of multi-edge footings was marginally smaller in comparison to regular-shaped skirted footings.

3. The BCR and SRF predicted by using ANN or regression analyses are in agreement with experimental values, although ANN provides superior predictions in all cases.

4. The developed equations can be used to predict the BCR and SRF of partly rough and completely rough skirted footings as the predictions are not much affected by interface roughness.

It is anticipated that the outcome of this study will help in emphasizing the use of multi-edge skirted footings wherever possible. Furthermore, the developed equations can be used to predict the bearing capacity and reduction in settlement of such footings at a given pressure without conducting the field tests. Neural network models, in general, cannot provide reasons and reasoning beyond the model (black boxes) thus obtained; once a network knows one set of weights, any new learning results in disastrous forgetting. The suitability of alternative techniques such as vector supporting machines, particle swarm optimization, or genetic programming may also be explored in the future.

References

Al-Aghbari, M. Y., and Khan, A. J. (2002). Behaviour of shallow strip foundations with structural skirts resting on dense sand. *Proceedings of challenges of concrete constructions, 6*, 737-746. 10.1680/cfec.31784.0072

Al-Aghbari, M. Y. (2007). Settlement of shallow circular foundations with structural skirts resting on sand. *The Journal of Engineering Research, 4*(1), 11-16. 10.24200/tjer.vol4iss1pp11-16

Al-Aghbari, M. Y. and Mohamedzein, Y. E.-A. (2020). The use of skirts to improve the performance of a footing in sand. *International Journal of Geotechnical Engineering, 14*(2), 134-141. 10.1080/19386362.2018.1429702

Armaghani, D. J., Faradonbeh R. S., Rezaei, H., Rashid A. S. A., and Amnieh, H. B. (2018). Settlement prediction of the rock-socketed piles through a new technique based on gene expression programming. *Neural Computing and Applications, 29*, 1115-1125. 10.1007/s00521-016-2618-8

Armaghani, D. J., Shoib, R. S. N. S. B. R., Faizi, K., and Rashid, A. S. A. (2017). Developing a hybrid PSO–ANN model for estimating the ultimate bearing capacity of rock-socketed piles. *Neural Computing and Applications, 28*, 391-405. 10.1007/s00521-015-2072-z

Chakraborty, D. and Kumar, J. (2013). Dependency of N_γ on footing diameter for circular footings. *Soils and Foundations, 53*(1), 173-180. 10.1016/j.sandf.2012.12.013

Chen, W., Sarir, P., Bui, X-N., Nguyen, H., Tahir, M. M., and Armaghani D. J. (2019). Neuro –genetic, neuro – imperialism and genetic programing modelsin predicting ultimate bearing capacity of pile. *Engineering with Computers, 36*, 1101-1115. 10.1007/s00366-019-00752-x

Dawarci, B., Ornek, M., and Turedi, Y. (2014). Analysis of multi-edge footings rested on loose and dense sand. *Periodica Polytechnica Civil Engineering, 58*(4), 355-370. 10.3311/PPci.2101

Dutta, R. K., Dutta K., Jeevanandham, S. (2015). Prediction of deviator stress of sand reinforced with waste plastic strips using neural network. International *Journal of Geosynthetics and Ground Engineering. 1*(2), 1-12. 10.1007/s40891-015-0013-7

Dutta, R. K., Rani, R., and Gnananandarao, T. (2018). Prediction of ultimate bearing capacity of skirted footing resting on sand using artificial neural networks. *Journal of Soft Computing in Civil Engineering, 2*(4), 34-46. 10.22115/SCCE.2018.133742.1066

Eid, H. T., Alansari, O. A., Odeh, A. M., Nasr, M. N., and Sadek, H. A. (2009). Comparative study on the behavior of square foundations resting on confined sand. *Canadian Geotechnical Journal, 46*, 438-453. 10.1139/T08-134

Erzin, Y. and Gul, T. (2014). The use of neural networks for the prediction of the settlement of one-way footings on cohesionless soils based on standard penetration test. *Neural Computing and Applications, 24*, 891-900. 10.1007/s00521-012-1302-x

Ghazavi, M. and Mokhtari, S. (2008). Numerical investigation of load-settlement characteristics of multi-edge shallow foundations. In ed. Jadhav, M. N. (Ed.) *Proceedings of The 12th International Conference of International Association for Computer Methods and Advances in Geomechanics* (*IACMAG*), Red Hook, NY: Curran. pp. 3344-3351.

Gnananandarao T., Khatri V. N., and Dutta R. K. (2018). Performance of multi-edge skirted footings resting on sand. *Indian Geotechnical Journal, 48*(3), 510-519. 10.1007/s40098-017-0270-6

Gnananandarao, T., Dutta, R. K. and Khatri, V. N. (2019). Application of artificial neural network to predict the settlement of shallow foundations on cohesionless soils. *Geotechnical Applications, 13*, 51-58. 10.1007/978-981-13-0368-5_6

Gnananandarao, T., Dutta, R. K. and Khatri, V. N. (2020). Model studies of plus and double box shaped skirted footings resting on sand. *International Journal of Geo-engineering, 11*(2), 1-17. 10.1186/s40703-020-00109-0

Hajihassani, M., Abdullah, S. S., Asteris, P. G., and Armaghani, D. J. (2019). A Gene Expression Programming Model for Predicting Tunnel Convergence. *Applied Sciences, 9*, 4650. 10.3390/app9214650

Harandizadeh, H., Armaghani, D. J., and Khari, M. (2019). A new development of ANFIS-GMDH optimized by PSO to predict pile bearing capacity based on experimental datasets. *Engineering with Computers*, 1-16. 10.1007/s00366-019-00849-3

Huang, L., Asteris, P. G., Koopialipoor, M., Armaghani, D. J., and Tahir, M. M. (2019). Invasive Weed Optimization Technique-Based ANN to the Prediction of Rock Tensile Strength. *Applied Sciences, 9*, 5372. 10.3390/app9245372

IS 1498 (1970). *Classification and identification of soils for general engineering purposes*. Delhi, India: Bureau of Indian Standards.

IS 6403 (1981). *Determination of bearing capacity of shallow foundation*. Delhi, India: Bureau of Indian Standards.

Kalinli, A., Acar, M. C., and Gunduz, Z. (2011).New approaches to determine the ultimate bearing capacity of shallow foun-dations based on artificial neural networks and ant colony optimization. *Engineering Geology, 117*, 29-38. 10.1016/j.enggeo.2010.10.002

Khari, M., Armaghani, D. J., and Dehghanbanadaki, A. (2020). Prediction of Lateral Defection of Small–Scale Piles Using Hy-brid PSO-ANN Model. *Arabian Journal for Science and Engineering, 45*, 3499-3509. 10.1007/s13369-019-04134-9

Khatri, V. N. and Kumar, J. (2019). Finite-Element Limit Analysis of Strip and Circular Skirted Footings on Sand. *International Journal of Geomechanics, 19*(3), 06019001. 10.1061/(ASCE)GM.1943-5622.0001370

Khatri, V. N., Debbarma, S. P., Dutta, R. K., and Mohanty, B. (2017). Pressure-settlement behavior of square and rectangular skirted footings resting on sand. *Geomechanical Engineering 12*(4),689-705. 10.12989/gae.2017.12.4.68

Khudier, A. S. (2018). Prediction of bearing capacity for soils in basrah city using artificial neural network (ANN) and multi-linear regression (MLR) models. *International Journal of Civil Engineering and Technology, 9*(4), 853-864.

Kumar, J. (2009). The variation of N_γ with footing roughness using the method of characteristics. *International Journal for Nu-merical and Analytical Methods in Geomechanics, 33*(2), 275-284. 10.1002/nag.716

Marto, A., Hajihassani, M., and Momeni, E. (2014). Bearing Ca-pacity of Shallow Foundation's Predic-tion through Hybrid Artificial Neural Networks. *Applied Mechanics and Materials, 567*, 681-686. 10.4028/www.scientific.net/AMM.567.681

Meyerhof, G. G. (1951). The ultimate bearing capac-ity of foun-dations. *Geotechnique, 2*(4) 301-332. 10.1680/geot.1951.2.4.301

Meyerhof, G. G. (1963). Shallow foundations. Journal of Soil Mechanics and Foundation Division. *ASCE, 91*(SM2), 21-31.

Meyerhof, G. G. (1963). Some recent research on bearing capacity of foundations. *Canadian Geotechnical Journal, 1*, 16-26. 10.1139/t63-003

Momeni, E., Nazir, R., Armaghani, D. J., and Maizir, H. (2015a). Application of Artifcial Neural Network for Predicting Shaft and Tip Resistances of Concrete Piles. *Earth Sciences Research Journal, 19*(1), 85-93. 10.15446/esrj.v19n1.38712

Momeni, E., Armaghani, D. J., Nazir, R., and Sohaie, H. (2015b). Bearing capacity of precast thin-walled foundation in sand. *Proceedings of the Institution of Civil Engineers Geotechnical Engineering, 168*(GE6), 539-550. 10.1680/jgeen.14.00177

Momeni, E., Armaghani, D. J., Fatemi, S. A., and Nazir, R. (2017). Prediction of bearing capacity of thin-walled foundation: a simulation approach. *Engineering with Computers, 3*(2), 319-327. 10.1007/s00366-017-0542-x

Nazir, R., Momeni, E., Marsono, K., and Sohaie, H. (2013). Pre-cast spread foundation in industrialized building system. In Hossain, M. Z. and Hossain, M. S. (Eds.) *Proceedings of the 3rd International Conference on Ge-otechnique, Construction Materials and Environment - GEO-MATE 2013* (pp. 47-52). Nagoya, Japan: Nagoya Institute of Technology.

Nazir, R., Momeni, E., Marsono, K., and Maizir, H. (2015a). An artificial neural network approach for prediction of bearing capacity of spread foundations in sand. *Journal Teknologi, 72*(3), 9-14. 10.11113/jt.v72.4004

Nazir, R., Momeni, E., and Marsono, K. (2015b). Prediction of bearing capacity for thin-wall spread foundations using ICA-ANN predictive model. In: *Proceedings of the International Conference on Civil, Structural, and Transportation Engineering*, Ottawa, Ontario-May, 4-5.

Prasanth, T. and Kumar, P.R. (2017). A study on load carrying capacity of skirted foundation on sand. *International Journal of Science and Research, 6*(6), 2231-2235.

Rezaei, H., Nazir, R., and Momeni, E. (2016). Bearing capacity of thin-walled shallow foundations: an experimental and artificial intelligence-based study. *Journal of Zhejiang University Sci-ence A: Applied Physics and Engineering, 17*(4), 273-285. 10.1631/jzus.A1500033

Shahin, M. A., Maier, H. R., and Jaksa, M. B. (2002). Predicting settlement of shallow foundations using neural networks. *Journal of Geotechnical and Geoenvironmental Engineering, ASCE, 128*(9), 785-793. 10.1061/(ASCE)1090-0241(2002)128:9(785)

Tang, C., Phoon, K.K., and Toh, K.C. (2014). Effect of footing width on Nγ and failure envelope of eccentrically and obli(6), 694-707. 10.1139/cgj-2013-0378

Terzaghi, K. (1943). *Theoretical soil mechanics*. New York: John Wiley and Sons. 10.1002/9780470172766

Vesic, A.S. (1973) Analysis of Ultimate Loads of Shallow Foundations. *Journal of the Soil Mechanics and Foundations Division, 99*, 45-73. https://cedb.asce.org/CEDBsearch/record.jsp?dockey=0020165

Xu, H., Zhou, J., Asteris, P. G., Armaghani, D. J., and Tahir, M. Md. (2019). Supervised Machine Learning Techniques to the Pre-diction of Tunnel Boring Machine Penetration Rate. *Applied sciences, 9*, 1-19. 10.3390/app9183715

Yong, W., Zhou, J., Armaghani, D. J., Tahir, M. M., Tarinejad, R., Pham, B. T., and Huynh, V. V. (2020). A new hybrid simulated annealing–based genetic programming technique to predict the ultimate bearing capacity of piles. *Engineering with Computers,* 1-17. 10.1007/s00366-019-00932-9

Ziaee, S. A., Sadrossadat, E., Alavi, A. H., and Shadmehri, D. M. (2015). Explicit formulation of bearing capacity of shallow foundations on rock masses using artificial neural networks: application and supplementary studies. *Environmental Earth Science, 73*(7), 3417-3431. 10.1007/s12665-014-3630-x

Research Article / Chemical – Food – Environmental Engineering http://dx.doi.org/10.15446/ing.investig.v40n3.83709

Adsorption Thermodynamics of Cr(VI) Removal by Using Agro-Industrial Waste of Oil Palm Bagasse and Plantain Peels

Termodinámica de Adsorción en la Remoción de Cr(VI) usando Residuos Agroindustriales Bagazo de Palma Aceitera y Cáscaras de Plátano

Angel Villabona-Ortíz[1], Candelaria Tejada-Tovar[2], and Rodrigo Ortega-Toro[3]

ABSTRACT

The effect of temperature on the adsorption of Cr (VI) was determined with waste materials from the agro-industry, such as palm bagasse and plantain peels, by determining the thermodynamic parameters which allow to establish the mechanisms that control the process. The methodology included an initial preparation of the biomass, its characterization, and subsequent adsorption tests by setting the initial concentration of the metal to 100 ppm, a particle size of 0,5 mm, a biomass amount of 0,325 g, pH 2, and a volume of 100 mL. The process temperature varied between 303,15, and 352,15 K. The concentration of the remaining metal in the solution was performed using the diphenyl carbazide colorimetric method through a UV-Vis spectroscopy at a wavelength of 540 nm. The results show that the effect of temperature does not present a defined trend for palm bagasse, whereas it is linear for adsorption using plantain peels. Furthermore, the process of adsorption of Cr (VI) with palm bagasse is more favorable at higher temperatures, since it is a spontaneous process with a physical adsorption mechanism. On the other hand, the adsorption process with plantain peels is thermodynamically feasible at temperatures from 40 to 55 °C and not spontaneous at higher temperatures. Also, a physicochemical adsorption mechanism was evinced. It is concluded that the use of the studied materials is possible in the removal of the Cr (VI) ion in aqueous solutions.

Keywords: adsorption, batch system, bio-adsorbent, heavy metal, temperature

RESUMEN

Se determinó el efecto de la temperatura en la adsorción de Cr (VI) usando materiales de desecho de la agroindustria, tales como bagazo de palma y cáscaras de plátano, determinando los parámetros termodinámicos que permiten establecer los mecanismos que controlan el proceso. La metodología incluyó una preparación inicial de la biomasa, su caracterización y pruebas posteriores de adsorción fijando la concentración inicial del metal a 100 ppm, el tamaño de partícula a 0,5 mm, una cantidad de biomasa de 0,325 g, pH 2 y un volumen de 100 mL. La temperatura del proceso varió entre 303,15, y 352,15 K. La concentración del metal remanente en la solución se realizó usando el método colorimétrico del difenil carbazida por espectroscopia UV-Vis a una longitud de onda de 540 nm. Los resultados muestran que el efecto de la temperatura no presenta una tendencia definida para el bagazo de palma, mientras que es lineal para la adsorción usando cáscaras de plátano. Además, el proceso de adsorción de Cr (VI) con el bagazo de palma se vuelve más favorable a temperaturas más altas, siendo este un proceso espontaneo con un mecanismo de adsorción física. Por otro lado, el proceso de adsorción con cáscaras de plátano es termodinámicamente factible a temperaturas de 40 hasta 55 °C y no espontáneo a temperaturas superiores. También se hace evidente un mecanismo de adsorción fisicoquímico. Se concluye que es posible el uso de los materiales estudiados en la remoción del ion Cr (VI) en soluciones acuosas.

Palabras clave: adsorción, sistema por lotes, bioadsorbente, metal pesado, temperatura

Received: November 25th, 2019
Accepted: July 13th, 2020

[1]Chemical Engineer, Universidad Industrial de Santander, Colombia. M.Sc. in Environmental Engineering, Universidad de Cartagena, Colombia. Affiliation: Full Time-Professor, Universidad de Cartagena, Colombia. Process design and Biomass Utilization Group (IDAB) E-mail: avillabonao@unicartagena.edu.co
[2]Chemical Engineer, Universidad Industrial de Santander, Colombia. M.Sc. in Environmental Engineering, Universidad de Cartagena, Colombia. Affiliation: Full Time-Professor, Universidad de Cartagena, Colombia. Process design and Biomass Utilization Group (IDAB) E-mail: ctejadat@unicartagena.edu.co
[3]Agroindustrial Engineer, Universidad del Cauca, Colombia. M.Sc. in Food Science and Engineering, Universidad Politécnica de Valencia. España. Ph.D. in Food Science and Technology, Universidad Politécnica de Valencia. España. Affiliation: Full Time Professor, Universidad de Cartagena, Colombia. Research Group in Complex Fluids Engineering and Food Rheology (IFCRA), and Food Packaging and Shelf Life Research Group (FP&SL).
E-mail: rortegap1@unicartagena.edu.co

How to cite: Villabona-Ortíz, A., Tejada-Tovar, C., and Ortega-Toro, R. (2020). Adsorption Thermodynamics of Cr(VI) Removal by Using Agro-Industrial Waste of Oil Palm Bagasse and Plantain Peels. *Ingeniería e Investigación*, *40*(3), 22-28. 10.15446/ing.investig.v40n3.83709

Introduction

Chromium and its compounds are toxic when introduced in natural water from a variety of industrial wastes. These materials accumulate in the food chain and affect living organisms in the ecosystem (Jain et al., 2018; Carreño-Sagayo, 2016). The primary sources come from dyeing, canning, electroplating, leather tanning, metal processing, paints and pigments, and the textile and steel manufacturing industries (Akram, Bhatti, Iqbal, Noreen, and Sadaf, 2017; Vargas-Niño, Carriazo, and Castillo, 2011). Contamination in water currents by Cr(VI) represents a threat to animal and human health, as it can cause lung cancer, ulcers, perforations of the nasal septum, and kidney damage (Razi, Al-Gheethi, and Za, 2018). Due to such mutagenic and carcinogenic properties in living tissues, Chromium is one of the most harmful metals. It is included in the priority list of hazardous substances of the World Health Organization (WHO), with permissible limits of 0,05 ppm and 0,1 ppm established in drinking and inland water, respectively (Jain et al., 2018; Haroon et al., 2016).

Traditional methods of contaminated effluent treatment with Cr(VI) consist of chemical precipitation, filtration, oxidation and chemical reduction, electrocoagulation, reverse osmosis, ultrafiltration, ion exchange, adsorption, evaporation, among others (Razi et al., 2018; Rico et al., 2018; Akram et al., 2016). However, these processes have some limitations due to high energy consumption and large amounts of input chemicals (Wassie and Srivastava, 2016), whereas bio-adsorption is a versatile and effective method to remove heavy metals because it involves the use of low cost and high-efficiency adsorbents. In recent years, several low-cost materials have been reported, such as rice husk (Brahmaiah, Spurthi, Chandrika, Ramanaiah, and Prasad, 2015; Lin et al., 2018), moringa seeds (Maina et al., 2016), walnut husk (Casarin et al., 2016), orange peels (Giza, 2017; Abdelhafez and Li, 2016), among others. These lignocellulosic materials are suitable heavy metal adsorbents due to the presence of polymeric groups that serve as active centers for the uptake of metals (Marimón-Bolívar, Tejeda-Benítez, and Herrera, 2018).

To understand the phenomenon of heavy metal bio-adsorption in solutions, isotherms, kinetics, and desorption have been used. All of them provide information on the amount of ions adsorbed by a given biomass, the interaction mechanisms between adsorbents and adsorbate, adsorption mechanisms, among others, which enables the optimization of adsorption mechanism pathways, the expression of surface properties and adsorbent capacities, and the production design of adsorption systems (Anastopoulos and Kyzas, 2016). It was found that the adsorption kinetics of Cr(VI) at different temperature values on oil palm bagasse was adjusted to the pseudo-second order model. The case of plantain peels was described by the Elovich model; for this reason, we can infer that the process is controlled by chemical adsorption (Villabona-Ortíz, Tejada-Tovar, and Ortega-Toro, 2020).

The study of adsorption thermodynamics allows us to identify the adsorption mechanisms involved in the process. This is due to the fact that there are three main steps during the adsorption of a compound to the adsorbent surface. In the first step, molecular mass is transferred from the solution to the adsorbing surface. Then, internal molecular diffusion to the adsorption sites placed in the adsorbent takes place. Adsorption is completed in the final step of the process by attraction between adsorbate and adsorbent, either by physiadsorption or chemosorption (Kecili and Hussain, 2018). Consequently, studies for the modification of thermodynamic parameters are the focus of evaluations of adsorbent efficiency, in order to establish adsorption mechanisms and optimize the process. The most common equations that relate thermodynamic parameters are the change in Gibbs free energy (ΔG^0), enthalpy (ΔH^0), and entropy (ΔS^0) in steady state (Tran, You, Hosseini-Bandegharaei, and Chao, 2017). As was reported previously, Cr(VI) adsorption by using coconut shells is endothermic, spontaneous, and irreversible (Ijeamaka, Christian, Fabian, MaryJane, and Joseph, 2018) and tamarind shells behave similarly (Bangaraiah, 2020). By contrast, adsorption by means of corn residues is reversible, favourable, spontaneous, and endothermic (Núñez-Zarur, Tejada-Tovar, Villabona-Ortíz, Acevedo, and Tejada-Tovar, 2018). Thus, the objective of the present work is to determine the influence of temperature on the adsorption of Cr(VI) present in aqueous solutions, using waste materials from the agro-industry such as palm bagasse and plantain peels. The thermodynamic parameters were calculated to establish the mechanisms that control the process.

Experimental

Biomass Preparation

The biomass, consisting of plantain peels and palm bagasse, was supplied by local businesses from the department of Bolívar, Colombia, where the biomass of interest was a residue of their production processes. It was prepared by applying a series of unit operations to adapt it to the necessary conditions for the bio-adsorption process. Firstly, the biomass was washed with deionized water to remove dirt and any pulp debris it could contain. Then, a size reduction was performed to facilitate its handling in later stages. A final wash with distilled water was carried out to remove tannins, resins or other compounds that could affect the process. The biomass was dried in an oven for 8 h at 60 °C, and then grounded and sieved until a particle size of 0,5 mm was obtained (Chieregato and Tapia, 2016). The bio-adsorbents were stored dry in airtight bags to keep them in good condition.

Solution Preparation

The solution was synthesized from 0,28288 g of potassium dichromate ($K_2Cr_2O_7$) per liter of water deionized for the Cr (VI) solution, thus ensuring a concentration of 100 ppm. The solutions were brought to a pH of 2 with 2 M hydrochloric acid and sodium hydroxide (Tejada, Quiñones, Tejeda, and Marimón, 2015).

Thermodynamic parameters

For the calculation of thermodynamic parameters, Cr(VI) adsorption tests were carried out in a 100 mL Erlenmeyer shaker, simulating a batch reactor with a stirring of 200 rpm for 24 hours at a constant temperature. After 24 hours, the samples were removed from the agitator, and then the final concentration measurements of the metal were made. These data were determined with a UV-VIS spectrophotometer of the brand Shimadzu, model UV 17 000. The standard method to determine the amount of Chromium in water was applied using 1,5-diphenylcarbazide at 540 nm (Tejada-Tovar, Herrera-Barros, and Villabona-Ortíz, 2020).

The effect of temperature on Cr(VI) adsorption was determined by calculating Gibbs free energy ($\Delta G°$), enthalpy ($\Delta H°$), and entropy change ($\Delta S°$), using Equations 1 and 2. The temperature values of 303,15, 328,15 and 352,15 K were taken with a particle size of 0,5 mm and a biomass quantity of 0,325 g.

$$\Delta G = -RT \ln k_c \tag{1}$$

$$\ln K_c = \frac{-\Delta H}{RT} + \frac{\Delta S}{R} \tag{2}$$

The total adsorption capacity qt for each temperature was initially calculated by using Equation 3

$$q_t = \frac{(C_0 - C_t) * V}{m} \tag{3}$$

where C_0 (mg/L) is the initial concentration of the metal, C_t (mg/L) is the metal concentration in the liquid phase at a time t, m is the mass of adsorbent used in (g), and V, the volume of the solution in (L).

Finally, the values of q_t for the different temperatures were obtained to perform the calculations of the adsorption equilibrium constants K_c described in Equation (4). The van't Hoff graphic method was used on the obtained data to determine the change in Gibbs free energy ($\Delta G°$) from the enthalpy change ($\Delta H°$) and entropy change ($\Delta S°$).

$$K_C = \frac{C_{ac}}{C_{se}} \tag{4}$$

where C_{ac} is the concentration of the adsorbate in equilibrium, contained in the surface of the adsorbent, C_{se} is the concentration in solution in the equilibrium, and R is the universal constant of the gases (8,314 J/mol.K).

Results and discussion

Adsorbent characterization

From the Brunauer-Emmett-Teller (BET) analysis, it was found that the plantain peel has a larger surface area: 3,0889 m^2/g compared to 2,7317 m^2/g of palm bagasse. However, lignocellulosic residues are expected to have a low surface area due to their cellulose and hemicellulose content (Asuquo, Martin, Nzerem, Siperstein, and Fan, 2017). As for the pore size, value: 16,410 nm compared to 6,4457 nm of plantain peels, which indicates that both are mesoporous materials, making them suitable for adsorption in their liquid phase, since this facilitates Cr (VI) diffusion into the adsorbent structure (Hubbe, Azizian, and Douven, 2019).

Figures 1 (a) and 2 (b) show the SEM micrographs. For the plantain peel, a cylindrical aspect is observed. In contrast, an entanglement structure with porosity is noted for the palm bagasse, typical of lignocellulosic materials (Martínez-Mendoza et al., 2020). This characteristic is extremely beneficial for metal ion transport which increases the adsorption capacity (Pradhan, Arora, and Mahajani, 2018).

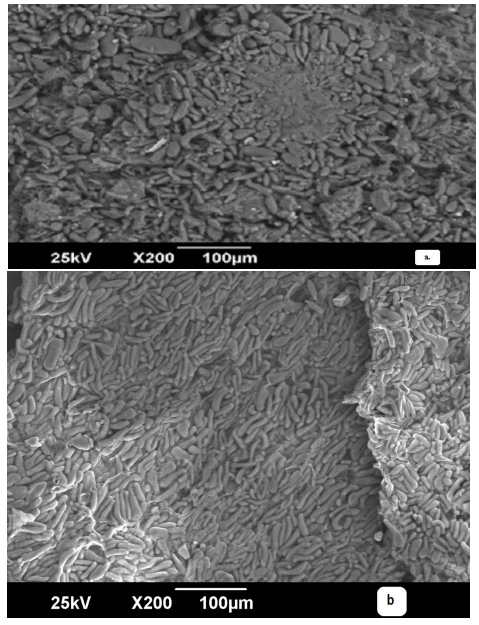


Figure 1. SEM image of plantain peel (a) before and (b) after Cr(VI) adsorption.
Source: Authors

In Figure 1 (b), the banana peel is shown after the adsorption process, where the cylindrical surface of the adsorbent was seen to be covered by Cr(VI) ions, manifested by white micro precipitated particles on the surface. Therefore, when Cr(VI) is captured, the ion adsorption mechanism in biomaterials is produced by the ion exchange between the metals under study and the active centers of the material. (Rodriguez-Narvaez, Peralta-Hernandez, Goonetilleke, and Bandala, 2017; Medellín-Castillo, 2017; Chen, An, Sun, Gao, and Qian, 2018).

After carrying out the Cr(VI) adsorption test using plantain peels and oil palm bagasse as bio-adsorbent, the effect of temperature, particle size, and amount of metal adsorbent were evaluated by calculating the capacity of Cr(VI) adsorption. Table 1 summarizes the results for each of the studied conditions. It was experimentally established that, at 55 °C, 0,5 mm, and 0,03 g of biomaterial, the highest adsorption capacity is obtained by using both bio-adsorbents. The results are 110,89 mg/g when using plantain peels, and 325,88 mg/g for palm bagasse oil can.

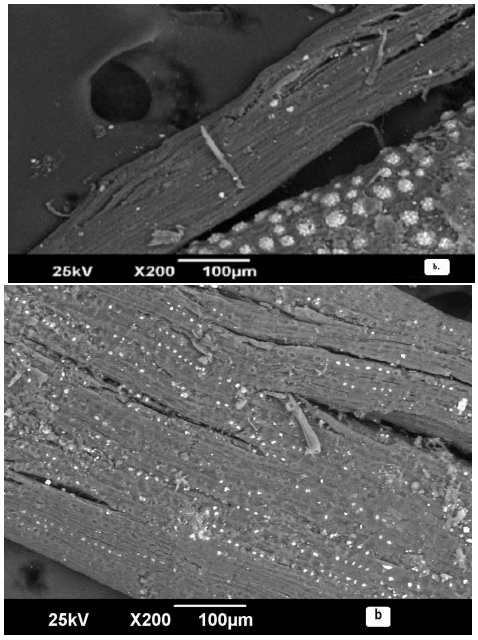


Figure 2. SEM image of oil palm bagasse (a) before and /b) after Cr (VI) adsorption.
Source: Authors

Table 1. The adsorption capacity of Cr (VI) on plantain peels and oil palm bagasse

Temperature (°C)	Particle size (mm)	Mass of adsorbent (g)	Plantain peel	Palm bagasse
40,0	0,355	0,15	52,93	38,36
70,0	1,0	0,15	66,24	15,82
40,0	1,0	0,5	19,95	18,01
55,0	0,5	0,62	15,99	14,89
55,0	0,13	0,325	30,46	7,15
80,2269	0,5	0,325	30,46	30,77
55,0	1,22	0,325	30,46	30,77
29,7731	0,5	0,325	30,46	25,22
55,0	0,5	0,03	110,89	325,88
70,0	0,355	0,15	66,24	66,67
40,0	1,0	0,15	53,11	34,21
55,0	0,5	0,325	30,46	29,48
40,0	0,355	0,5	20	19,99
70,0	1,0	0,5	19,88	4,69
70,0	0,355	0,5	19,90	19,99
55,0	0,5	0,325	30,61	29,48

Source: Authors

Figure 3 initially shows a gradual increase in adsorption capacity with an increase in temperature. This is due to the increase in interactions between adsorbate and adsorbent.However, at high temperatures, its effect is diminished (Sakulthaew, Chokejaroenrat, Poapolathep, Satapanajaru, and Poapolathep, 2017), with a noticeable change at 55 °C.

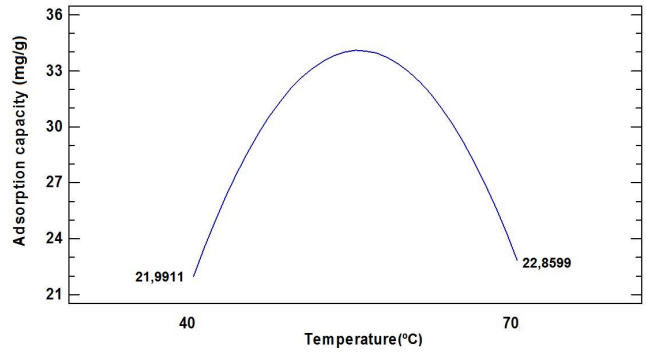


Figure 3. Temperature effect over Cr(VI) adsorption using oil palm bagasse.
Source: Authors

The reason for the descending behaviour of Cr(VI) removal at a high temperature and the immediate adsorption without positive effect due to its increase is that the metal ions in the solution could interact more with the binding sites at a lower temperature. As the temperature increases, so does the mobility of the ions, and the attractive forces between the sorbent and the metal ions decrease, thus decreasing the sorption efficiency of the adsorbent (Abbas et al., 2017). This behaviour suggests that the system is spontaneous and endothermic up to the evaluated intermediate temperature (55 °C). Also, when the temperature tends to higher values, the process becomes unfeasible under these conditions (Ajmani, Shahnaz, Subbiah, and Narayanasamy, 2019).

Figure 4 exhibits an increasing trend in adsorption capacity with an increase in temperature, which could be due to the chemical interaction between adsorbate and adsorbent, and the subsequent creation of new adsorption sites (Jain et al., 2018).

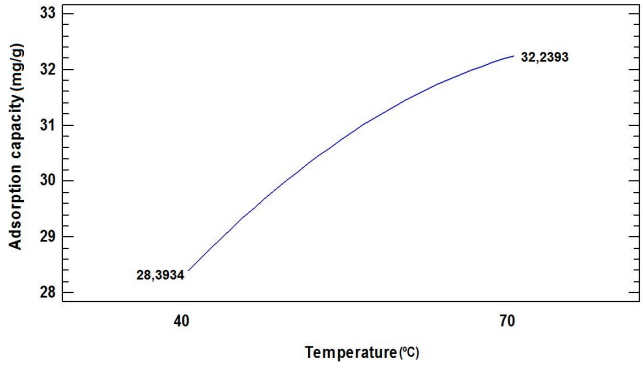


Figure 4. Temperature effect over Cr(VI) adsorption using plantain peels.
Source: Authors

The possibility of presenting a higher diffusion rate of chromium ions in the shells is due to the formation of bonds between the ions and the active functional groups on the adsorbent. This overcomes the activation energy barrier and improves the rate of intraparticle diffusion, which can also

interfere with the process (Aksu and Kabasakal, 2004) and increases the collision sequences between the adsorbate and the bio-adsorbent. Therefore, a more efficient adsorption and greater content of Cr(VI) ions are obtained in the shells (Soniya and Krishnakumar, 2015). This behaviour coincides with the values reported when using coconut shell; it was found that an increase in temperature between 50 °C and 90 °C benefits the Cr(VI) removal process. This phenomenon is due to its endothermic nature and the increase in the kinetic rate of adsorption (Ijeamaka et al., 2018).

Enthalpy values of adsorption $\Delta H°$, adsorption entropy $\Delta S°$, and Gibbs energy $\Delta G°$ were calculated to establish the type of Cr(VI) adsorption by the studied materials. They allow setting the favourable process and the effect that temperature has on it (Figueroa, Moreno and Hormanza, 2014).

Table 2. Thermodynamic parameters for (VI) adsorption in a batch system using oil palm bagasse and plantain peels

Equations/Biomass	Temperature (K)	$\Delta G°$ (KJ/mol)	$\Delta H°$ (KJ/mol)	$\Delta S°$ (KJ/mol K)
Oil Palm Bagasse	303,15 (40 °C)	0,0820	22,719	0,0746
	328,15 (55 °C)	-1,7848		
	353,15 (70 °C)	-3,6516		
Plantain peel	303,15 (40 °C)	-26,0468	-206,320	-0,5946
	328,15 (55 °C)	-11,1801		
	328,15 (55 °C)	398,494	240,349	0,7665
	353,15 (70 °C)	240,805		

Source: Authors

From Table 2, it was established that the process of Cr(VI) adsorption with oil palm bagasse becomes more favorable at higher temperatures, as it is a spontaneous process. Likewise, the enthalpy change indicates that the adsorption is endothermic and has a physical adsorption mechanism.

Besides, the positive value in entropy reveals the increase in randomness in the solid-liquid interface and evinces the high affinity of Cr(VI) ions with the residual material, as well as the probability of some structural changes due to the formation of Cr(VI) links -functional groups of bagasse in the interface. This value also hints to the possible reversibility of the process, which suggests that Cr(VI) ions replace some water molecules previously adsorbed on the adsorbent surface. These replaced water molecules reach a higher entropy by the translation of Cr(VI) ions, which allows the prevalence of randomness (Bedin, Martins, Cazetta, Pezoti, and Almeida, 2016; Jaiswal, Mani, Banerjee, Gautam, and Chattopadhyaya, 2015; Mthombeni, Onyango, and Aoyi, 2015). Similar values of enthalpy and entropy in the adsorption process were determined and recorded in the study on Cr(VI) adsorption by modified rice straw (Lin et al., 2018).

On the other hand, the positive values of the change of Gibbs free energy ($\Delta G°$) in adsorption with plantain peels indicate that the process is thermodynamically feasible (spontaneous) at temperatures from 303,15 to 328,15 K and unfavorable (not spontaneous) at temperatures from 328,15 to 353,15 K (Khan, Nazir, Ali, and Kumar, 2017). Regarding the enthalpy values for the different stages of the Cr(VI) removal process, the negative value indicates that the process is exothermic in nature. In contrast, the positive value reveals the endothermal nature at temperatures below 353,15 K. It also suggests that the adsorption mechanism was physicochemical (Khan et al., 2017; Yang, Yu, and Qiu, 2014).

Conclusions

The effect of temperature on the Cr (VI) adsorption process was studied using plantain peels and oil palm bagasse as adsorbents, by determining the thermodynamic adsorption parameters and their physical significance on metallic removal. It was found that the biomaterials have a fibrous and cylindrical structure after the elimination of the ion micro-precipitates. It was determined that the process that controls the adsorption is the ion exchange between the adsorbent and the solution. The increase in temperature had a significant incidence on the removal, as it favors the adsorption capacity of the metal. From the thermodynamic parameters, it was determined that the adsorbents have a high affinity for Cr(VI) ions. The process is spontaneous up to 328,15 K. The removal is reversible on palm bagasse, exothermic up to 328,15 K on plantain peels and endothermic at higher temperatures. From the high adsorption capacities, these two materials are suggested as Cr(VI) adsorbents in aqueous solution.

Acknowledgements

The authors thank the Universidad de Cartagena (Colombia) for the support in the development of this work regarding laboratory, software use, and time for their researchers.

References

Abbas, A., Ajaz, M., Sher, M., Imran, M., Nawaz, M., Tremel, W., Zajif, S., and Hussain, I. (2017). Design, characterization and evaluation of hydroxyethylcellulose based novel regenerable supersorbent for heavy metal ions uptake and competitive adsorption. *International Journal of Biological Macromolecules*, *102*, 170-18. j.ijbiomac.2017.04.024

Abdelhafez, A. A., and Li, J. (2016). Removal of Pb (II) from aqueous solution by using biochars derived from sugar cane bagasse and orange peel. *Journal of the Taiwan Institute of Chemical Engineers, 61*, 367-375. 10.1016/j.jtice.2016.01.005

Ajmani, A., Shahnaz, T., Subbiah, S., and Narayanasamy, S. (2019). Hexavalent chromium adsorption on virgin, biochar, and chemically modified carbons prepared from Phanera vahlii fruit biomass: equilibrium, kinetics, and thermodynamics approach. *Environmental Science and Pollution Research, 26*(31), 32137-32150. 10.1007/s11356-019-06335-z

Akram, M., Bhatti, H. N., Iqbal, M., Noreen, S., and Sadaf, S. (2017). Biocomposite efficiency for Cr (VI) adsorption: kinetic, equilibrium and thermodynamics studies. *Journal of Environmental Chemical Engineering, 5*(1), 400-411. 10.1016/j.jece.2016.12.002

Anastopoulos, I. and Kyzas, G. Z. (2016). Are the thermodynamic parameters correctly estimated in liquid-phase adsorption phenomena? *Journal of Molecular Liquids, 218*, 174-185. 10.1016/j.molliq.2016.02.059

Asuquo, E., Martin, A., Nzerem, P., Siperstein, F., and Fan, X. (2017). Adsorption of Cd(II) and Pb(II) ions from aqueous solutions using mesoporous activated carbon adsorbent: Equilibrium, kinetics and characterisation studies. *Journal of Environmental Chemical Engineering, 5*(1), 679-698. 10.1016/j.jece.2016.12.043

Bangaraiah, P. (2020). Kinetic and Equilibrium Study on Biosorption of Chromium using Tamarind Fruit Shell. *Research Journal of Pharmacy and Technology, 13*(5), 2340-2344. 10.5958/0974-360X.2020.00421.7

Bedin, K., Martins, A., Cazetta, A., Pezoti, O., and Almeida, V. (2016). KOH-activated carbon prepared from sucrose spherical carbon: Adsorption equilibrium, kinetic and thermodynamic studies for Methylene Blue removal. *Chemical Engineering Journal, 286*, 476-484. 10.1016/j.cej.2015.10.099

Brahmaiah, T., Spurthi, L., Chandrika, K., Ramanaiah, S., and Prasad, K. S. (2015). Kinetics of heavy metal (Cr & Ni) removal from the wastewater by useing low cost adsorbent. *World Journal of Pharmacy and Pharmaceutical Sciences, 4*(11), 1600-1610. https://www.researchgate.net/publication/291102751_KINETICS_OF_HEAVY_METAL_Cr_Ni_REMOVAL_FROM_THE_WASTEWATER_BY_USEING_LOW_COST_ADSORBENT

Carreño-Sayago, U. F. (2016). Diseño y evaluación de un biosistema de tratamiento a escala piloto de aguas de curtiembres a través de la Eichhornia crassipes. *Revista Colombiana de Biotecnología, 18*(2), 74-81. 10.15446/rev.colomb.biote.v18n2.52271

Casarin, J., Jr, A. C. G., Coelho, G. F., Corazza, M. Z., de Oliveira, F. M., Ricardo, C., Pinheiro, A., Meier, M., and Dragunski, D. C. (2016). Adsorption of metallic ions Cd2+, Pb2+, and Cr3+ from water samples using Brazil nut shell as a low-cost biosorbent. In Mishra A. K. (ed.), *Smart Materials for Waste Water Applications* (pp. 311-334), Hoboken, NY: Wiley-Scrievener. 10.1002/9781119041214.ch12

Chen, Y., An, D., Sun, S., Gao, J., and Qian, L. (2018). Reduction and removal of chromium VI in water by powdered activated carbon. *Materials, 11*(2), 269. 10.3390/ma11020269

Chieregato, B. and Tapia, D. (2016). Isolation and characterization of starch from babassu mesocarp. *Food Hydrocolloids, 55*, 47-55. 10.1016/j.foodhyd.2015.11.001

Figueroa, D., Moreno, A., and Hormaza, A. (2015). Equilibrio, termodinámica y modelos cinéticos en la adsorción de Rojo 40 sobre tuza de maíz. *Revista Ingenierías Universidad de Medellín, 14*(26), 105-120. 10.22395/rium.v14n26a7

Guiza, S. (2017). Biosorption of heavy metal from aqueous solution using cellulosic waste orange peel. *Ecological Engineering, 99*, 134-140. 10.1016/j.ecoleng.2016.11.043

Haroon, H., Ashfaq, T., Gardazi, S. M. H., Sherazi, T. A., Ali, M., Rashid, N., and Bilal, M. (2016). Equilibrium kinetic and thermodynamic studies of Cr (VI) adsorption onto a novel adsorbent of Eucalyptus camaldulensis waste: Batch and column reactors. *Korean Journal of Chemical Engineering, 33*(10), 2898-2907. 10.1007/s11814-016-0160-0

Hubbe, M. A., Azizian, S., and Douven, S. (2019). Implications of apparent pseudo-second-order adsorption kinetics onto cellulosic materials: A review. *BioRes, 14*(3), 7582-7626. https://bioresources.cnr.ncsu.edu/resources/implications-of-apparent-pseudo-second-order-adsorption-kinetics-onto-cellulosic-materials-a-review/

Ijeamaka, E. C., Christian, O. C., Fabian, O. O., MaryJane, I. I., and Joseph, K. T. (2018). Isotherm Studies of Adsorption of Cr (VI) Ions onto Coconut Husk. International Journal of Biochemistry, *Biophysics and Molecular Biology, 3*(2), 38-44. 10.11648/j.ijbbmb.20180302.13

Jain, M., Yadav, M., Kohout, T., Lahtinen, M., Garg, V., and Sillanpää, M. (2018). Development of iron oxide/activated carbon nanoparticle composite for the removal of Cr(VI), Cu(II) and Cd(II) ions from aqueous solution. *Water Resources and Industry, 20*, 54-74. 10.1016/j.wri.2018.10.001

Jaiswal, A., Mani, R., Banerjee, S., Gautam, R., and Chattopadhyaya, M. (2015). Synthesis of novel nano-layered double hydroxide by urea hydrolysis method and their application in removal of Chromium (VI) from aqueous solution: Kinetic, thermodynamic and equilibrium studies. *Journal of Molecular Liquids, 202*, 52-61. 10.1016/j.molliq.2014.12.004

Kan, C., Ibe, A., Rivera, K., Arazo, R., and de Luna, M. (2017). Hexavalent chromium removal from aqueous solution by adsorbents synthesized from groundwater treatment residuals. *Sustainable Environment Research, 27*(4), 163-171. 10.1016/j.serj.2017.04.001

Kecili, R. and Hussain, C. M. (2018). Mechanism of Adsorption on Nanomaterials. In Hussain, C. M. (Ed.) *Nanomaterials in Chromatography* (pp. 89-115), Amsterdam, Netherlands: Elsevier. 10.1016/b978-0-12-812792-6.00004-2

Khan, T., Nazir, M., Ali, I., and Kumar, A. (2017). Removal of Chromium (VI) from aqueous solution using guar gum–nano zinc oxide biocomposite adsorbent. *Arabian Journal of Chemistry, 10*(2), s2388-s2398. 10.1016/j.arabjc.2013.08.019

Lin, C., Luo, W., Luo, T., Zhou, Q., Li, H., and Jing, L. (2018). A study on adsorption of Cr (VI) by modified rice straw: Characteristics, performances and mechanism. *Journal of Cleaner Production, 196*, 626-634. 10.1016/j.jclepro.2018.05.279

Maina, I. W., Obuseng, V., and Nareetsile, F. (2016). Use of Moringa oleifera (Moringa) seed pods and Sclerocarya birrea (Morula) nut shells for removal of heavy metals from wastewater and borehole water. *Journal of Chemistry, 2016*. 10.1155/2016/9312952

Marimón-Bolívar, W., Tejeda-Benítez, L., and Herrera, A. (2018). Removal of mercury (II) from water using magnetic nanoparticles coated with amino organic ligands and yam peel biomass. *Environmental Nanotechnology, Monitoring and Management, 10*, 486-493. 10.1016/j.enmm.2018.10.001

Martínez-Mendoza, K. L., Barraza, J. M., Marraiaga, Cabrales, N., Machuca-Martínez, F., Barajas, M., and Romero, M. (2020) Production and characterization of activated carbon from coal for gold adsorption in cyanide solutions. *Ingeniería e Investigación, 40*(1). 10.15446/ing.investig.v40n1.80126

Medellín-Castillo, N. A., Hernández-Ramírez, M. G., Salazar-Rábago, J. J., Labrada-Delgado, G. J., and Aragón-Piña, A. (2017). Bioadsorción de Plomo (II) presente en solución acuosa sobre residuos de fibras naturales procedentes de la industria ixtlera (Agave lechuguilla Torr. Y Yucca carnerosana (Trel.) McKelvey). *Revista internacional de contaminación ambiental, 33*(2), 269-280. 10.20937/rica.2017.33.02.08

Mthombeni, N., Onyango, M., and Aoyi, O. (2015). Adsorption of hexavalent chromium onto magnetic natural zeolite-polymer composite. *Journal of the Taiwan Institute of Chemical Engineers, 50*, 242-251. 10.1016/j.jtice.2014.12.037

Núñez-Zarur, J., Tejada-Tovar, C., Villabona-Ortíz, A., Acevedo, D., and Tejada-Tovar, R. (2018). Thermodynamics, kinetics and equilibrium adsorption of Cr (VI) and Hg (II) in aqueous solution on corn husk (Zea Mays). *International Journal of ChemTech Research, 11*(5), 265-280. 10.20902/ijctr.2018.110529

Razi, M. A. M., Al-Gheethi, A., and Za, I. A. (2018). Removal of heavy metals from textile wastewater using sugarcane bagasse activated carbon. *International Journal of Engineering and Technology, 7*(430), 112-115. 10.14419/ijet.v7i4.30.22066

Rico, I. R., Carrazana, R. C., Karna, N. K., Iáñez-Rodríguez, I., and de Hoces, M. C. (2018). Modeling the mass transfer in biosorption of Cr (VI) y Ni (II) by natural sugarcane bagasse. *Applied Water Science, 8*(2), 55. 10.1007/s13201-018-0692-z

Rodriguez-Narvaez, O. M., Peralta-Hernandez, J. M., Goonetilleke, A., and Bandala, E. R. (2017). Treatment technologies for emerging contaminants in water: a review. *Chemical Engineering Journal, 323*, 361-380. 10.1016/j.cej.2017.04.106

Tejada, C., Quiñones, E., Tejeda, L., and Marimón W. (2015). Absorción de Cromo Hexavalente en soluciones acuosas por cascaras de naranja (Citrus sinensis). *Producción + Limpia, 10*(1), 9-21. 10.22507/pml.v10n1a1

Tejada-Tovar, C., Herrera-Barros, A., and Villabona-Ortíz, A. (2020). Assessment of Chemically Modified Lignocellulose Waste for the Adsorption of Cr (VI). *Revista Facultad de Ingeniería, 29*(54), e10298. 10.19053/01211129.v29.n54.2020.10298

Tran, H. N., You, S.-J., Hosseini-Bandegharaei, A., and Chao, H.-P. (2017). Mistakes and inconsistencies regarding adsorption of contaminants from aqueous solutions: A critical review. *Water Research, 120*, 88-116. 10.1016/j.watres.2017.04.014

Sakulthaew, C., Chokejaroenrat, C., Poapolathep, A., Satapanajaru, T., and Poapolathep, S. (2017). Hexavalent chromium adsorption from aqueous solution using carbon nano-onions (CNOs). *Chemosphere, 184*, 1168-1174. 10.1016/j.chemosphere.2017.06.094

Soniya, M. and Krishnakumar, G. (2015). Biosorption of Heavy Metals from Aqueous solution using Mangrove fern Acrostichumaureum L. leaf Biomass as a Sorbent. *International Research Journal of Environment Sciences, 4*(11), 25-31. http://www.isca.in/IJENS/Archive/v4/i11/5.ISCA-IRJEVS-2015-162.pdf

Vargas-Niño, C., Carriazo, J. G., and Castillo, E. (2011). A study of low–cost adsorbent ma-cost adsorbent ma-cost adsorbent ma-cost adsorbent materials for removing Cr(VI) from materials for removing Cr(VI) from terials for removing Cr(VI) from terials for removing Cr(VI) from aqueous waste effluent aqueous waste effluent. *Ingeniería e Investigación, 31*(1), 154-162. https://revistas.unal.edu.co/index.php/ingeinv/article/view/20537

Villabona-Ortíz, A., Tejada-Tovar, C. N., and Ortega-Toro, R. (2020). Modelling of the adsorption kinetics of chromium (VI) using waste biomaterials. *Revista Mexicana de Ingeniería Química, 19*(1), 401-408. 10.24275/rmiq/ia650

Wassie, A. B. and Srivastava, V. C. (2016). Teff straw characterization and utilization for chromium removal from wastewater: Kinetics, isotherm and thermodynamic modelling. *Journal of Environmental Chemical Engineering, 4*(1), 1117-1125. 10.1016/j.jece.2016.01.019

Yang, J., Yu, M., and Qiu, T. (2014). Adsorption thermodynamics and kinetics of Cr(VI) on KIP210 resin. *Journal of Industrial and Engineering Chemistry, 20*(2), 480-486. 10.1016/j.jiec.2013.05.005

Research Article / Electrical-Electronic Engineering http://dx.doi.org/10.15446/ing.investig.v40n3.89091

Power Quality in AC Islanded Microgrids: Technical Framework and State of the Art Review

Calidad de Potencia en Microrredes Aisladas AC: Marco de Referencia Técnico y Revisión del Estado del Arte

Vanessa Quintero-Molina[1], Ana María Blanco[2], Miguel Romero-L[3], Jan Meyer[4], and Andrés Pavas[5]

ABSTRACT

In recent years, operation and control strategies in distribution systems have changed due to the increase in the connection of distributed generation sources (DGs). Small local networks with electricity users and DGs are known as microgrids. These microgrids can operate independently (islanded) or in collaboration (interconnected) with the main network or other microgrids. Some advantages of interconnected microgrids include the reduction of losses, an increase in reliability, and decentralized operation under fault conditions. Nevertheless, when a microgrid operates in islanded mode, its electrical characteristics change and, consequently, the severity of power quality disturbances can increase, as well as their negative impact on electronic devices (loads and DG devices). This paper presents a comprehensive literature review of existing studies on power quality disturbances in islanded microgrids and identifies the most relevant needs for future research on this topic. Detailed information is analyzed to compare the differences between disturbance levels in both interconnected and islanded microgrids. In the case of harmonic disturbances, the impact of the different microgrid configurations is also analyzed.

Keywords: AC islanded microgrid, disturbances, field measurements, power quality

RESUMEN

En los últimos años, las estrategias de operación y control en los sistemas de distribución han cambiado debido al aumento en la conexión de las fuentes de generación distribuidas (DGs). Las pequeñas redes locales con usuarios de electricidad y DGs se conocen como microrredes. Estas microrredes pueden funcionar de forma independiente (aislada) o en colaboración (interconectadas) con la red principal u otras microrredes. Algunas ventajas de las microrredes interconectadas incluyen la reducción de pérdidas, aumento en la confiabilidad y la operación descentralizada bajo condiciones de falla. No obstante, cuando una microrred opera en modo aislado, sus características eléctricas cambian y, en consecuencia, la gravedad de las perturbaciones de calidad de potencia podrían aumentar, así como su impacto negativo en dispositivos electrónicos (cargas y dispositivos DG). Este artículo presenta una revisión literaria exhaustiva de los estudios existentes sobre perturbaciones de calidad de potencia en microrredes aisladas e identifica las necesidades más relevantes de investigación futura sobre este tema. Se analiza información detallada para comparar las diferencias entre los niveles de perturbaciones en microrredes interconectadas y aisladas. En el caso de perturbaciones causadas por armónicos, se analiza también el impacto de las diferentes configuraciones de microrredes.

Palabras clave: microrred aislada AC, perturbaciones, mediciones de campo, calidad de potencia

Received: July 12th, 2020
Accepted: November 9th, 2020

Introduction

Increased involvement of renewable energy sources in distribution systems has grown in recent years as a response to environmental and energy policies, which has led to substantial changes in the traditional operation of power systems. Instead of big generators and extended transmission lines, now there are several small generators distributed around the loads in the distribution system. This new system configuration begets a new concept called 'microgrid'. According to the literature, a microgrid is an electricity distribution system containing generation and consumption that can operate in a controlled and coordinated way with the main grid (interconnected mode) or in islanded mode (Bollen et al., 2017a; CIGRE, 2018). Islanded mode implies that the microgrid can be disconnected and operated autonomously.

[1]B.Eng. and M.Eng. in Electrical Engineer, Universidad Nacional de Colombia, Colombia. Affiliation: Research Assistant, Universidad Nacional de Colombia, Colombia. E-mail: vmquinterom@unal.edu.co
[2]B.Eng. and M.Eng. in Electrical Engineer, Universidad Nacional de Colombia, Colombia. Dr.-Ing. in Electrical Engineering, Technische Universität Dresden, Germany. Affiliation: Scientific Researcher, Technische Universität Dresden, Germany. E-mail: ana.blanco@tu-dresden.de
[3]B.Eng., M.Eng. and PhD in Electrical Engineer, Universidad Nacional de Colombia, Colombia. Affiliation: Research Assistant, Universidad Nacional de Colombia, Colombia. E-mail: mfromerol@unal.edu.co
[4]B.Eng., M.Eng. and Dr.-Ing, Habil. Electrical Engineering, Technische Universität Dresden, Germany. Affiliation: Senior Research Assistant, Technische Universität Dresden, Germany. E-mail: jan.meyer@tu-dresden.de
[5]B.Eng., M.Eng., and PhD in Electrical Engineer, Universidad Nacional de Colombia, Colombia. Affiliation: Associate Professor, Universidad Nacional de Colombia, Colombia. E-mail: fapavasm@unal.edu.co

How to cite: Quintero-Molina, V., Blanco, A., Romero-L, M., Meyer, J., Pavas, A. (2020). Power Quality in AC Islanded Microgrids: Technical Framework and State of the Art Review. *Ingeniería e Investigación*, *40*(3), 29-37. 10.15446/ing.investig.v40n3.89091

As described in Prabaakaran, Chitra, and Kumar (2013) and Kumar and Venkateshwarlu (2013), the increased penetration of non-linear loads and power electronic-interfaced distributed generation systems (DGs) can cause power quality (PQ) issues in microgrids. PQ issues in microgrids such as voltage magnitude variation (voltage band), voltage sags and swells, flicker, harmonics, unbalance, or real and reactive power problems have been analyzed in several papers in recent years. Figure 1 presents the number of publications related to microgrids and PQ studies based on the abstract and citation from the Scopus database.

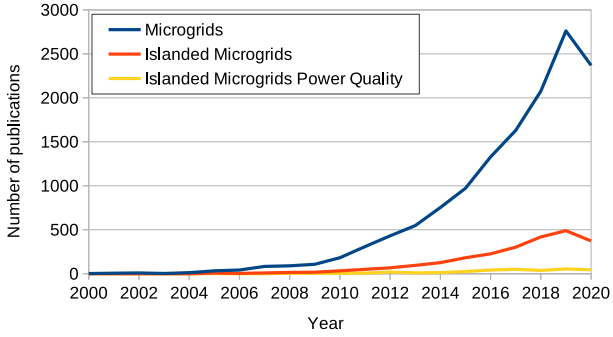


Figure 1. Papers related to power quality in microgrids per year.
Source: Authors

Figure 1 also compares the number of papers that deal with microgrids (all topics), islanded microgrids (all topics), and PQ issues related to islanded microgrids in the time period between 2000 and 2020. Although the number of publications related to microgrids and islanded microgrids per year has increased (8 615 and 1 551 total papers up to 2020 respectively), it is evident that research on power quality issues on islanded microgrids is still scarce (219 papers up to 2020), thus representing 2,56% of all research about microgrids.

Additionally, Figure 2 shows the number of papers produced per country between 2005 and 2020. The countries with the highest publication of articles about islanded microgrids are the United States (454 papers), China (452 papers), and Iran (272 papers). The average ratio between publications analyzing PQ issues of islanded microgrids and the general issues of islanded microgrids is 11% (44 papers) for these three countries. Although countries such as Brazil have made less research (91 papers) than the leading countries, papers about power quality on islanded microgrids represent 31,8% of that country's total.

Research on power quality disturbances in islanded microgrids could be of great importance for countries such as Colombia due to two main aspects: first, Colombia has large non-interconnected areas that could hold potential for the implementation of remote islanded microgrids (López-García, Arango-Manrique, and Carvajal-Quintero, 2018); and second, in recent years, the government has developed policies to encourage the connection of renewable resources (UPME, 2014).

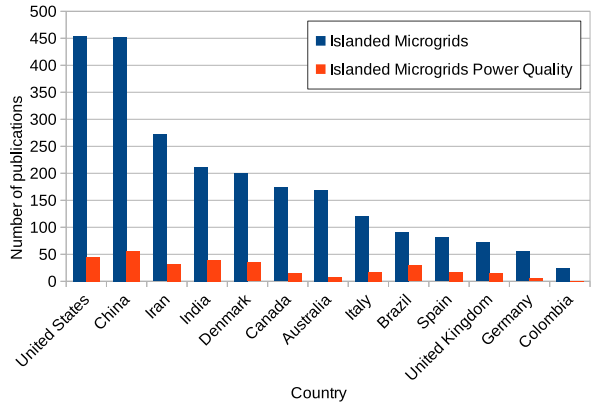


Figure 2. Papers related to power quality in microgrids per Country.
Source: Authors

In this paper, several publications are classified and analyzed with the purpose of systematically identifying the impact of the most common PQ disturbances in microgrids in islanded mode, as well as the levels that can be reached for every disturbance and determining possible future research needs on this topic. Additionally, considering the fast growth of DGs with an electronic interface, this paper focuses only on microgrids containing non-inertial generation sources.

In the next section, the general characteristics of microgrids and their main elements are described. Subsequently, the reviewed papers are classified according to the PQ disturbances in islanded microgrids, and some comparative values for voltage and current harmonics, voltage variations, voltage unbalance, and frequency variations are given. Finally, relevant findings are summarized in a conclusion section.

General Characteristics of Microgrids

In Marnay et al. (2015), microgrids are classified into four types: customer microgrids, utility microgrids, isolated remote microgrids, and virtual microgrids. This classification is shown in Figure 3.

1. *Customer microgrids* are self-governed local distribution networks. Usually, their generation and consumption are downstream of a single point of common coupling (PCC). Customers own these microgrids; therefore, they can operate them relatively unrestricted by the network operator, but they usually cannot transfer (sell) energy to the distribution network.

2. *Utility microgrids* are technically similar to customer microgrids, but the main difference is that the utility owns the microgrid. Accordingly, customers must comply with existing utility regulation policies.

3. *Isolated remote microgrids* are similar to customer microgrids, but these grids are not able to operate while interconnected with the main grid.

4. *Virtual microgrids* cover DGs at multiple sites, but are centrally coordinated so that they can be presented to the power system as a single controlled entity.

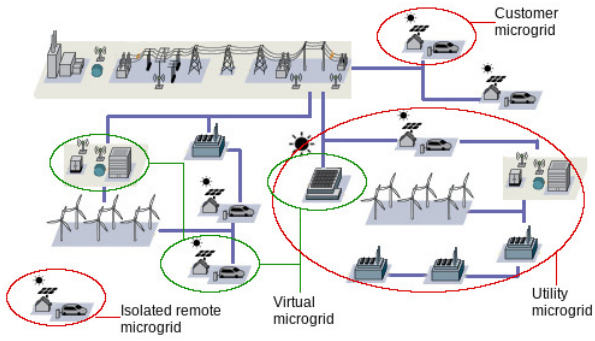


Figure 3. Microgrid Classification.
Source: Authors

For customer and utility microgrids, two operation modes are identified: the interconnected and the islanded modes. During interconnected mode, the main grid can improve their efficiency and reliability due to the integration of DGs near the consumers, as well as making them more resilient. On the other hand, the islanded mode supplies energy to the microgrids during main grid faults, but if this mode operates for a long time, reliability and power quality issues should be taken into consideration (Marnay et al., 2015; Sood and Abdelgawad, 2019).

PQ conditions are related to the main elements of microgrids, which are: DGs, loads, control systems, and energy storage systems.

1. DGs are the energy resources that feed the loads. These sources can be classified into equipment without inertia (100% electronic-based generators, such as photovoltaic systems and small wind turbines), or equipment with inertia (based on synchronous or asynchronous machines, e.g. mini-hydro generators). Microgrids with high amounts of generators without inertia may suffer from voltage and frequency stability problems. Moreover, photovoltaic systems and wind turbines are also sources of waveform distortion (Harmonics, interharmonics and supraharmonics) and other power quality disturbances (CIGRE, 2016).

2. Loads of residential, industrial, and commercial customers can be linear or non-linear devices. Non-linear devices are those that draw a current waveform that is not proportional to the applied voltage. The most common non-linear loads are electronic equipment, like computers, compact fluorescent lamps, battery chargers, and adjustable frequency drives. Electronic loads are common sources of waveform distortion and flicker, which are the major power quality problems in distribution systems.

3. Control systems can be classified into three categories: i) control of the devices themselves (generators and/or loads); ii) control of the interaction between source and consumption (e.g. demand management); and iii) connection and disconnection schemes between the microgrid and the main grid. A control system has a great influence on the interaction between the elements of the microgrid, and, consequently, on the power quality status.

4. Energy storage systems (ESS) are a fundamental aspect of isolated microgrids, which allows the compensation of imbalances between generation and load to ensure the reliability of the energy supply. Storage options include batteries, flywheels, super (or ultra) capacitors, superconducting magnetic energy storage (SMES), compressed air, pumped hydroelectric, and thermal energy storage. Depending on the characteristics of the energy storage system, the behavior of the microgrid may vary.

Assessment of Power Quality Disturbances in AC Islanded Microgrids

The interaction between loads, DGs, and energy storage systems during islanded operations and transition states can lead to adverse impacts on the power quality of microgrid. A bibliographic review of the most common PQ disturbances studied in islanded microgrids during the last decade is shown in Figure 4. Some of the assessed PQ disturbances include harmonic and supraharmonic distortions, voltage variation as long-term and short-term flicker severity, voltage unbalance, and frequency variations. Given that the generation and load equipment are mainly electronic-based, voltage and current waveform could be affected.

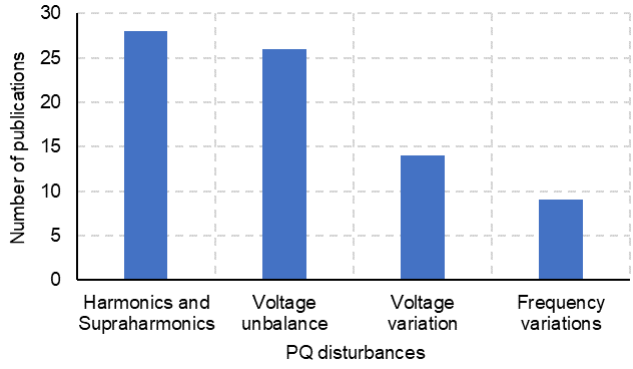


Figure 4. Papers related to islanded microgrid between 2008 and 2018 per PQ disturbance.
Source: Authors

Furthermore, higher volatility and stronger interactions between loads and distributed resources change the value of microgrid impedances in larger ranges with a tendency to weaken them, which will result in more pronounced, more frequent, and longer voltage variations, as well as higher susceptibility to PQ issues. In some cases, the microgrid can also present resonances at low frequencies, which might result in an undesirable increase of voltage harmonics (Bollen et al., 2017a, 2017b). According to Kaushal and Basak (2020), the effect of line impedance on power quality parameters is especially important when the microgrid's status changes

from interconnected to islanded mode. As a result, voltage deviations are the main affected parameter.

In the next subsections, several studies on PQ issues in islanded microgrids are presented. The aim of these sections is to compare the PQ indexes of microgrids in interconnected and islanded modes and to identify which operation mode is more critical for the microgrid.

Harmonics and Supraharmonics

Waveform distortion is one of the most common electromagnetic disturbances in electrical systems. It is the deviation of voltages and currents from the ideal sine waveform. Periodic distorted voltages and currents are usually studied using the Fourier transform, which decomposes the original distorted signals into a linear combination of pure sinusoids at different frequencies. Using the Fourier transform, the total waveform distortion can be separated into:

1. *Harmonics*: spectral components with frequencies that are integer multiples of the fundamental frequency (50 Hz or 60 Hz in most electrical systems).
2. *Interharmonics*: spectral components with frequencies that are non-integer multiples of the fundamental frequency.
3. *Subharmonics*: spectral components with frequencies below the fundamental frequency.
4. *Supraharmonics*: spectral components in the frequency range between 2 and 150 kHz.

In order to assess those disturbances, indexes such as total harmonic distortion (THD) for voltage and current, and the individual magnitude for each harmonic are used. The THD is calculated as:

$$THD_u = \frac{\sqrt{\sum_{h=2}^{40} {U_h}^2}}{U_1} * 100 \quad (1)$$

where U_h is the magnitude of the harmonic component of order h, and U_1 is the magnitude of the fundamental component (component with the fundamental frequency). The THD can be also applied to current signals (THD_I).

Tenti, Paredes, and Mattavelli (2011); Padayattil, Thobias, Thomas, Sebastian, and Pathirikkat (2016); Agundis-Tinajero et al. (2018); and Garde, Casado, Santamaria, and Aguado (2015) report the THD_U from simulations of different islanded microgrids with different load compositions (only non-linear loads, only linear loads, and a mix between linear and non-linear loads). These references show that microgrids with non-linear loads present higher THD_U values than microgrids composed of only linear loads (see Table 1). According to Padayattil et al. (2016) the variation of the THD_U values depends on the type of electronic-based generation and the amount of non-linear load connected.

Table 1. THD_U (%) on islanded microgrid for different configuration generation - load

Generation	Load	THD_U [%]
Electronic-based	Non-linear	29,9
	Linear	3,48 to 15,76
	Mix	3,57 to 41,67

Source: Authors based on references (Tenti et al., 2011; Padayattil et al., 2016; Agundis-Tinajero et al., 2018; Garde et al., 2015).

Some studies implement control strategies in order to reduce the THD_U in those kinds of microgrids. Some of the proposed strategies include adding virtual impedance (Savaghebi, Shafiee, Vasquez, and Guerrero, 2015; Golsorkhi, Lu, Savaghebi, Vasquez, and Guerrero, 2016; Yang, Chen, Dong, and Han, 2015; Micallef, Apap, Spiteri-Staines, Guerrero, 2012; Blanco, Tardelli, Reigosa, Zanchetta, and Briz, 2019), harmonic controller (Zeng, Yi, Zhai, Zhuo, and Wang, 2017; Thai, Raisz, Monti, Ponci, and Ahmadifar 2018; Dehghani, Vahedi, Savaghebi, and Guerrero, 2012), compensations (Phan and Lee, 2019) Fuzzy-PI (Lhachimi, Sayouti, and Kouari, 2020) and artificial neural network (ANN) controllers (Kaushal and Basak, 2020). Table 2 shows the THD_U variation of five islanded microgrids when different control strategies are implemented. In most of the cases, the THD_U is reduced.

Additionally, some approaches, such as the one in Lhachimi et al. (2020), apply control strategies integrating IEEE standards in order to reduce the THD_U levels in microgrids and accomplish regulation limits in both islanded and connected modes.

Table 2. THD_U (%) on islanded microgrid for different configuration generation-load and different control strategies

Generation	Load	Control strategy	THD_U [%]	
			Base case	Control
Electronic-based	Non-linear	Harmonic controller	11,38	4,85
	Non-linear	Compensation	10,3	3,48
	Mix	Virtual impedance	0,6 - 8,7	0,1 - 4,7
	Non-linear	Fuzzy-PI	1,82	4,66
	Non-linear	ANN	12,0	0

Source: Authors based on references (Golsorkhi et al., 2016; Yang et al., 2015; Micallef et al., 2012; Blanco et al., 2019; Zeng et al., 2017; Thai et al., 2018; Dehghani et al., 2012; Phan and Lee, 2019; Kaushal and Basak, 2020).

Studies by Laaksonen and Kauhaniemi (2008); Nomm, Ronnberg, and Bollen (2018a); Kulia, Molinas, Lundheim, and Fosso (2017); Zhou, Yuan, and Wang (2012); and Algaddafi, Brown, Gammon, Altuwayjiri, and Alghamdi (2016) discuss harmonic distortion in a microgrid when it operates in islanded and interconnected modes. A summary of the results is presented in Table 3 for the THD_U index, Table 4 for the THD_I index, and Table 5 for the first odd harmonic voltage magnitudes. For all the studied microgrid systems, generation sources are electronic-based and could include

DGs and ESS. Additionally, four types of load are considered: non-linear load, linear load, motor, and mixed loads (linear and non-linear loads).

As shown in Table 3, THD_U in islanded mode increases by up to 14,6% compared to the values in interconnected mode. Microgrids with only non-linear loads show the highest THD_U values in both operation modes. According to Laaksonen and Kauhaniemi (2008), microgrid components such as filters and the switching frequency of converters directly affect harmonic distortion, especially if the load of the microgrids is completely non-linear.

Voltage distortion levels in islanded microgrids depend not only on the type of loads, but also on their size. In Hicks, Baghzouz, and Haddad (2018), THD_U is 2,3% in a no-load condition islanded microgrid. When a non-linear load (980W) is connected, the THD_U increases to 3,6%. However, when a larger linear load (1,9 kW) is connected to the microgrid, it reaches up to 6,9%.

Table 3. Comparison of THD_U (%) in interconnected (interc) and islanded (island) modes for different configuration generation - load

Generation	Load	THD_U [%]		
		Interc	Island	Difference
Electronic-based	Non-linear	0,4 - 10,4	1,1 - 25,0	0,7 - 14,6
	Motor	0,4	1,1	0,7
	Mix	0,4 - 1,88	1,3 - 8,0	0,9 - 6,12
	Linear	1,6	2,5	0,9

Source: Authors based on references (Laaksonen and Kauhaniemi, 2008; Nomm et al., 2018a; Kulia et al., 2017; Zhou et al., 2012; Algaddafi et al., 2016).

In Table 4, the THD_I values for interconnected and islanded modes are presented. Interestingly, THD_I values not always increase during islanded mode. This is caused by the change in network impedance, which results in a different behavior between loads and generators. According to Laaksonen and Kauhaniemi (2008), Nomm et al. (2018a), and Zhou et al. (2012), when the microgrid has a mixed load, THD_I values increase up to 0,87%, but if the load were completely linear, the THD_I would have similar behavior or could even be reduced, as in the results obtained by Algaddafi, et al. (2016), in interconnected and islanded modes.

Table 4. Comparison of THD_I (%) in interconnected (interc) and islanded (island) modes for different configuration generation - load.

Generation	Load	THD_I [%]		
		Interc	Island	Difference
Electronic-based	Non-linear	1,08	0,61	-0,47
	Motor	0,27	0,29	0,02
	Mix	0,24	1,11	0,87
	Linear	1,60	1,12	-0,48

Source: Authors based on references (Laaksonen and Kauhaniemi, 2008; Nomm et al., 2018a; Kulia et al., 2017; Zhou et al., 2012; Algaddafi et al., 2016).

Another way to assess waveform distortions is through the harmonic voltage for each harmonic component. Table 5 presents the assessed harmonic voltage for the 3rd to the 15th harmonic. As a general result all harmonic components increase when the microgrid operates in islanded mode, up to 1,44 V (Nomm et al., 2019).

Table 5. Comparison of harmonic voltage (hv3 - hv15) in interconnected (interc) and islanded (island) modes for different configuration generation-load.

Generation	Load		Harmonic magnitude [V]		
		Harmonic	Interc	Island	Difference
Electronic-based	Mix	3^{rd}	0,52	1,96	1,44
		5^{th}	0,90	2,00	1,10
		7^{th}	0,60	1,45	0,85
		9^{th}	0,44	1,38	0,94
		11^{th}	0,17	0,66	0,49
		15^{th}	0,15	0,18	0,03

Source: Authors based on references (Nomm et al., 2019).

In Nomm et al. (2018a), THD_U values are compared against regulatory limits. PQ standards as IEEE Std 519 (IEEE, 2014) and SS-EN 50160 (SEK, 2011) are applied. The study shows that, while in islanded mode, THD_U exceeds the established limit, not only in magnitude but also in duration. This means that, within a 48-hour period, THD_U values exceed the 8% limit for 10 hours. A comparison against compatibility levels provided in IEC 61000-2-2 as basis for EMC coordination was not performed.

Table 6 summarizes the maximum voltage magnitude in the supraharmonic frequency range interconnected and islanded operation mode in a microgrid with non-linear loads. Laaksonen and Kauhaniemi (2008) and Ronnberg et al. (2017) identified supraharmonics while in interconnected and islanded modes. Both studies show frequency components near the switching frequency (10 - 12 kHz) of the converters, thus jeopardizing microgrid operations due to possible resonances. Also, it is noted that during islanded operation, more components are present, and the magnitudes become higher by up to 3,5% compared to the interconnected mode.

Table 6. Comparison of the maximum harmonic magnitude (V) in the supraharmonic range in interconnected (interc) and islanded (island) modes for different configuration generation - load.

Generation	Load	Max harmonic magnitude [V] frequency range [2-150] kHz		
		Interc	Island	Difference
Electronic-based	Non-linear	1,3	4,8	3,5

Source: Authors based on references (Laaksonen and Kauhaniemi, 2008; Ronnberg et al., 2017).

Voltage variations

The most common indexes to assess fast voltage variations include short-term flicker severity (P_{ST}), long-term flicker severity (P_{LT}), and the 3-second very short variations index (Ronnberg et al., 2017; Nomm et al., 2019), while slow voltage variations are usually quantified by the variation of 10 min values. Tables 7, 8, and 9 show the maximum value for

each index when the microgrid operates in interconnected and islanded modes.

According to Tables 7 and 8, higher values for short-term and long-term flicker were observed while in islanded mode. The P_{ST} can increase up to 4,66 and, once again, the increase depends on the type of loads.

Table 7. Comparison of maximum P_{ST} value in interconnected (interc) and islanded (island) modes for different configuration generation - load

Generation	Load	maximum P_{ST} value		
		Interc	Island	Difference
Electronic-based	Non-linear	0,45	0,55	0,1
	Mix	2,34	7,00	4,66

Source: Authors based on references (Ronnberg et al., 2017; Nomm et al., 2019).

Table 8. Comparison of P_{LT} value in interconnected (interc) and islanded (island) modes for different configuration generation - load

Generation	Load	maximum P_{LT} value		
		Interc	Island	Difference
Electronic-based	Mix	1,20	2,99	1,79

Source: Authors based on references (Ronnberg et al., 2017; Nomm et al., 2019).

The results for 10-min variations and 3-second very short variations (VSV) obtained by Nomm et al. (2019) show that the 10-minute VSV values are less severe during islanded operation (see Table 9). However, the 3-second VSV values reach higher maximums during islanded operation. The possible cause of the 3s very short variations in islanded mode is the switching behavior of inverter operation due to solar irradiance variation.

Table 9. Comparison of maximum very short variations (10 minutes and 3 seconds) in interconnected (interc) and islanded (island) modes for different configuration generation - load.

Generation	Load	Interconnected		Islanded	
		Max 10 min	Max 3 s	Min 10 min	Min 3 s
Electronic-based	Mix	7,89	15,7	6,34	18,3

Source: Authors based on references (Ronnberg et al., 2017; Nomm et al., 2019).

In Ronnberg et al. (2017) and Nomm et al. (2019), microgrid measurements in interconnected and islanded operation modes are compared against regulatory limits, namely the values proposed in IEEE Std 519 (IEEE, 2014) or EN 50160 (SEK, 2011). For Ronnberg et al. (2017), voltage variation indexes are below regulatory limits, but Nomm et al. (2019) states that P_LT values exceed the EN 50160 limit for a maximum of three weeks during a complete year. Both studies measured the same type of loads but received different results. For that reason, it is important to develop further and systematic measurement studies.

Voltage unbalance

The studies presented in Ronnberg et al. (2017) and Nomm et al. (2019) show that voltage unbalance in islanded microgrids is higher than in the interconnected mode. The results of both studies are summarized in Table 10.

Table 10. Comparison of the maximum voltage unbalance (%) in interconnected (interc) and islanded (island) modes for different configuration generation - load

Generation	Load	Maximum voltage unbalance [%]		
		Interc	Island	Difference
Electronic-based	Non-linear	0,95	1,1	0,15
	Mix		1,77	2,83

Source: Authors based on references (Ronnberg et al., 2017; Nomm et al., 2019).

Two operational modes, A and B, in islanded operation are studied in Nomm et al. (2019), depending on their short circuit impedance: mode A fluctuates between 0,1 and 1,0 Ω; and mode B between 0,1 and 2,0 Ω. Voltage unbalance is higher in operational mode B than in mode A with an increase of up to 2 times the values while in mode A. The same analysis for the other PQ parameters, namely voltage variation and harmonic distortion, showed similar results.

Frequency variations

Due to the interactions between loads and DGs, frequency variations in an islanded microgrid can be more pronounced, more frequent, and longer (Bollen et al., 2017b). For example, in Nomm et al. (2018b) measurements were performed in a microgrid for 48 weeks. As a result, 89% of the frequency variations were outside the 42,5-57,5 Hz band defined in EN 50160 (SEK, 2011) for islanded grids. Consequently, short interruptions occurred commonly, which might have some adverse effects on certain equipment.

In a network with high DG penetration, some of the DGs may be disconnected from it after a fault. During periods of high DG generation, this might result in local variations in system frequency. This is an evolving issue, and current distribution systems are not necessarily equipped to properly support system frequency, but depend on the interconnected system for load-generation balance (Bollen et al., 2017b). Studies such as Rodrigues et al. (2020) propose a novel frequency control, that allows the frequency setpoint to change adaptively throughout the evolution of the microgrid's islanded operation.

Conclusions

This paper presents a comprehensive literature review on studies about PQ disturbances in islanded microgrids. An increasing interest in this topic has been identified, especially regarding interconnected microgrids. Although the number of studies on this topic grows, there are still no systematic studies that analyze the impact of the different microgrid elements on PQ levels.

In most of the reviewed papers, the PQ of microgrids becomes worse during islanded operation, especially voltage distortion (harmonics and supraharmonics), voltage unbalance, and the frequency variations; these seem to have the highest difference between the interconnected and islanded modes. However, there is no systematic study to generalize the results obtained from the papers, as some contradict others.

The increase or decrease of the power quality indexes is linked to the type of generators, type of loads, control schemes, and the presence of ESS, filters, and compensation devices. In most cases, generation units are electronic-based, and the loads are non-linear. It seems that this generation-load scenario results in high voltage and current disturbances. However, not all reviewed cases specify the exact type of generator nor the power and the types of the supplied loads, so it cannot be determined which generation-load combination produces the highest distortions.

Changes in microgrid operation mode, such as from interconnected to islanded, cause variations in network impedance. None of the reviewed references thoroughly analyze the variation in network impedance, their frequency dependent characteristics. Consequently, typical values for microgrids are still unknown. Network impedance analysis is indispensable to understand microgrid behavior and the interaction between devices, as well as to propose limits for power quality disturbances and EMC coordination.

Most of the reviewed papers based their analysis on simulations or small laboratory-based microgrids. There is a lack of studies based on measurements of real islanded microgrids.

Finally, current international standards have yet established a comprehensive framework for EMC coordination in islanded microgrids. However, defining the corresponding limits is an paramount to ensure the reliable operation of electrical devices in islanded microgrids without loss of performance or damage.

Acknowledgements

This project is supported by DAAD with funds from the German Federal Ministry of Education and Research (Project ID: 57447921), as well as by COLCIENCIAS with funds from the National Fund for Science, Technology, and Innovation Francisco José de Caldas (Project ID: 64996).

References

Agundis-Tinajero, G., Segundo-Ramírez, J., Peña-Gallardo, R., Visairo-Cruz, N., Núñez-Gutiérrez, C., Guerrero, J. M., and Savaghebi, M. (2018). Harmonic issues assessment on PWM VSC- based controlled microgrids using Newton methods. *IEEE Transactions on Smart Grid, 9*(2), 1002-1011. 10.1109/TSG.2016

Algaddafi, A., Brown, N., Gammon, R., Altuwayjiri, S. A., and Alghamdi, M. (2016). Improving off-grid PV system power quality, and comparing with grid power quality. Paper presented at the International Conference on Electronics, Information, and Communications, ICEIC 2016, (Dc). 10.1109/ELINFOCOM.2016.7562928

Blanco, C., Tardelli, F., Reigosa, D., Zanchetta, P., and Briz, F. (2019). Design of a cooperative voltage harmonic compensation strategy for islanded microgrids combining virtual admittance and repetitive controller. *IEEE Transactions on Industry Applications, 55*(1), 680-688. 10.1109/TIA.2018.2868691

Bollen, M., Zhong, J., Zavoda, F., Meyer, J., McEachern, A., and Lopez, F. C. (2017a). Power Quality aspects of Smart Grids. *Renewable Energy and Power Quality Journal, 1*(8), 1061-1066. 10.24084/repqj08.583

Bollen, M. H. J., Das, R., Djokic, S., Ciufo, P., Meyer, J., Ronnberg, S. K., and Zavodam, F. (2017b). Power Quality Concerns in Implementing Smart Distribution-Grid Applications. *IEEE Transaction on Smart Grid, 8*(1), 391-399. 10.1109/TSG.2016.2596788

CIGRE Working Group JWG C4/C6.29. (2016). Power Quality Aspects of Solar Power. (Technical Brochures; Vol. 672). CIGRE. ISBN: 978-2-85873-375-0 https://e-cigre.org/publication/672-power-quality-aspects-of-solar-power

CIGRE Working group JWG C4.24/CIRED. (2018). Power quality and EMC issues with future electricity networks. In Technical Brochures; Vol. 719). CIGRE. ISBN: 978-2-85873-421-4 http://cired.net/uploads/default/files/final-report-C4.24-CIRED.pdf

Dehghani, M. T., Vahedi, A., Savaghebi, M., and Guerrero, J. M. (2012). Voltage quality improvement in islanded microgrids supplying nonlinear loads. 2012 3rd Power Electronics and Drive Systems Technology (PEDSTC), 3, 360-365. 10.1109/PEDSTC.2012.6183356

Garde, R., Casado, S., Santamaria, M., and Aguado, M. (2015). Power quality and stability analysis during islanded mode operation in a microgrid based on master-slave configuration. In IEEE (Eds.) 2015 Saudi Arabia Smart Grid (SASG) (pp. 1-8). New York, NY: IEEE. 10.1109/SASG.2015.7449288

Golsorkhi, M. S., Lu, D., Savaghebi, M., Vasquez, J. C., and Guerrero, J. M. (2016). A GPS-based control method for load sharing and power quality improvement in microgrids. In IEEE (Eds.) 2016 IEEE 8th International Power Electronics and Motion Control Conference (IPEMC-ECCE Asia)(pp.3734-3740)New York, NY. IEEE. 10.1109/IPEMC.2016.7512893

Hicks, C., Baghzouz, Y., and Haddad, S. (2018). Power quality of residential PV system under low solar irradiance and off-grid operation. 2018 18th International Conference on Harmonics and Quality of Power (ICHQP), 4026, 1-5. 10.1109/ICHQP.2018.8378937

IEEE. (2014). IEEE Std 519-2014 (Revision of IEEE Std 519-1992) Recommended Practice and Requirements for Harmonic Control in Electric Power Systems. New York, NY: IEEE. 10.1109/IEEESTD.2014.6826459

Kaushal, J. and Basak, P. (2020). Power quality control based on voltage sag/swell, unbalancing, frequency, THD and power

factor using artificial neural network in PV integrated AC microgrid. Sustainable Energy, Grids and Networks, 23, 100365. 10.1016/j.segan.2020.100365

Kulia, G., Molinas, M., Lundheim, L. M., and Fosso, O. B. (2017). Simple model for understanding harmonics propagation in single-phase microgrids. In IEEE (Eds.) 6th International Conference on Clean Electrical Power (ICCEP), pp. 354-358. New York, NY: IEEE. 10.1109/ICCEP.2017.8004839

Kumar, K. and Venkateshwarlu, S. (2013). A Review on Power Quality in Grid Connected Renewable Energy System. *CVR Journal of Science and Technology, 5, 6*. 10.32377/cvrjst0510

Laaksonen, H. and Kauhaniemi, K. (2008). Voltage and current THD in microgrid with different DG unit and load configurations. CIRED Seminar 2008: SmartGrids for Distribution, (0057), 71-71. 10.1049/ic:20080476

Lhachimi, H., Sayouti, Y., and Kouari, Y. El. (2020). Control of a flexible microgrid during both modes of operations with presence of nonlinear loads. *Journal of the Franklin Institute, 357*(11), 6498-6538. 10.1016/j.jfranklin.2020.03.046

López-García, D., Arango-Manrique, A., and Carvajal-Quintero, S. X. (2018). Integration of distributed energy resources in isolated microgrids: the Colombian paradigm. *Revista TecnoLógicas, 21*(42), 13-30. 10.22430/22565337.774

Marnay, C., Chatzivasileiadis, S., Abbey, C., Iravani, R., Joos, G., Lombardi, P., Mancarella, P., and Appen, J. V. (2015). Microgrid Evolution Roadmap. In IEEE (Eds.) International Symposium on Smart Electric Distribution Systems and Technologies (EDST) (pp. 139-144). New York, NY: IEEE. 10.1109/SEDST.2015.7315197

Micallef, A., Apap, M., Spiteri-Staines, C., and Guerrero, J. M. (2012). Cooperative control with virtual selective harmonic capacitance for harmonic voltage compensation in islanded microgrids. IECON 2012 - 38th Annual Conference on IEEE Industrial Electronics Society, 3, 5619-5624. 10.1109/IECON.2012.6389037

Nomm, J., Ronnberg, S., and Bollen, M. (2018a). Harmonic voltage measurements in a single house microgrid. 18th International Conference on Harmonics and Quality of Power (ICHQP), 2018(May), 1-5. 10.1109/ICHQP.2018.8378921

Nomm, J., Ronnberg, S., and Bollen, M. (2018b). An analysis of frequency variations and its implications on connected equipment for a nanogrid during Islanded operation. *Energies, 11*(9), 1-13. 10.3390/en11092456

Nomm, J., Ronnberg, S., and Bollen, M. (2019). An Analysis of Voltage Quality in a Nanogrid during Islanded Operation. *Energies, 12*(4), 614. 10.3390/en12040614

Padayattil, G. M., Thobias, T., Thomas, M., Sebastian, J., and Pathirikkat, G. (2016). Harmonic analysis of microgrid operation in islanded mode with nonlinear loads. In IEEE (Eds.) 2016 International Conference on Computer Communication and Informatics, ICCCI 2016 (pp. 1-5). New York: IEEE. 10.1109/ICCCI.2016.7480031

Phan, D. M. and Lee, H. H. (2019). Interlinking Converter to Improve Power Quality in Hybrid AC-DC Microgrids with Nonlinear Loads. *IEEE Journal of Emerging and Selected Topics in Power Electronics, 7*(3), 1959-1968. 10.1109/JESTPE.2018.2870741

Prabaakaran, K., Chitra, N., and Kumar, A. S. (2013). Power quality enhancement in microgrid - A survey. In IEEE (Eds.) Proceedings of IEEE International Conference on Circuit, Power and Computing Technologies ICCPCT 2013 (pp. 126-131). New York, NY: IEEE. 10.1109/ICCPCT.2013.6528830

Rodrigues, Y., Monteiro, M., Abdelaziz, M., Wang, L., de Souza, A. Z., and Ribeiro, P. (2020). Improving the autonomy of islanded microgrids through frequency regulation. *International Journal of Electrical Power & Energy Systems, 115*(May 2019), 105499. 10.1016/j.ijepes.2019.105499

Ronnberg, S., Bollen, M., and Nomm, J. (2017). Power quality measurements in a single-house microgrid. *CIRED - Open Access Proceedings Journal, 2017*(1), 818-822. 10.1049/oap-cired.2017.0240

Savaghebi, M., Shafiee, Q., Vasquez, J. C., and Guerrero, J. M. (2015). Adaptive virtual impedance scheme for selective compensation of voltage unbalance and harmonics in microgrids. *IEEE Power & Energy Society General Meeting, 2015*(Sept), 1-5. 10.1109/PESGM.2015.7286092

SEK Svensk Elstandard. (2011). SS-EN 50160, Voltage characteristics of electricity supplied by public distribution systems. https://elstandard.se/standard/1002101

Sood, V. K. and Abdelgawad, H. (2019). Chapter 1 - Microgrids architectures. In Chauhan, K. and Chauhan, Kalpana (Eds.) Distributed Energy Resources in Microgrids (pp. 1-31). Amsterdam, Netherlands: Elsevier. 10.1016/B978-0-12-817774-7.00001-6

Thai, T. T., Raisz, D., Monti, A., Ponci, F., and Ahmadifar, A. (2018). Voltage harmonic reduction using virtual oscillator-based inverters in islanded microgrids. In IEEE (Eds.) Proceedings of International Conference on Harmonics and Quality of Power, (ICHQP, 2018), (pp. 1-6). New York, NY: IEEE. 10.1109/ICHQP.2018.8378896

Tenti, P., Paredes, H. K. M., and Mattavelli, P. (2011). Conservative power theory, a framework to approach control and accountability issues in smart microgrids. *IEEE Transactions on Power Electronics, 26*(3), 664-673. 10.1109/TPEL.2010.2093153

Ley 1715. Por medio de la cual se regula la integración de las energías renovables no convencionales al sistema energético nacional (D.O. No. 49.150). 13 de mayo de 2014.

Yang, J., Chen, Y., Dong, M., and Han, H. (2015). A harmonic and reactive power sharing control strategy for islanded microgrids. In IEEE (Eds.) Proceedings - 2015 Chinese Automation Congress (CAC), (pp. 1422-1427). New York, NY: IEEE. 10.1109/CAC.2015.7382722

Zeng, Z., Yi, H., Zhai, H., Zhuo, F., and Wang, Z. (2017). Harmonic power sharing and PCC voltage harmonics compensation in islanded microgrids by adopting virtual

harmonic impedance method. In IEEE (Eds.) Proceedings IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society (pp. 263-267). New York, NY: IEEE. 10.1109/IECON.2017.8216048

Zhou, N., Yuan, C., and Wang, Q. (2012). Control strategies for microgrid power quality enhancement with back-to-back converters connected to a distribution network. In IEEE (Eds.) 2012 IEEE 15th International Conference on Harmonics and Quality of Power (pp. 384-389). New York, NY: IEEE. 10.1109/ICHQP.2012.6381166

Research Article / Electrical-Electronic Engineering http://dx.doi.org/10.15446/ing.investig.v40n3.80124

Control of a Direct Current Motor Using Time Scaling

Control de un Motor de Corriente Continua usando Escalamiento Temporal

José Danilo Rairán Antolines[1]

ABSTRACT

Humans naturally control their surrounding space. However, that capacity has not been fully used to build better intelligent controllers, mainly because the reaction time of a person limits the number of industrial applications. In this paper, the author proposes a method to overcome the problem of reaction time for a human in the control loop. This method, called Time Scaling Control, starts by modifying the constant times of the plant's model to the point where control is comfortable for a human. Then, the controller acquires the knowledge that was expressed during the human control stage and places it in a Neural Network, which controls both scaled and original plants. Time Scaling Control highly improves the control performance compared with a PID, in this case demonstrated by the control of a direct current motor, which cannot be controlled by a human without time scaling control due to the speed of the system.

Keywords: DC motor, time scaling control, plant identification, neural network, intelligent control

RESUMEN

Los humanos controlan el espacio que los rodea de manera natural. Sin embargo, esta capacidad no se ha usado completamente para construir mejores controladores inteligentes, principalmente porque el tiempo de reacción de una persona limita el número de posibles aplicaciones industriales. En este artículo se propone un método para eliminar el problema del tiempo de reacción de un humano en un lazo de control. Este método, llamado Control con Escalamiento Temporal, comienza por modificar las constantes de tiempo del modelo de la planta, hasta el punto en el que el control sea cómodo para un humano. Entonces, el controlador adquiere el conocimiento que fue expresado durante la etapa de control humano y lo ubica en una red neuronal, la cual controla tanto la planta escalizada como la planta original. El Control con Escalamiento Temporal mejora bastante el desempeño del control en comparación con un PID, demostrado en este caso por el control de un motor de corriente directa, el cual no puede ser controlado por una persona sin el uso de escalamiento temporal por la velocidad del motor.

Palabras clave: motor DC, control con escalamiento temporal, identificación de plantas, redes neuronales, control inteligente

Received: June 06th, 2019
Accepted: September 3rd, 2020

Introduction

Traditional control strategies look for a control law that properly satisfies certain conditions, such as stability, controllability and observability in a control system. These strategies can be more or less complex, according to the restrictions and requirements of the problem. However, most strategies are purely mathematical and ignore the fact that the human brain is a highly effective adaptive controller (Van der El, Pool, Van Paassen, and Mulder, 2018). Only recently have scientists begun to measure human control performance (Huang, Chen, and Li, 2015; Laurense, Pool, Damveld, Paassen, and Mulder, 2015; Inga, Flad, and Hohmann, 2017). These studies show that one of the main limitations to having a human in a control loop is time delay. Brain and body take time to provide an actuating signal and that delay could be too long for some applications.

An area of science that studies the human brain as a controller is neuroscience. Through this discipline, it is possible to explain the abilities of humans in motion planning and decision making (Mackie, Van Dam, and Fan, 2013). The majority of studies focus on the prefrontal cortex and other structures like the amygdala (Duverne and Koehlin, 2017). However, other studies show that there is a coordination among multiple brain areas, especially when the brain has last-minute decisions (Xu et al., 2017), as it happens during the control of dynamic systems. The brain is so prone to control that a new area called network control theory explains some aspects of the brain that may help in the treatment of neurological diseases (Medaglia, Pasqualetti, Hamilton, Thompson, and Bassett, 2017).

Regardless of the successes or failures of different areas of science while proving the human capacity to control, people control their environment and transform it every day. In addition, it is interesting to see that, when a human influences

[1]Electrical Engineer, Universidad Nacional de Colombia, Colombia. Ph.D. in Engineering – Systems and Computation, Universidad Nacional de Colombia, Colombia. Affiliation: Professsor at Universidad Distrital Francisco José de Caldas, Colombia. E-mail: drairan@udistrital.edu.co

How to cite: Rairán, D. (2020). Control of a Direct Current Motor Using Time Scaling. *Ingeniería e Investigación*, *40*(3), 38-46. 10.15446/ing.investig.v40n3.80124

the environment it also affects the human in return. Thus, a human controlling also becomes the object of control, as can be seen in Corno, Giani, Tanelli, and Savaresi (2015). Humans have also served as inspiration for several control strategies, such as teaching robots how to control articulation to achieve natural motion (Lee, 2015), or teaching autonomous vehicles to make better decisions (Suresh and Manivannan, 2016). A broad area that has generated several proposals is human-robot collaboration, because a robot working with a human should predict human behavior, in other words, it should emulate the control actions of a human (Robla et al., 2017).

This proposal, as well as other techniques in intelligent control, aims to build better controllers by the emulation of a human during a control task. However, and first of all, the author proposes a method to eliminate the limitations caused by the time delay from the brain and body. That delay makes some systems too fast to control, whereas systems that are too slow may cause attention problems. In both cases, there may be a reduction in control performance, which limits the number of dynamic systems that a human can control. Instead, and using the proposal in this paper, a person can control any dynamic system by properly changing its speed.

The intuitive idea of the proposal, called Time Scaling Control (TSC), consists of changing the constant times of a plant until its control is comfortable enough for humans. In general, this is impossible to do with a real plant, but it is possible through the use of this model. Thus, a human controls a scaled version of the model instead of the real plant. Finally, the knowledge that was captured during the control task is learned by a neural network. An important feature of some neural network architectures, such as a Multilayer Perceptron, is that they are blind to changes in the time scale (as will be explained later), so they can be equally used for the scaled model and in the real plant.

There are three main advantages of Time Scaling Control compared with any other intelligent controller:

1. TSC captures the ability of the human brain to control dynamic systems, even when that knowledge remains hidden for the human controller, since most of the motor control activity happens unconsciously. In contrast, for instance, in Fuzzy Control, the so-called expert must be able to verbalize the knowledge to achieve control, but that is not always possible.
2. TSC allows a person to control any system, fast or slow. Current uses of the computational power of the human brain are limited to systems that evolve at a pace that matches human possibilities. Those limits disappear when using the model.
3. TSC not only allows the human to control any system, even when the person can not consciously describe the control rule, but also provides a method to transfer that knowledge into an automatic algorithm, which makes TSC ready to use in industrial applications.

The implementation of TSC requires the following three steps: *Scaling*, *Training*, *Running*:

1. *Scaling* starts by defining a model of the plant, which is then scaled, as defined in Equation 1. The application of this Equation is the only math operation during the whole control process, which is part of the beauty of this proposal.
2. During *Training*, the person controls the scaled model of the plant, and those control actions are recorded as data to train a neural network.
3. During the *Running* step, the neural network with encapsulated knowledge controls not only the scaled version of the model, but the plant itself. The three main sections of this paper (Scaling, Training, Running) detail those three steps to implement a Time Scaling Controller.

The author uses an example to show the application of the proposed control strategy: the control of the angular position of a shaft. This problem required the construction of a position sensor, as explained in the next Section. The following Section presents a central procedure in the controller design; it is the identification of the plant. The next three sections thoroughly show the step by step of the design for the proposed controller: the *Scaling* (S), the *Training* (T), and the *Running* of the controller (R). This paper ends with a discussion section, conclusions, and future work.

Angular Position Measurement

The basic configuration of a system to control the angular position of a motor requires, in addition to the controller, three components: a driver, the motor, and a sensor. The driver uses two operational LM386 amplifiers set as a bridge to allow the motor to rotate clockwise or counter clockwise. The actuating signal coming from the controller feeds the amplifier card, so that, in the steady state, the motor rotates at a speed proportional to the actuating signal.

There is a huge variety of sensors used to measure the angular position of a motor. The usual selection is an incremental encoder, but other popular choices are the absolute optical encoder and the resolver. Applications with low precision requirements, as well as small angle variations, can be measured using potentiometers. Given the high cost of an absolute encoder or a resolver, the low precision of a potentiometer, and the undesirable dependency on the initial conditions of an incremental encoder, this paper presents the design of a sensor based on the Hall effect. This sensor aims to be absolute, as well as cheap and precise.

The sensor has two components: one rotates while the other remains still. The static part has two Hall effect sensors in quadrature, plus the electronic to normalize the signal. The rotating part uses two magnets in the shaft, which produces sine and cosine type output signals in the sensors, as described in (Rapos, Mechefske, and Timusk, 2016), which is the configuration used in this paper. However, there are other interesting proposals to explore in future applications. For instance, Wu and Wang (2016) use three or six hall sensors instead of two to improve accuracy at the expense

of increasing the complexity of signal processing. On the other hand, Anoop and George (2013) use two hall sensors, but propose the use of linear outputs instead of sinusoidal outputs, which may boost measurement accuracy.

The signal conditioning component shown in Figure 1 starts by filtering the noise that may appear due to mechanical vibrations or electrical interferences. The implementation of this filtering uses a traditional low-pass filter of the first order $out/in = 1/(s\tau + 1)$, which can be written using a discrete set of blocks in Simulink, following the rule $out = (1/s)(1/\tau)(in - out)$. This configuration allows the definition of an initial output for the filter as the initial condition of the Integrator $(1/s)$. That initial condition equals the initial measurement in each sensor, whereas the subtraction $in - out$ corresponds to a negative feedback connection, which passes through the constant $1/\tau$ in order to define the cut frequency of the filter ($\tau = 2$ ms). On the other hand, the constants in the signal conditioning section in Figure 1 serve to set the average signal to zero and the amplitude to one. These values were experimentally set at $c_1 = 1{,}094$, $c_2 = 1{,}0144$, $c_3 = 0{,}277$, $c_4 = 0{,}664$. This normalization facilitates the computation of the angle made by means of a $atan2$ function. This function uses x and y to come up with and angle between $\pm\pi$, instead of the traditional $\pm\pi/2$ with the function $atan$.

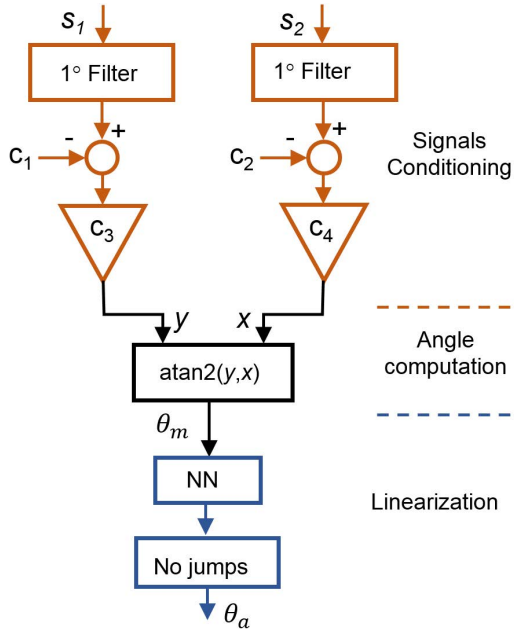


Figure 1. Block diagram of the angle estimation.
Source: Authors

Any mechanical misalignment in the measurement configuration or between the sensor gains causes a discrepancy between the actual angle, θ, and the output of the function $atan2$, also called θ_m. This difference was as high as 10%, so the author used a neural network (NN) to reduce that error. The basic idea to reduce the error is to train a NN to learn the inverse relationship between θ and θ_m. Thus, when a mechanical position θ produces an estimation equal to θ_m, that estimation ideally produces an output equal to θ in the NN. The author used a feed-forward backpropagation network with 8 neurons in the input layer and 4 neurons in the hidden layer. The input data θ cover angles between $-180°$ and $180°$, with data points every $1°$. Figure 2 shows the error in the estimation using the NN to linearize the angle measurement. The error always remains under $0{,}3°$.

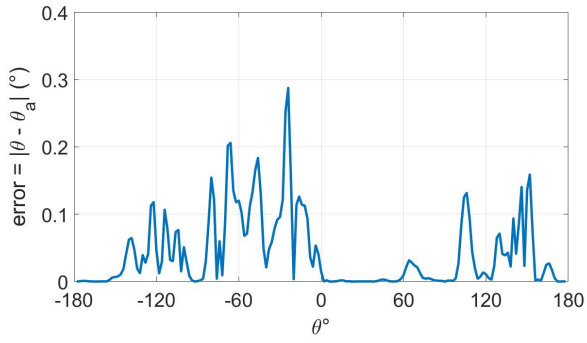


Figure 2. Error in the angle estimation.
Source: Authors

The estimation of the angle covers a single turn from -180° to 180°. Going further means generating discontinuities in the approximation due to the behavior of the function $atan2$. The final step in angle approximation implies avoiding those discontinuities, as depicted in Figure 1 by the block "No jumps". The author starts by using two new variables to detect the discontinuities: cd for jumps from $180°$ to $-180°$, and cu in the other case. Those constants start at zero and increase by one every time the difference between the current and the previous angle goes under $-300°$ for cd, or over $300°$ for cu. The value $\pm 300°$ was defined experimentally under the consideration that it should be higher than the delta angle produced at the maximum speed between two consecutive angle samples. Thus, the final approximation can be written as $\theta_a \leftarrow 360(cd - cu) + \theta_a$.

System Identification

The plant in this application is a permanent magnet direct current motor (PMDC) with a nominal current of 500 mA and nominal voltage of 12 V. This motor is widely used in robotics and it is the most popular option to test control algorithms. However, its parameters are unknown to the author. We could apply some tests to approximate the parameters of the motor, but we preferred to use the approach in this section. Firstly, because those parameters assume a linear system, when they do not; the motor can be seen as linear only around the nominal values of current, voltage, speed, and temperature. Secondly, the process used in this paper requires a single set of data to capture a model that resembles the whole dynamic instead of running several tests to obtain a model.

A traditional DC motor model uses linear differential equations to represent the relationship between the states in the motor. In this linear model, the input voltage produces an electrical current which generates a mechanical torque

resulting in the rotation of the motor. Identification, in addition to the equations, requires the computation of electrical, electromechanical, and mechanical constants, such as resistance, inductance, the constant of friction, and the constant of inertia, as is presented in Wahyunggoro and Saad (2010). This type of model is used to study the motor as a dynamic system. It can also be part of a first attempt to design a controller. However the approximation given by a set of linear equations cannot fully emulate the richness of the dynamic in a real machine.

Instead of a pure linear model, Aquino and Velez (2006) propose a gray-box model. In this model, the electrical part is linear and known, whereas the mechanical part is nonlinear and unknown, as a result of the effect of changes in the load. The author proposes the use of Radial Basis Neural Networks to capture the nonlinear behavior of the model, and compare the results using a recursive two-stage method. Another approach in the identification assumes that the whole model is unknown. Thus, the linear part of the model can be represented by an ARX model, whereas the nonlinear part is chosen to be a polynomial of known order, as presented in Kara and Eker (2004). These authors study the effect of nonlinear friction and other nonlinearities in the model under the assumption that it is possible to separate the nonlinearities in two groups: static and dynamic. Another proposal by Rios and Makableh (2011) uses a Feed Forward Neural Network to capture the nonlinearities of the motor. That model is then used to simulate the effect of another neural network working as a controller of the motor. This last proposal focuses on two of the main nonlinearities studied in this paper: dead zone and saturation.

The identification of the motor in this paper uses some characteristics of the proposals in the references above, but instead of looking for a very accurate model, this paper shows that a model that approximately captures the main features of the dynamic is enough to get good control performances using time scaling. The core of the model presented in Figure 3 emulates the transient behavior of the motor using two linear blocks: one for the relation between the input voltage and the speed in the shaft (a second order transfer function, $H(s)$), and the other for the relation between speed and angular position (an Integrator, $1/s$). Given the author' experience controlling the motor, there are two main nonlinearities that should be included in its model: 1) the dead zone, and 2) the gain of the steady-state as a function of the rotation sense. The dead zone is represented by the first block in Figure 3, while the gains correspond to the triangles. The model includes another nonlinearity, the saturation of the Integrator, in other words, the natural limits for the rotation of the motor. The block between the gains and the Integrator symbolizes the selection of one gain or the other depending on the sense of rotation. If the speed is greater than or equal to zero, then the gain for the model is k_{ccw}; if the speed is negative, the gain is k_{cw}.

This is the list of parameters for the model, according to the model in Figure 3:

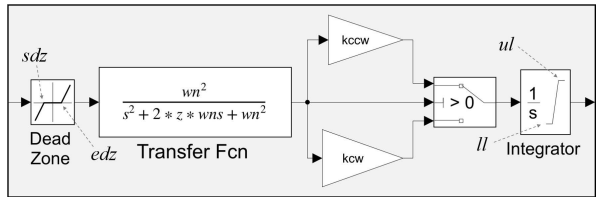


Figure 3. Motor model scheme.
Source: Authors

- Inside the block "Dead Zone", the start of dead zone, sdz, and the end of the dead zone, edz.
- Transient, in the block "Transfer Fcn": natural frequency, ω_n, damping ratio, $z = \zeta$.
- Gains: counter clockwise sense, k_{ccw}, clockwise sense, k_{cw}.
- Saturation, inside the block "Integrator", the upper limit, ul, and the lower limit, ll.

The computation of these parameters starts by defining the last two: ul and ll. Even when the control rank equals ±90°, the limits are greater than that rank. A first extra 10° allows the human controller to make mistakes at the ends of the rank and still generate good data to train the controller. A second extra 10° bounds the parameter estimation algorithm to vary the parameters of the model. Thus, the control rank is ±90°, but the human control covers ±100°, while the estimation algorithm has a window of ±110°. Thus, ul = 110° and ll = −110°. The next two parameters correspond to the transfer function $H(s)$ in the block "Transfer Fcn". These two parameters define the shape of the transient when the input of the model changes. A way to evaluate these values implies generating sudden changes in the input of the motor and then measuring the speed of the shaft. The values that match those transients better are rounded to ω_n = 100 rad/s, and ζ = 1.

The four remaining parameters of the model were computed using the tool Parameter Estimation of Simulink. Given that the main goal of the model is to emulate the system when the controller is working, then a proportional controller with k_p = 1/10 leads the motor during the acquisition of the input and output data. The whole experiment lasts a minute and uses samples every millisecond, so there are 60 001 data points for the input and the same number for the output. The reference for the control system is a step function randomly changing from 0,3 to 0,7 s with also random amplitude from −100° to 100°. This signal was smoothed using a first order filter with constant time of 0,1 s, which better emulates the work of the motor during normal conditions. The optimization method is the Nonlinear Least Squares and the algorithm is Levenberg-Marquardt. The tolerance in the optimization is 0,001. A final estimation was run, including all the parameters except the saturation constants. The natural frequency changes to ω_n = 300 rad/s. It is important to report that the change from 100 to 300 in ω_n didn't alter the quality of the approximation in more than 0,1%, which

means that the nonlinearities affect the dynamic of the motor more than the linear part. The results of the optimization are shown in Figure 4, where the lower part corresponds to the input voltage, while the upper part is the output of the plant and the model.

A traditional modelling process for a DC motor defines the model as a linear transfer function, and that approximation may be useful when working around the nominal conditions of the machine. However, the DC motor for this application changes the nominal speed depending on the rotation sense, as expressed by means of the gains k_{cw} and k_{ccw}. These gains can be evaluated as the quotient between the speed in °/s at nominal voltage in V, for both positive and negative cases of voltage, but, in this case, Simulink made the approximation based on the data in Figure 4. The result is that $k_{cw} = 940°/s/V$ whereas $k_{ccw} = 530°/s/V$. Thus, the motor is almost two times faster in a clockwise sense than in the counter clockwise sense.

The last two constants of the model express the fact that the voltage in the motor requires to pass a certain threshold before the shaft starts to rotate, mainly given the friction effect. An initial estimation of those constants can be made by increasing the voltage to the point that the shaft starts to rotate, but, once more, that approximation was made by Simulink based on the data in Figure 4. The rotation sense changes the threshold, so the model requires the estimation of that value in both senses, as defined by using the gains sdz and edz inside the block "Dead Zone". The values for those gains are $sdz = -0{,}9$ V and $edz = 0{,}6$ V.

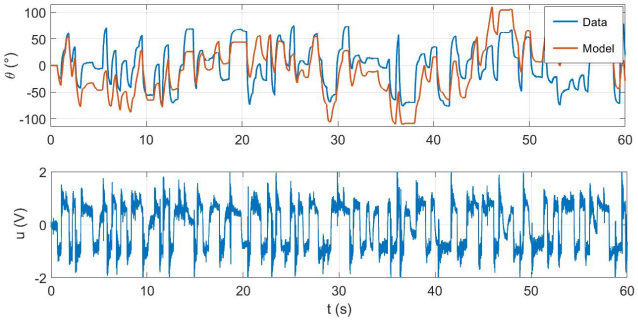


Figure 4. Output data and simulated output under the same input.
Source: Authors

In summary, the model in Figure 3 has dead zone $sdz = -0{,}9$, $edz = 0{,}6$; transient $\omega_n = 300$, $\zeta = 1$; gains $k_{cw} = 940$, $k_{ccw} = 530$; and saturation: $ul = 110°$, $ll = -110°$.

Scaling Stage (S)

The model in Figure 3 uses the traditional definition of time, but that dynamic is too fast for a human to control. Thus, this section starts by scaling the linear components of the motor, as defined in Equation (1), according to the presentation in Rairán (2017). It can be seen that the nonlinear components are not time-dependent, but constant, so the nonlinear part remains equal regardless of the time scale. The constant k_t in Equation (1) corresponds to the scaling factor. For instance, $k_t = 5$ means that the new dynamic H' is five times faster than the dynamic H. The new plant, H', keeps amplitude and shape of H and only changes how long a transient lasts.

$$H = \frac{a_n S^n + a_{n-1} S^{n-1} + \cdots + a_1 S + a_0}{b_n S^n + b_{n-1} S + \cdots + b_1 S + b_0}$$

$$H' = \frac{a_n S^n + a_{n-1} S^{n-1} k_t + \cdots + a_1 S k_t^{n-1} + a_0 k_t^n}{b_n S^n + b_{n-1} S k_t + \cdots + b_1 S k_t^{n-1} + b_0 k_t^n} \quad (1)$$

Defining the scaling factor implies an experimental procedure where the human controlling the system changes that factor until achieving a value that makes control of the system comfortable. In this case, this value is set to $k_t = 1/30$, which means that the scaled system is 30 times slower than the original. In this way, for instance, a transient of two seconds in the motor lasts a whole minute in the scaled model. The transient $H = 300^2/(s^2 + 600s + 300^2))$ corresponds to $H' = 100/(s^2 + 20s + 100)$, and the integral $\theta(s)/Vel(s) = 1/s$ corresponds to $\theta'(s)/Vel'(s) = (1/30)/s$. Both input and output in each transfer function need to be scaled as indicated by the notation. The slow human reactions produce a slow velocity, which in turn produces a slow change in the angular position. The scaling additionally requires the scaling of the sampling time at which the data is recorded or generated, from t_s to $t'_s = t_s/k_t$ in this case, $t_s = 1$ ms, so $t'_s = 30$ ms.

This stage in the control process requires interaction between the human and the scaled system. The human senses and evaluates the behavior of the plant, and, with experience in the manipulation of the system, compares his or her perceptions with the set point. As a result, the human makes two decisions: 1) whether to increase, decrease or maintain the actuating signal, and 2) how significant this change will be. The application of the actuating signal as input for the scaled system produces changes in the output of the scaled plant in such a way that, ideally, the perceived output matches the set point in the shortest time possible. The instantiation of the control stage in this paper uses a computer and a human, as shown in Figure 5.

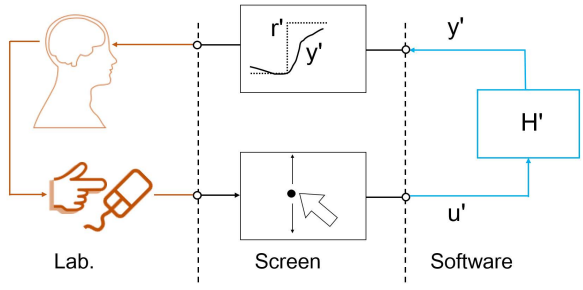


Figure 5. Control Scheme.
Source: Authors

This stage aims to generate the best data to train a NN in the next stage. Thus, the definition of the reference signal should make the system as dynamic as possible. In this

case, the author defines a square signal oscillating between 100° and −100°, lasting between 12 and 15 seconds in each value. That signal passes through a low pass filter of the first order with $\tau = 2$ s, as shown in Figure 6. This filter produces references along the rank of ±100°, instead of the extreme values alone. The constant time τ aims to emulate the normal work of the motor during a traditional control application, and at the same time its stabilization time falls just below the transition time, which allows the system to reach its steady state. The data acquisition lasts 2,5 minutes. Thus, every reference value appears about five times, which is enough data to train a NN.

Figure 6 also shows that the information presented to the human controller through the scope is not the scaled error, e' (as it is done in traditional control), but its negative, $-e'$. It can be seen that, regardless of the value of the output y', the output will eventually match the sign of the input u'. For instance, if u' increases, then y' (as well as $-e' = y' - r'$) will eventually also increase. Thus, if $-e'$ is, for instance, negative, a good control action will consist of setting u' as a positive value. This action is intuitive for a human: it is performed to counteract the signal in the screen in order to have null error. On the other hand, the block "Mouse Pointer Reading" in Figure 6 runs a function based on the function "gpos" of Matlab, which provides the coordinates of the pointer of the mouse. The y value of the pointer corresponds to the value of u'.

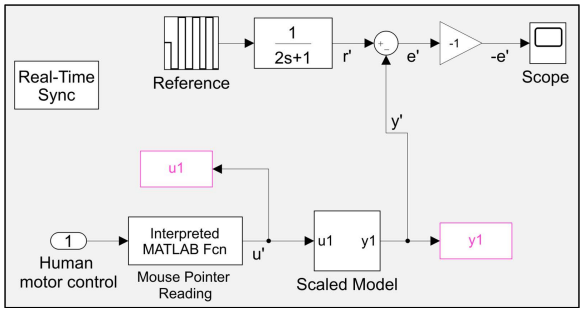


Figure 6. Control system in Simulink.
Source: Authors

Blocks $u1$ and $y1$ in Figure 6 save the data generated during the running of the simulation and make it available for use in the workspace of Matlab. The signal u' corresponds to the actions of the human controller, while y' provides the corresponding emulated position given by the scaled model. Finally, the block "Real-Time Sync" allows the running of the simulation in real time using the Simulink Desktop Real-Time toolbox.

Training Stage (T)

This stage uses the data coming from the previous stage to train a NN. Thus, instead of the human controlling the motor, the system uses the NN to define the actuating signal. The training stage starts by normalizing the data, which implies cutting and scaling. By cutting, the author understands triming and deleting the first 15 seconds of the experiment, given that this is the time required for a person to accommodate, concentrate, and be focused on the control problem. The scaling consists in dividing the input e by 180, as shown in Figure 7. This scaling makes the rank of the input [−1.1, 1.1], which is close to the rank of the activation function of each neuron [−1, 1]. Nevertheless, the effectiveness of the training remains equal under the scaling or without it in this application. The scaling makes a difference when looking at multiple inputs, as could be the case of future work, when the problem requires (for instance, in addition to e', its derivative). Normalization makes all the inputs have a similar influence during the training.

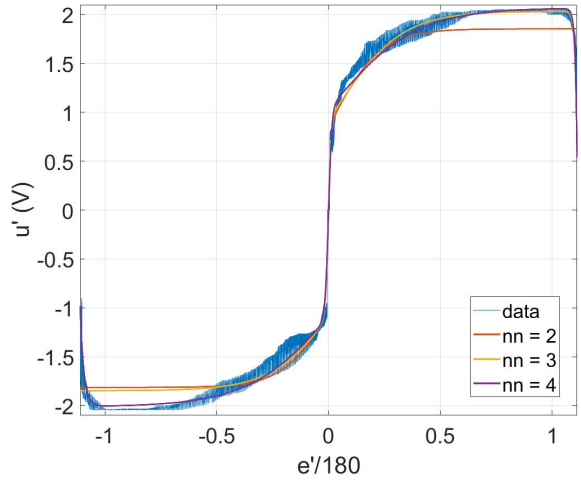


Figure 7. Effect of the number of neurons in the training.
Source: Authors

An analysis of the sorted data in Figure 7 shows the relationship between the error and the actuating signal. That relation looks like a tangent sigmoid function. This observation favors the use of a Multilayer Perceptron Neural Network to learn the given data, so that each of its neurons may have a tangent sigmoid transfer function. It can be seen that sorting error evinces the control strategy of the human, which was hidden inside the brain of the human controller until now. It is possible to divide that strategy in three zones: 1) small errors, from 0 to 0,2; 2) medium errors, from 0,2 to 0,8; 3) major errors, from 0,8 to 1,1. Zone 1 presents the highest rate of change and happens where the majority of control takes place. Zone 2 has a smaller slope and resembles the saturation of a tangent sigmoid. Zone 3, contrary to the intuition, decreases the actuating signal. This last behavior is caused by the reaction time of a human due to large errors (positive or negative). Unlike a pure mathematical algorithm, the control interface (including the brain itself) does not allow the human to change a decision instantly. However, a positive aspect of that delay is that it reduces the overshoot in the response.

The author used the Levenberg-Marquardt algorithm to train the NN, with a maximum of 1 000 epochs, and data randomly divided into three sets: 70% to train, 15% to validate the training, and 15% to test the NN. Figure 7 shows the result of changing one of the most influential parameters during the learning process: the number of neurons in the network.

Results in Figure 7 regard networks with a single hidden layer with two, three, or four neurons. Two neurons miss the third zone. Three neurons resulted better than two, but the generalization for major and negative errors fails. The best result in Figure 7 uses four neurons. That network properly matches and generalizes the data across the three zones.

Figure 8 shows the results when training 100 neural networks with four neurons in the hidden layer. The lower part of the figure shows the final mean square error mse for each training. This error was used as the measure of quality during the learning process. It can be seen that the variation of mse is not big: the best mse is 0,015, while the worst is 0,022. The upper part of Figure 8 presents the results for the NN with the best performance out of the 100 trials. This part of the figure shows the output data $u1$ and the output of the network ($u1$NN) when fed with the same input. The network closely follows human control behavior, especially in Zones 2 and 3.

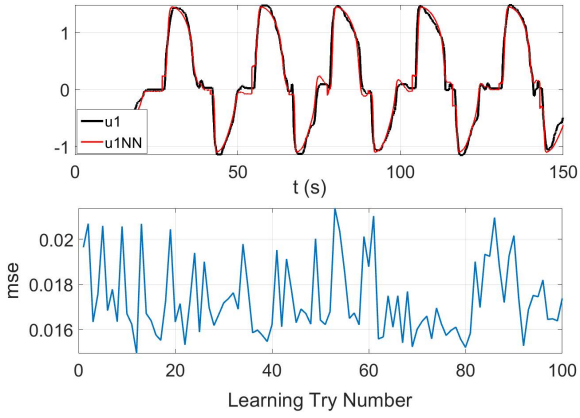


Figure 8. Selection of the best neural network.
Source: Authors

Running Stage (R)

This final time scaling control stage uses the best Neural Network of the previous stage. The best network defines an actuating signal based on the error value in order to control the models and the motor itself. The same network can control models and motor since the Multilayer Perceptron Neural Network is blind to changes in the scale of time. This type of NN produces the same outputs if the same inputs are given, regardless of time variation. If an input changes, the corresponding output is updated after a few mathematical operations. These operations are instantaneous in practice, given the speed of the systems in contrast with the time it takes to propagate a network. Thus, the same network controls the simulated system and the real plant.

Testing the trained NN starts with running the control over the scaled model H' because the controller was trained using data from H'. The test during the scaling stage lasts 2,5 minutes, which is enough time to train the network. However, the simulation in this stage may last longer, that is, 30 minutes or more. It can be seen that there is no risk that the network will stop paying attention, as it actually happens with humans after a few minutes. In addition, instead of two unique values ±100°, it is better to test several and random amplitudes in the rank ±90°. If the control works properly, the next step in testing consists of using the same network over system H. This last simulation could take a single minute, because the system runs at its normal speed (remember that a minute at the unscaled time corresponds to 30 minutes using the scaled time, $kt = 1/30$). Thus, the transitions between one amplitude and another may last from 0,3 s to 0,6 s. If this test succeeds, it is possible to run the final test, which is to control the real plant. If not, the *Training* stage should be run again.

The final and definite test uses the trained NN to control the real plant. The block diagram of the connections inside the computer is shown in Figure 9. It can be seen that the input of the network is scaled dividing the error by 180. The output of the controller reaches the plant using a data acquisition card, the National Instruments PCI-6024E. This card translates the actuating signal inside the computer into voltage to feed the power amplifier that drives the motor. Figure 9 also shows a traditional PID controller with the same tuning that was used during the *Scaling* stage, it is a proportional controller with $k_p = 1/10$.

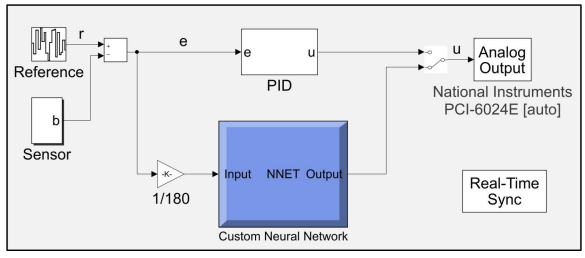


Figure 9. Block diagram to control the real plant.
Source: Authors

Figure 10 shows the performance of the Time Scaling Controller during the control of the real plant. The upper part of the figure presents the reference, the output with a PID controller, and the output with the proposed controller in this paper. TSC presents a lower overshoot and reaches the reference better, as can also be seen in the middle figure. It has a null error for the steady state after almost all the transients, but the PID has a steady state error of about 4°, this is due to the effect of the dead zone. The lower part of Figure 10 shows the actuating signal for both controllers. It is important to remark that TSC does not reach the limits of ±2, but the PID controller does it. This characteristic reveals that the trained NN uses the energy to control the plant better than the PID for the experiment in the figure. The switching of the actuating signal for TSC happens because the controller reaches null error, so that the signal jumps alternatively from negative (or positive) to a null value. The actuating signal for the PID does not switch because it almost never reaches the null error. Finally, a measure of the performance for both controllers is the index Integral of the Absolute Error (IAE). For a trial of ten seconds, the result for the PID is 56,8 and 37,8 for TSC. Thus, the ratio between the two controllers is about 1,5 in favor of the Time Scaling Controller.

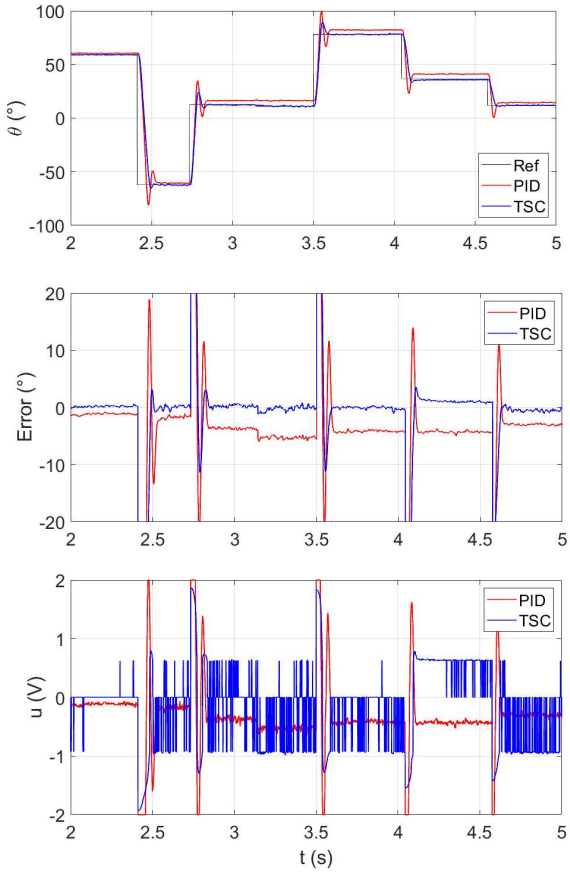


Figure 10. Control performance comparison.
Source: Authors

Discussion

Results in Figure 10 show the advantage of using TSC to regulate the position of a motor in contrast with a PID controller. However, a tuning process could improve the PID control performance, so that the apparent advantage of TSC could decrease. On the other hand, it is also true that the model of the plant could be improved, as well as the training of the NN. Thus, the ratio of 1,5 when comparing controller performances in the previous section is not definitive. Instead of a numerical ratio, this paper aims to show how to use the scaling time as a key concept in controlling dynamic systems, more specifically a motor. Both PID and TSC can be optimized, but that is not the purpose of this paper.

An aspect that can be studied to improve the performance of TSC is the influence of the human controller in the generation of data to train a NN. The data used in this paper comes from the author as the controller of the scaled model, but another person could generate different control actions, thus resulting in a different NN. Thus, TSC performance can be improved by looking at data coming from different human controllers. Another aspect that affects TSC performance is the model used to get the data. The model in Figure 3 was enough to train a NN to control the real plant, as can be seen in Figure 10. However, the difference between the output of H', H, and the real plant shows that it is possible to consider another model. It was evident that the real plant presented overshoot, while the models do not. In addition, the real plant was faster than its simulated counterpart.

The oscillation of the control signal $u\,(t)$ in the lower part of Figure 10 shows another feature of the proposed controller that was not programmed intentionally but emerged from the definition of the system itself. If the absolute value of the error is smaller than 10°, the actuating signal $u\,(t)$ switches between two values, which resembles the work of a PWM when switched to increase the efficiency of a system. TSC applies energy to the system only when it is necessary and the error is larger than 1° in this application. By doing so, the controller avoids system oscillation. However, if the absolute value grows larger than 10°, the actuating signal has a traditional continuous shape. These two control actions will be the focus of future studies on the application of TSC.

Finally, an important aspect of TSC is how the trained NN controls systems at different time scales. You can see that the network itself does not have any reference to time. However, given that the network in this paper is implemented in a computer, the network updates its output at the sample rate t_s. In this paper, t_s = 1 ms, whereas t_s' = 30 ms. Thus, and given that the stabilization time of the motor is about 20 ms, or 600 ms for H', then the updating of the network can be considered to happen instantly. Consequently, the actuating signal changes at the same speed of the error, and the error changes at the same speed of the feedback signal coming from the sensor ($e = y - r$). Thus, the actuating signal changes at the rate of the plant's output y. In summary, the controlled plant itself defines the speed of the whole system, which is not defined by the NN. If the plant is fast, the actuating signal changes accordingly; on the contrary, if the plant is slow, the network also produces a slow actuating signal.

Conclusions

The experiments in this paper show that Time Scaling Control properly controls the angular position of a motor and even results in better control performance than a PID controller. The main reason for this is that TSC uses the control knowledge of a human, which daily handles nonlinear situations and is adaptive and robust in comparison with a fixed and linear rule such as the PID controller.

The human plays the main role in TSC because he directly influences the outcome of three processes: first, defining the model of the system, as shown in Figure 3 and Figure 4; second, setting the experiment to acquire data, as shown in Figure 5 and Figure 6; and third, being the source of the knowledge to train a NN, as shown in Figure 7 and Figure 8. In the case of stages 1 and 2, the knowledge is explicit: the human is an expert in the system. In 3, TSC extracts hidden knowledge inside the brain, which may be the main contribution of TSC.

There are still many interesting topics to explore in TSC, but the author will provide just three of them as an example:

1. The definition of the minimal components of a model in order to have enough information to run TSC.
2. The definition of a deterministic method to select the best scaling factor k_t, according to the system model.
3. The analysis of data coming from different persons and the study of the effect of those differences in TSC performance.

References

Anoop, C. S. and George, B. (2013). New Signal Conditioning Circuit for MR Angle Transducers with Full-Circle Range. *IEEE Transactions on Instrumentation and Measurement, 62*(5), 1308-1317. 10.1109/TIM.2012.2236778

Aquino, A. and Velez, M. (2006). *Gray-Box Modeling of Electric Drives using Recursive Identification and Radial Basis Functions.* Paper presented at IECON 2006 - 32nd Annual Conference on IEEE Industrial Electronics, Paris, 1447-1452. 10.1109/IECON.2006.348111

Corno, M., Giani, P., Tanelli, M., and Savaresi, S.M. (2015). Human-in-the-Loop Bicycle Control via Active Heart Rate Regulation. I*EEE Transactions on Control Systems Technology, 23*(3), 1029-1040. 10.1109/TCST.2014.2360912

Duverne, S. and Koehlin, E. (2017). Rewards and Cognitive Control in the Human Prefrontal Cortex. *Cerebral Cortex, 27*(1), 5024-5039. 10.1093/cercor/bhx210

Huang, J., Chen, Y., and Li, Z. (2015). *Human operator modeling based on fractional order calculus in the manual control system with second-order controlled element*. Paper presented at The 27th Chinese Control and Decision Conference (2015 CCDC), Qingdao, 4902-4906, 10.1109/CCDC.2015.7162802

Inga, J., Köpf, F., Flad, M., and Hohmann, S. (2017). *Individual human behavior identification using an inverse reinforcement learning method*. Paper presented at 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Banff, AB, 99-104. 10.1109/SMC.2017.8122585

Kara, T. and Eker, I. (2004). Nonlinear modeling and identification of a DC motor for bidirectional operation with real time experiments. *Energy Conversion and Management, 45*(1), 1087-1106. 10.1016/j.enconman.2003.08.005

Laurense, V. A., Pool, D. M., Damveld, H. J., Paassen, M. R. M., and Mulder, M. (2015). Effects of Controlled Element Dynamics on Human Feedforward Behavior in Ramp-Tracking Tasks. *IEEE Transactions on Cybernetics, 45*(2), 253-265. 10.1109/TCYB.2014.2324037

Lee, D. (2015). *Incremental robot skill learning by human motion retargetting and physical human guidance*. Paper presented at the 12th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), Goyang, 585-586. 10.1109/URAI.2015.7358837

Mackie, M. A., Van Dam, N. T., and Fan, J. (2013). Cognitive Control and Attentional Functions. *Brain Cognitions, 82*(3), 301-312. 10.1016/j.bandc.2013.05.004

Medaglia, J., Pasqualetti, F., Hamilton, R., Thompson, S., and Bassett, D. (2017). Brain and Cognitive Reserve: Translation via Network Control Theory. *Neuroscience and Biobehavioral Reviews, 75*(1), 53-64, https://10.1016/j.neubiorev.2017.01.016

Rairán, J. D. (2017). *Control of Dynamic Systems Using Time Scaling*. *ASME 2017 Dynamic Systems and Control Conference, 1*(1), 1-10. 10.1115/DSCC2017-5048

Rapos, D., Mechefske, C., and Timusk, M. (2016). *Dynamic sensor calibration: A comparative study of a Hall Effect sensor and an incremental encoder for measuring shaft rotational position*. Paper presented at 2016 IEEE International Conference on Prognostics and Health Management (ICPHM), Ottawa, ON, 1-5. 10.1109/ICPHM.2016.7542858

Rios, F. and Makableh, Y. F. (2011). *Efficient position control of DC Servomotor using backpropagation Neural Network*. Paper presented at 2011 Seventh International Conference on Natural Computation, Shanghai, 653-657. 10.1109/ICNC.2011.6022230

Robla, S., Becerra, V. M., Llata, J. R., González, E., Torre, C., and Pérez, J. (2017). Working Together: A Review on Safe Human-Robot Collaboration in Industrial Environments. *IEEE Access, 5*(1), 26754-26773. 10.1109/ACCESS.2017.2773127

Suresh, P. and Manivannan, P. V. (2016). *Human driver emulation and cognitive decision making for autonomous cars*. Paper presented at the International Conference on Robotics: Current Trends and Future Challenges (RCTFC), Thanjavur, 1-6. 10.1109/RCTFC.2016.7893411

Van der El, K., Pool, D. M., Van Paassen, M. R. M., and Mulder, M. (2018). Effects of Preview on Human Control Behavior in Tracking Tasks with Various Controlled Elements. *IEEE Transactions on Cybernetics, 48*(4), 1242-1252. 10.1109/TCYB.2017.2686335

Wu, S. T. and Wang, Z. L. (2016). Equilateral Measurement of Rotational Positions with Magnetic Encoders. *IEEE Transactions on Instrumentation and Measurement, 65*(10), 2360-2368. 10.1109/TIM.2016.2578579

Wahyunggoro, O. and Saad, N. (2010). *Analysis and evaluation of real-time and s-domain model of A DC servomotor*. Paper preseted at 2010 International Conference on Intelligent and Advanced Systems, Kuala Lumpur, Malaysia, 1-5. 10.1109/ICIAS.2010.5716206

Xu, K. Z., Anderson, B. A., Emeric, E. E., Sali, A. W., Stuphorn, V., Yantis, S., and Courtney, S. M. (2017). Neural Basis of Cognitive Control over Movement Inhibition: Human fMRI and Primate Electrophysiology Evidence. *Neuron, 96*(6), 1447-1458. 10.1016/j.neuron.2017.11.010

Research Article / Electrical – Electronic Engineering http://dx.doi.org/10.15446/ing.investig.v40n3.82665

Optimal Coordination of Active Generators in a Grid-Connected Microgrid

Coordinación Óptima de Generadores Activos en una Microrred Interconectada

Adriana C. Luna[1], Nelson L. Díaz[2], and Eider Alexander Narvaez[3]

ABSTRACT

In a microgrid composed of distributed active generators based on renewable energy sources, with heterogeneous features and generation profiles, the availability of the energy resource, the energy reserve capacity, and the degradation of the storage unit, define the constraints for the management and dispatch of each active generator. This can result in sub-optimal use of distributed energy resources in comparison with the operation of a single generation unit. However, under the current trend oriented to distributed installations, the overall operation could be improved if an aggregated operation is considered within the management level. This paper proposes a coordinated operation of the storage units associated with distributed active generators for a hybrid grid-connected microgrid. In order to optimize the use of the active generators, including the equalization of the state of charge of the storage units, a mathematical model is proposed. This model tries to avoid uneven degradation of the storage units, and, consequently, enhance the reserve capacity and reduce the depth of discharge by achieving the operation of the distributed system as a unified system. The simulations are carried out in GAMS and MATLAB in order to validate the system's operation. The results show a better performing grid-connected microgrid with the proposed approach.

Keywords: distributed active generators, energy storage equalization, energy management systems, microgrids

RESUMEN

En una microrred compuesta por generadores activos distribuidos basados en fuentes de energía renovables con características heterogéneas y diferencias en sus perfiles de generación, la disponibilidad del recurso energético, la capacidad de reserva de energía y la degradación de la unidad de almacenamiento definen las limitaciones para la gestión y despacho de cada generador activo. Esto puede resultar en un uso subóptimo de los recursos de energía distribuida en comparación con la operación de una unidad de generación única. Sin embargo, bajo la tendencia actual orientada a instalaciones distribuidas, la operación general podría mejorarse si se considera una operación agregada dentro del nivel de gestión. Este documento propone una operación coordinada de las unidades de almacenamiento asociadas con generadores activos distribuidos para una microrred híbrida conectada a la red. Se propone un modelo matemático para optimizar el uso de los generadores activos, incluida la ecualización del estado de carga de los sistemas de almacenamiento. Este modelo intenta evitar la degradación desigual de dichas unidades de almacenamiento y, en consecuencia, mejorar la capacidad de reserva y reducir su profundidad de descarga al lograr que el sistema distribuido opere como un sistema unificado. Las simulaciones se desarrollan en GAMS y MATLAB con el objetivo de validar la operación del sistema. Los resultados muestran un mejor desempeño de la microrred interconectada con el enfoque propuesto.

Palabras clave: generadores activos distribuidos, ecualización de sistemas de almacenamiento de energía, sistemas de gestión de energía, microrredes

Received: October 4th, 2019
Accepted: August 7th, 2020

Introduction

The use of Renewable Energy Sources (RESs) such as photovoltaic (PV) generators or Wind Turbines (WT) have allowed the diversification of the energy generation matrix and the substitution of conventional sources based on fossil fuels, which produce a high environmental impact (Keyhani 2016). In this sense, electrical microgrids appear as a solution to integrate Distributed Energy Resources (DERs) with heterogeneous characteristics to feed local electrical loads both in isolated and grid-connected modes (IEEE 2011,

[1]Electronics Engineer, Universidad Nacional de Colombia, Colombia. M.Sc. in Indrustial Automation, Universidad Nacional de Colombia, Colombia. Ph.D. in Energy Technology Aalborg University, Denmark. Affiliation: Faculty of Mechanical, Electronics, and Biomedical Engineering, Universidad Antonio Nariño, Colombia. E-mail: aluna10@uan.edu.co
[2]Electronics Engineer, Universidad Distrital Francisco José de Caldas, Colombia. M.Sc. Industrial Automation, Universidad Nacional de Colombia, Colombia. Ph.D. in Energy Technology Aalborg University, Denmark. Affiliation: Associate Professsor, Universidad Distrital Francisco José de Caldas, Colombia. E-mail: nldiaza@udistrital.edu.co
[3]Electrical Engineer, Universidad Nacional de Colombia, Colombia. M.Sc. Electrical Engineering, Universidad de los Andes, Colombia. Affiliation: Associate Professor, Universidad Distrital Francisco José de Caldas, Colombia. E-mail: anarvaez@udistrital.edu.co

How to cite: Luna, A. C., Díaz N. L., and Narváez E. A. (2020). Optimal Coordination of Active Generators in a Grid-Connected Microgrid. *Ingeniería e Investigación*, *40*(3), 47-54. 10.15446/ing.investig.v40n3.82665

Lasseter 2002, Guerrero, Vasquez, Matas, de Vicuna, and Castilla 2011). The main characteristic of microgrids is that they allow controlling all their components and distributed resources as a single and coordinated entity. Therefore, an Energy Management System (EMS) is required to optimize the use of local resources according to pre-established operating objectives (CEN-CENELEC-ETSI 2014, Katiraei, Iravani, Hatziargyriou, and Dimeas 2008, Toro, Baron, and Mojica-Nava 2019).

The high variability of RESs based on PV and WT, in addition to their dependence on environmental conditions, impose additional technical challenges in terms of reliability and availability of the energy resource. One of their main characteristics is that these types of sources, by themselves, are commonly used as non-dispatchable sources and, therefore, do not participate in the regulation and management of the microgrid (Rafique and Jianhua 2018), since they usually operate following a Maximum Power Point Tracking (MPPT) algorithm that seeks to obtain the maximum available energy (Thongam and Ouhrouche 2011, de Brito, Galotto, Sampaio, e Melo, and Canesin 2013). Therefore, within the microgrid models renewable energy is usually considered as an uncontrollable input (Arcos-Aviles, Pascual, Guinjoan, Marroyo, Sanchis, and Marietta 2017).

In order to include a certain degree of flexibility in the EMS of the networks that incorporate RESs, a possible approach to deal with the surplus of energy is to manage the power exchange between the generation unit and an additional storage system (Blaabjerg, Yang Y., Yang D., and Wang 2017). This strategy constitutes the operation of active generators, which integrate variable energy resources with Energy Storage Systems (ESSs), power conditioning systems and a dedicated local control algorithm (Kanchev, Colas, Lazarov, and Francois 2014). In this way, active generators can deliver energy in a smoothed way in the short term, and balance the generation and demand for energy in the medium term. Thus, active generators based on highly variable RESs, such as PV and WT generators, can be used as dispatchable units as long as their storage systems are operated under a partial State of Charge (SoC) (Azmi, Kolhe, and Imenes 2015, Hill, Such, Chen, Gonzalez, and Grady 2012).

In the scheduling process for the operation of energy devices in microgrids based on active generators, a set point of operation for each active generator is usually assigned from the EMS. In other words, the EMS considers distributed active generators as independent and dispatchable units (Choudar, Boukhetala, Barkat, and Brucker 2015). Therefore, in order to achieve the requirements assigned to the active generator, each ESS independently compensates the generated energy of its associated RES, by storing or delivering energy. The management of microgrids with active generators has already been addressed. For instance, in (Yan, Abbes, Francois, and Bevrani 2016) an optimization algorithm is developed by taking into account the restrictions of each one of the storage units. In Azmi and Kolhe (2015), an algorithm is proposed for the management of active generators in a microgrid by means of state flow control. In Limouchi, Taher, and Ganji (2016), the dispatch of active and reactive power for microgrids with active generators is carried out. Although these works define restrictions related to the operation of ESSs, no strategies have been proposed in order to achieve an aggregated or collaborative operation between active generators intended to reduce the overall degradation of the storage units due to uneven cycles of operation and and different depths of discharge.

In Diaz, Luna, Vasquez, and Guerrero (2017), the coordination of the ESSs integrated into the active generators for an islanded microgrid is presented. In this work, the concept of SoC equalization for distributed ESSs is introduced with the aim to minimize the difference between their charging profiles by adjusting parameters on the primary controllers and the power-sharing process. As a result, the overall operation of distributed low-capacity ESSs was coordinated and their charge profiles were equalized, which in the end, achieved that the distributed energy storage units operate just as a single aggregate system with greater capacity would. This approach of managing distributed units as a single and aggregated one, adds to the electrical system the inherent advantages of a distributed system, such as redundancy, the easy expansion, and elimination of a single point of failure, without losing the simplicity offered in the management process when dealing with a single centralized storage unit. However, the work in Diaz, Luna, Vasquez, and Guerrero (2017), does not aim to optimize the operation of the microgrid and not include exchanges of energy with the utility grid.

In this work, a schedule is proposed for the charging and discharging profiles of the ESSs which are integrated into the active generators of a grid-connected microgrid, in order to equalize the SoC profiles in an optimal and coordinated way. In consequence, the main aim is to achieve the operation of the distributed ESSs as a single with a greater capacity. Firstly, a general optimization problem is defined considering the active generators and the power exchange with the main grid. After that, the model is improved by equalizing the SoCs for the distributed ESSs. As a result, the ESSs would work in an analogous way to which a centralized storage system based on batteries does, if possible. In the end, the proposed strategy manages to reduce the overall degradation of the distributed energy storage systems based on batteries.

This document is organized as follows: section II presents the general description of the microgrid considered for the case study; section III defines the energy management system of the microgrid and the mathematical model of the equalization for the SoCs of the ESSs; section IV, the validation of the proposed strategy is presented; and, finally, section V presents conclusions.

Microgrid Description

The microgrid established as a case study is connected to the conventional electrical network and is composed of n_i active generators based on RES with banks of batteries based ESSs, as shown in Figure 1. The main grid defines the voltage and frequency of the microgrid, while the DERs operate as power

sources (Katiraei, Iravani, Hatziargyriou, and Dimeas 2008). The primary controllers are not the object of this study, so interested readers can refer to other sources such as Farrokhabadi, König, Cañizares, Bhattacharya, and Leibfried (2018); Diaz, Vasquez, and Guerrero (2018); and Luna, Diaz, Graells, Vasquez, and Guerrero (2017).

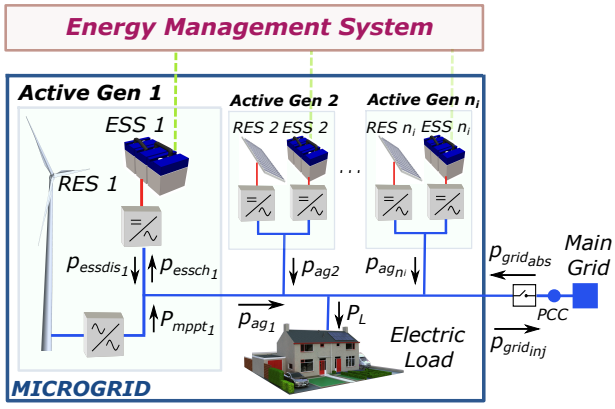


Figure 1. Hybrid grid-connected microgrid based on active generators.
Source: Authors

The microgrid includes an EMS to schedule the charging and discharging of the batteries based on the prediction of the available renewable energy and the energy demand of the load for a 24-hour time horizon. In this work, a reliable prediction for this time horizon is assumed (Negnevitsky, Johnson, and Santoso 2007). Therefore, the effects of uncertainty over the proposed strategy related to the generation/demand forecast are not considered. Interested readers may consult related literature such as Li, Vrakopoulou, and Mathieu (2018), and Gigoni et al. (2018).

For the development of the proposal presented in this document, the simplified mathematical models defined by Fathima and Palanisamy (2015), and Moradi, Esfahanian, Abtahi, and Zilouchian (2018) have have been used to determine the output power of RESs integrated into the active generators.

In the case of WT, the defined model sets the wind speed as input in a piece-wise function that can be written as

$$P_w(v) = \begin{cases} c_w * v^3 & 0 \leq v \leq v_{nom} \\ c_w & v_{nom} \leq v \leq v_{stop} \\ 0 & v \geq v_{stop} \end{cases} \quad (1)$$

where, $P_w(v)$ is the WT power as a function of the wind speed v; $c_w = 1500\ W/(m/s)^3$ is a constant related to the geometrical characteristics of the rotor; $v_{nom} = 10\ m/s$ is the nominal wind speed; and $v_{stop} = 24\ m/s$ is the cut-off wind speed (Wais 2017).

For the PV model, the irradiance I is defined as input. In this way, the power is established as a linear relation, so that

$$P_v(I) = \begin{cases} 0 & I \leq I_{min} \\ c_v * I & I_{min} \leq I \leq I_{max} \end{cases} \quad (2)$$

where, $P_v(I)$ is the output power of the PV in terms of the irradiance, I; $I_{min} = 200\ W/m^2$ is the minimum irradiance from which the solar cells start generating energy; $c_v = 1,415\ m^2$ is the generation coefficient of the PV arrays; and $I_{max} = 1000\ W/m^2$ is established as the standard irradiance on the terrestrial surface, for which the rated PV power is defined.

On the other hand, the ESSs of the active generators considered in this paper are banks of lithium-ion (Li-ion) batteries, since this technology is displacing the traditional batteries of Lead-acid in stationary applications, due to its longer duration and better performance (Marra and Yang 2015). The batteries operate under a constant power load strategy, and their SoC should be maintained between a 20% and 90% to avoid operation in unsafe regions (Linden and Reddy 2001, Diaz, Luna, Vasquez, and Guerrero 2017, Luna, Meng, Diaz, Graells, Vasquez, and Guerrero 2018, Marra and Yang 2015).

Energy Management System (EMS)

The implemented EMS seeks to schedule the use of the ESSs of the different active generators in such a way that the consumption of electrical energy absorbed from the main grid is minimized, and, at the same time, maximizes the use of the generated renewable energy within the microgrid. To this effect, the optimization problem poses a general model that allows the scalability of the number of active generators and the flexibility related to the characteristics of their DERs. It should be noted that the model focuses on the ESSs management and does not consider any demand response program.

Formulation of the Mathematical Model

The optimization model is defined for a time horizon of $T = 24$ h. The discrete time is represented by the index $h = 1, 2, \ldots, H$, in intervals of $\Delta h = 1$ hour. In this way, the number of time slots is $H = \frac{T}{\Delta h}$. On the other hand, the index $i = 1, \ldots, n_i$ is used to define the active generators. The inputs of the model are the prediction of the RES generation, $P_{mppt_i}(h)$ and the electric demand of the microgrid, $P_L(h)$.

The variables defined in the model are the average values at each time interval and have been denoted with lowercase letters so that they can easily be recognized within the optimization problem.

The definition of a set of constraints is required to confine the solution to a feasible region. First, the microgrid must comply with the energetic balance (as shown in Figure. 1) which can be defined as,

$$\sum_{i=1}^{n_i} p_{ag_i}(h)\Delta h + \left\{p_{grid_{abs}}(h) - p_{grid_{inj}}(h)\right\}\Delta h = P_L \Delta h,\ \forall h \quad (3)$$

where $p_{ag_i}(h)$ is the power of the active generators, and $p_{grid_{abs}}(h)$ and $p_{grid_{inj}}(h)$ are the absorbed/injected power from/to the main grid.

In turn, the power provided by the active generators, $p_{ag_i}(h)$, corresponds to the sum of the power generated by the associated RES and ESS, so that,

$$p_{ag_i}(h) = P_{mppt_i}(h) + \left\{p_{essdis_i}(h) - p_{essch_i}(h)\right\}, \forall h, i \quad (4)$$

where $p_{essch_i}(h)$ and $p_{essdis_i}(h)$ are the amount of power charged and discharged by the ESSs, respectively.

The SoC of the ESSs, $soc_i(h)$, can be estimated using the Ah-counting method as a function of the previous time and the energy stored/discharged at the current time, so that,

$$soc_i(h) = \begin{cases} SoC_i(0) - \frac{\{p_{essdis_i}(h)\Delta h\}*100\%}{\eta_{dis_i}*Cap_{ESS_i}} \\ \quad + \frac{\eta_{ch_i}*\{p_{essch_i}(h)\Delta h\}*100\%}{Cap_{ESS_i}}, \forall i, h = 1 \\ soc_i(h-1) - \frac{\{p_{essdis_i}(h)\Delta h\}*100\%}{\eta_{dis_i}*Cap_{ESS_i}} \\ \quad + \frac{\eta_{ch_i}*\{p_{essch_i}(h)\Delta h\}*100\%}{Cap_{ESS_i}}, \forall i, 2 \le h \le H \end{cases} \quad (5)$$

where $\eta_{dis_i} = 0,95$ and $\eta_{ch_i} = 0,98$ correspond to the discharging and charging efficiencies of the ESSs, respectively, and Cap_{ESS_i} are their capacities in kWh.Ft

In addition, all variables have been defined as positive and have been upper-bounded. This has been modeled as

$$0 \le p_{ag_i}(h) \le P_{RES_{nom_i}}, \quad \forall h, i \quad (6)$$

$$0 \le p_{grid_{abs}}(h) \le P_{grid_{max}}, \quad \forall h \quad (7)$$

$$0 \le p_{grid_{inj}}(h) \le P_{grid_{max}}, \quad \forall h \quad (8)$$

$$0 \le p_{essch_i}(h) \le P^{charge}_{ESS_{max_i}} * x_{bat_i}(h), \quad \forall h, i \quad (9)$$

$$0 \le p_{essdis_i}(h) \le P^{discharge}_{ESS_{max_i}} * (1 - x_{bat_i}(h)), \quad \forall h, i \quad (10)$$

where $x_{bat_i}(h)$ is a binary variable to establish the status of charge of the ESSs. In Equation (6), the upper bound of $p_{ag_i}(h)$ is the nominal power of its associated RES, $P_{RES_{nom_i}}$. The power exchange between the microgrid and the utility is limited to not exceed $P_{grid_{max}} = 2$ kW, which corresponds to physical and/or market constraints. Likewise, The boundaries of the charging and discharging powers for the ESSs defined in Equations (9) and (10) are $P^{charge}_{ESS_{max_i}}$ kW and $P^{discharge}_{ESS_{max_i}}$ kW, respectively, as recommended by the manufacturers (Linden and Reddy 2001, Battery University n.d.).

Regarding the SoC of the ESSs, in the case of Li-ion technologies, even when they can be within the range [0%, 100%], the SoCs should not exceed the limits $\left[SoC_{min_i}, SoC_{max_i}\right] = [20\%, 90\%]$ to avoid an accelerated degradation (Linden and Reddy 2001). This restriction has been considered as a soft constraint by including the variables $\delta_{lo_i}(h)$ and $\delta_{lo_i}(h)$. In this way, the boundaries of the SoCs can be written as,

$$soc_i(h) \ge SoC_{min_i} - \delta_{lo_i}(h), \quad \forall h, i \quad (11)$$

$$soc_i(h) \le SoC_{max_i} + \delta_{up_i}(h), \quad \forall h, i \quad (12)$$

The deviations are bounded in such a way that,

$$\delta_{lo_i}(h), \delta_{up_i}(h) \ge 0, \quad \forall h, i \quad (13)$$

$$\delta_{lo_i}(h) \le SoC_{min_i}, \quad \forall h, i \quad (14)$$

$$\delta_{up_i}(h) \le 100\% - SoC_{max_i}, \quad \forall h, i \quad (15)$$

The decision variables are $p_{ag_i}(h)$, $p_{essch_i}(h)$, $p_{essdis_i}(h)$, $soc_i(h)$, $p_{grid_{abs}}(h)$, and $p_{grid_{inj}}(h)$, and have been included in the vector x. In light of the above,, the objective function, $J(x)$, is established as,

$$\min_x J(x) = \sum_{h=1}^{H} C_{grid} * \left\{p_{grid_{abs}}(h)\Delta h\right\} + \sum_{h=1}^{H} C_{penalty_{inj}} * \left\{p_{grid_{inj}}(h)\Delta h\right\} + \sum_{h=1}^{H}\sum_{i=1}^{n_i} \xi_{lo} * \delta_{lo_i}(h) + \sum_{h=1}^{H}\sum_{i=1}^{n_i} \xi_{up} * \delta_{up_i}(h) \quad (16)$$

where C_{grid} is the price of electricity per kWh; $C_{penalty_{inj}}$, is a penalty for injecting power into the main grid; and ξ_{lo} and ξ_{up} correspond to the penalty coefficients related to the soft constraints to lower and upper boundaries of the SoC presented in Equations (11) and (12). Thus, the first term in Equation (16) is related to the cost of using energy from the main grid. The second term is a penalty for injecting energy into the main grid (de la Hoz et al. 2019), which manages to prioritize the local use of RES energy and, consequently, store the energy surplus in the ESSs, whenever possible. Meanwhile, the last two terms in Equation (16) are penalties for injecting energy into the main grid, which help the ESSs to work mainly with partial SoC, to avoid their early degradation.

To sum up, the optimization model defined for the management of the microgrid is composed of Equations (3)-(16).

SoC Equalization

In this section, the coordination of the ESSs is proposed in such a way that they behave as a single ESS (aggregate storage system), i.e., their SoCs are equalized so that the charging/discharging cycles are unified. This condition has been modeled through the error variable between the SoCs, $\epsilon_{soc}(h)$ defined as,

$$\epsilon_{soc}(h) = \sum_{i=1}^{n_i - 1} \left(SoC_i(h) - SoC_{i+1}(h)\right)^2, \quad \forall h \quad (17)$$

The error of equalization should be minimized, so this condition has been included in the objective function:

$$\min_{\hat{x}} \hat{J}(\hat{x}) = \sum_{h=1}^{H} C_{grid} * \left\{p_{grid_{abs}}(h)\Delta h\right\} + \sum_{h=1}^{H} C_{penalty_{inj}} * \left\{p_{grid_{inj}}(h)\Delta h\right\} + \sum_{h=1}^{H}\sum_{i=1}^{n_i} \xi_{lo} * \delta_{lo_i}(h) + \sum_{h=1}^{H}\sum_{i=1}^{n_i} \xi_{up} * \delta_{up_i}(h) + \sum_{h=1}^{H} C_{eq} * \epsilon_{soc}(h) \quad (18)$$

where $\hat{x}$ is the previously-defined decision vector plus the equalization errors of the ESSs, $\epsilon_{soc}(h)$. Also, $C_{eq} = 0,00001$

DKK is the penalty coefficient related to those equalization errors. It has been adjusted by trial and error.

Validation

To evaluate the proposed strategies, GAMS v. 24.2.3 is used as an algebraic modeling language, selecting the BONMIN solver (GAMS 2013b, GAMS 2013a), while the input parameters and results are processed and analyzed by using Matlab R2018b (Mathworks 2016). The available power profiles of the renewable resources were obtained by means of the Equations (1) and (2), using wind speed and irradiance data from Aalborg, Denmark, which were acquired in (Aalborg University n.d.). Pseudo-code 1 summarizes the validation process.

Pseudo-code 1 Validation Process

```
function Data Conditioning(MATLAB)
    Load data of wind speed, solar irradiance, and Consumption;
    Obtain power generation profiles based on RES models;
    Load the parameters related to ESSs;
    Arrange the input data in a .gms file;
    Call GAMS using the input data.
end function
function Optimization Process(GAMS)
    Load the predefined optimization model;
    Load the data (.gms file);
    Solve the optimization problem (BONMIN solver);
    Create the file with the scheduled solutions (.gdx file);
end function
function Graphics and Results(MATLAB)
    Load the (.gdx file);
    Arrange the output data;
    Obtain Figures and histograms;
end function
```

As a particular case, $C_{grid} = 0,1DKK$, which is convenient to apreciate the efects of the proposed equalization scheme. Also, $C_{penalty_{inj}}$ is set to C_{grid}, while $\xi_{lo} = 1$ and $\xi_{up} = 1$.

Comparison of strategies in a 24-hour time horizon

The behavior of the system with the optimization strategy, described in section III.A. (optimization strategy from now on), and with the optimization plus equalization, presented in section III.B. (equalization scheme hereafter), is simulated for a specific day, so that the differences in operation of the proposed management systems can be observed. For this, the power profiles of the RES and the demand curve presented in Figure 2 and the parameters related to ESS included in Table 1 are established. In Figure 2 (P_w) corresponds to the generation profile for the WT generator in the active generator 1 (AG1), (P_{v1}) corresponds to generation profile for the PV generator in the active generator 2 (AG2), (P_{v2}) corresponds to generation profile for the PV generator in the active generator 3 (AG3), and (P_{v2}) corresponds to the load profile.

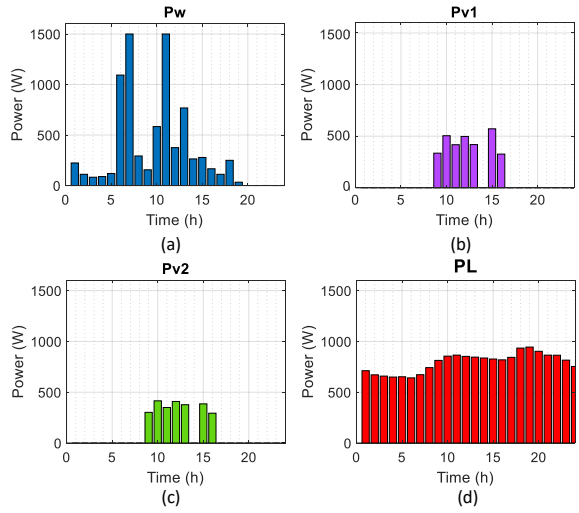


Figure 2. Generation and demand profiles used to compare the proposed strategies.
Source: Authors

Table 1. Parameters related to ESSs included in the optimization models.

	AG1	AG2	AG3
Cap_{ESS} (kWh)	10,75	6,72	6,72
$P^{charge}_{ESS_{max}}$ (W)	600	600	300
$P^{discharge}_{ESS_{max}}$ (W)	2000	2000	1000
$SoC(0)(\%)$	90	50	70

Source: Authors

The scheduling of the active generators considering the selected generation scenario using the two optimization strategies is shown in Figure 3(a) and (b), while the expected SoCs are presented in Figure 3(c) and (d). In this case, it is not necessary to absorb energy from the main grid.

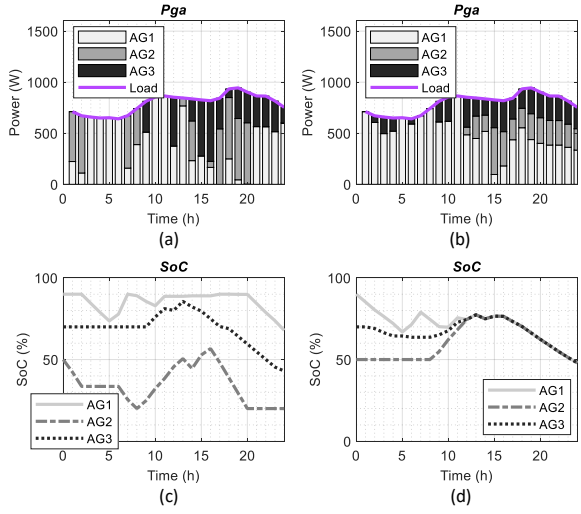


Figure 3. Simulation results for the microgrid with the optimization strategy ((a) and (c)) and with the optimization plus equalized strategy ((b) and (d)).
Source: Authors

It can be seen in Figures 3(a) and 3(c) that the contribution of the active generators under the optimization strategy does not follow a clear pattern to fulfill the demand requirement while the SoCs are within the safety range. Also, small micro-cycles of charge and discharge are generated that affect the lifetime of the batteries. The energy in each ESS (Figure 3(c)) does not correlate with the other ESSs in the microgrid. For this reason, the SoCs at the end of the day are very different. In addition, the maximum Depth of Discharge (DoD) in the ESSs, understood as the inverse of the SoC experienced by the ESSs, is 80%, which is the worst case for the life span of the ESSs based on batteries.

When the ESS equalization is applied, the scheduling of the active generators shown in Figure 3(b) and the SoC behavior observed in Figure 3(d) are obtained. In this case, it is also not required to absorb energy from the main grid.

It is observed that, in the equalized case, the active generator with the highest stored energy tends to provide more energy to the microgrid when the $SoCs$ do not equalize. When the equalization is achieved, the power scheduling is equal for the active generators with the same capacity (AG2 and AG3) while the unit with more capacity supplies more power. Therefore, the second strategy makes the ESS reaches lower DoDs, in this case, 50 %. Addionally, with the equalization, the shape of the SoCs is smoother with less micro-cycles.

In addition, the differences between the SoCs of the ESSs remain below 40% and decrease along T, reaching zero from the middle of the day (Figure 4).

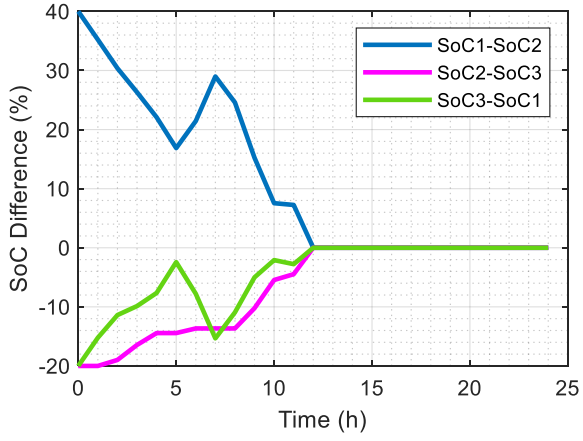


Figure 4. Differences among SoCs.
Source: Authors

In general, with the second strategy, the maximum DoD of the ESS is reduced and the behavior of the SoC profiles is softer, which means less degradation. The equalization manages the SoCs of the ESSs of the active generators in the fulfillment of the schedule in order to improve their performance.

Histogram of the minimum SoCs of the ESSs

In order to have a more representative sample of generation scenarios that show the impact of the ESS equalization of the active generators on the optimization of a hybrid grid-connected microgrid, eight days were randomly taken and the minimum SoC reached by the ESSs were obtained following the two proposed optimization strategies. An initial SoC of 50% was defined so that the effect of this parameter on results was decoupled. In Figure 5, the histogram of the minimum SoC reached every selected day for the three ESSs following both strategies is presented.

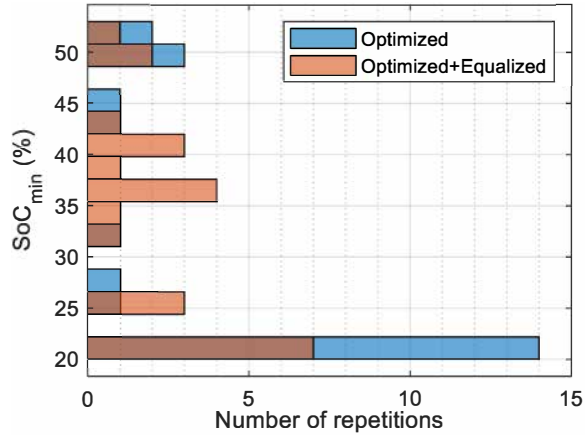


Figure 5. Histogram of the minimum SoCs of the ESSs following the optimized (blue bars) and the equalized (orange bars) strategies.
Source: Authors

For the selected scenarios, the optimized strategy caused DoDs greater than 65% in most cases, (17 times of 24), whereas, with the equalized strategy, this condition was evident 12 times. Additionally, a full discharge ($SoC = SoC_{min}$) was reached 14 times with the optimized strategy, while this occurred only 7 times when the equalization was performed.

Since the lifetime of the batteries is inversely related to the DoD (Linden and Reddy 2001), it is evident that the equalized strategy manages better the overall operation of the ESS and avoids the uneven degradation caused by an uneven DoD.

It should be noted that the costs associated with the implementation of each strategy $J(x)$ were recorded under the selected scenarios and were the same for both strategies in most cases, except one, where the cost with the optimized strategy was 1,3662 DKK and with the equalized one was 1,389 DKK. Therefore, it is concluded that the proposed equalization strategy has few effect on the cost of operation of the microgrid (1,64% in the worst case of the selected scenarios).

Conclusion

In this work, a novel energy management system is developed to minimize the exchange of energy between microgrids, based on active generators, and the electricity grid. In the proposal, optimal coordination of the ESSs of distributed active generators is carried out, so that they can work in an equalized manner, behaving as a unified storage unit, whenever possible, and avoiding uneven degradation. The mathematical formulation is defined in a generalized way that

allows considering several active generators with different sizes and storage characteristics. From the results of the simulation, it can be seen that the scheduling of the active generators is better distributed by means of SoC equalization, since their use depends on the energy stored and the energy capacities of their ESS. Meanwhile, in a conventional optimization approach, the use of the active generators is made indistinctly regardless of their current capacities and energy reserves. Additionally, smoother SoCs profiles and lower DoDs values are obtained in the ESSs of the active generators through their optimal coordination, which in the end is reflected in less degradation of the batteries, without affecting the operating costs of the microgrid. As future work, a quantitative analysis of the degradation of batteries using this strategy could be performed while also considering the uncertainty of the forecasting of renewable energy sources and loads.

Acknowledgements

This research has been co-financed by Colciencias and the Universidad de los Andes under the research project Nr. C145I005000000885-5 within the call 784-2017 and The Center for Research and Scientific Development of the Universidad Distrital Francisco José de Caldas (CIDC) (Project number: 2-5-596-19).

References

Aalborg University (n.d.). *Research Program: Photovoltaic System*. http://www.et.aau.dk/research-programmes/photovoltaic-systems/

Arcos-Aviles, D., Pascual, J., Guinjoan, F., Marroyo, L., Sanchis, P., and Marietta, M. P. (2017). Low complexity energy management strategy for grid profile smoothing of a residential grid-connected microgrid using generation and demand forecasting. *Applied Energy, 205*, 69-84. 10.1016/j.apenergy.2017.07.123

Azmi, A. N. and Kolhe, M. L. (2015). *Photovoltaic based active generator: Energy control system using stateflow analysis*. Paper presented at the 2015 IEEE 11th International Conference on Power Electronics and Drive Systems, Sydney, Australia. 10.1109/PEDS.2015.7203433

Azmi, A. N., Kolhe, M. L., and Imenes, A. G. (2015). *Review on photovoltaic based active generator*. Paper presented at the 2015 9th International Symposium on Advanced Topics in Electrical Engineering (ATEE), Bucharest, Romania. 10.1109/ATEE.2015.7133914

Battery University (n.d.). *Secondary* (*Rechargeable*) *Batteries - Battery University*. https://batteryuniversity.com/index.php/

Blaabjerg, F., Yang-Y., Y., Yang-D., D., and Wang, X. (2017). Distributed power-generation systems and protection, *Proceedings of the IEEE, 105*(7), 1311-1331. 10.1109/JPROC.2017.2696878

CEN-CENELEC-ETSI (2014). *Sgcg-m490-g smart grid set of standards version 3.1,* https://ec.europa.eu/energy/sites/ener/files/documents/\xpert_group1_sustainable_processes.pdf

Choudar, A., Boukhetala, D., Barkat, S., and Brucker, J.-M. (2015). A local energy management of a hybrid pv-storage based distributed generation for microgrids, *Energy Conversion and Management, 90*, 21-33. 10.1016/j.enconman.2014.10.067

Díaz, N. L., Luna, A. C., Vásquez, J. C., and Guerrero, J. M. (2017). Centralized control architecture for coordination of distributed renewable generation and energy storage in islanded ac microgrids. *IEEE Transactions on Power Electronics, 32*(7), 5202-5213. 10.1109/TPEL.2016.2606653

de Brito, M. A. G., Galotto, L., Sampaio, L. P., e Melo, G. d. A., and Canesin, C. A. (2013). Evaluation of the main MPPT techniques for photovoltaic applications. *IEEE Transactions on Industrial Electronics*, 60(3), 1156-1167. 10.1109/TIE.2012.2198036

de la Hoz, J., Martn, H., Alonso, A., Carolina Luna, A., Matas, J., Vasquez, J. C., and Guerrero, J. M. (2019). Regulatory-framework-embedded energy management system for microgrids: The case study of the spanish self-consumption scheme. *Applied Energy, 251*, 113374. 10.1016/j.apenergy.2019.113374

Diaz, N. L., Vasquez, J. C., and Guerrero, J. M. (2018). A communication-less distributed control architecture for islanded microgrids with renewable generation and storage. *IEEE Transactions on Power Electronics. 33*(3), 1922-1939. 10.1109/TPEL.2017.2698023

Farrokhabadi, M., König, S., Cañizares, C. A., Bhattacharya, K., and Leibfried, T. (2018). Battery energy storage system models for microgrid stability analysis and dynamic simulation. *IEEE Transactions on Power Systems. 33*(2), 2301-2312. 10.1109/TPWRS.2017.2740163

Fathima, A. H. and Palanisamy, K. (2015). Optimization in microgrids with hybrid energy systems a review. *Renewable and Sustainable Energy Reviews, 45*, 431-446. 10.1016/j.rser.2015.01.059

GAMS (2013a). *GAMS - The Solver Manuals, GAMS Release 24.2.1.* http://www.gams.com/dd/docs/solvers/allsolvers.pdf

GAMS (2013b). *General Algebraic Modeling System* (*GAMS*) *Release 24.2.1.* http://www.gams.com/

Gigoni, L., Betti, A., Crisostomi, E., Franco, A., Tucci, M., Bizzarri, F., and Mucci, D. (2018). Day-ahead hourly forecasting of power generation from photovoltaic plants. *IEEE Transactions on Sustainable Energy, 9*(2), 831-842. 10.1109/TSTE.2017.2762435

Guerrero, J. M., Vasquez, J. C., Matas, J., de Vicuna, L. G., and Castilla, M. (2011). Hierarchical control of droop-controlled ac and dc microgrids; a general approach toward standardization. *IEEE Transactions on Industrial Electronics, 58*(1), 158-172. 10.1109/TIE.2010.206653

Hill, C. A., Such, M. C., Chen, D., Gonzalez, J., and Grady, W. M. (2012). Battery energy storage for enabling integration of distributed solar power generation. *IEEE Transactions on Smart Grid, 3*(2), 850-857. 10.1109/TSG.2012.2190113

IEEE (2011). *Guide for Design, Operation, and Integration of Distributed Resource Island Systems with Electric Power Systems*. 10.1109/IEEESTD.2011.5960751

Kanchev, H., Colas, F., Lazarov, V., and Francois, B. (2014). Emission reduction and economical optimization of an urban microgrid operation including dispatched PV-based active generators. *IEEE Transactions on Sustainable Energy, 5*(4), 1397-1405. 10.1109/TSTE.2014.2331712

Katiraei, F., Iravani, R., Hatziargyriou, N., and Dimeas, A. (2008). Microgrids management,*IEEE Power and Energy Magazine 6*(3), 54-65. 10.1109/MPE.2008.918702

Keyhani, A. (2016). *Design of Smart Power Grid Renewable Energy Systems*. IEEE Press Series on Power Engineering, Hoboken, NJ:, Wiley.

Lasseter, R. H. (2002). MicroGrids, *2002 IEEE Power Engineering Society Winter Meeting. Conference Proceedings,* 305-308. 10.1109/PESW.2002.985003

Li, B., Vrakopoulou, M., and Mathieu, J. L. (2018). Chance constrained reserve scheduling using uncertain controllable loads part ii: Analytical reformulation. *IEEE Transactions on Smart Grid, 10*(2), pp. 1618-1625. 10.1109/TSG.2017.2773603

Limouchi, E., Taher, S. A., and Ganji, B. (2016). *Active generators power dispatching control in smart grid,* 21st Conference on Electrical Power Distribution Networks Conference (EPDC). 10.1109/EPDC.2016.7514778

Linden, D. and Reddy, T. (2001). *Handbook of Batteries. New York, NY,* McGraw-Hill Education.

Luna, A. C., Diaz, N. L., Graells, M., Vasquez, J. C., and Guerrero, J. M. (2017). Mixed-integer-linear-programming-based energy management system for hybrid pv-wind-battery microgrids: Modeling, design, and experimental verification. *IEEE Transactions on Power Electronics, 32*(4), 2769-2783. 10.1109/TPEL.2016.2581021

Luna, A. C., Meng, L., Diaz, N. L., Graells, M., Vasquez, J. C., and Guerrero, J. M. (2018). Online energy management systems for microgrids: Experimental validation and assessment framework. *IEEE Transactions on Power Electronics, 33*(3), 2201-2215. 10.1109/TPEL.2017.2700083

Marra, F. and Yang, G. (2015). Chapter 10 - decentralized energy storage in residential feeders with photovoltaics.In P. D. Lu (ed.), *Energy Storage for Smart Grids* (pp. 277-294), Boston, MA: Academic Press. 10.1016/B978-0-12-410491-4.00010-5

Mathworks (2016). *MATLAB version 9.0.0.341360* (*R2016a*), The Mathworks, Inc., Natick, Massachusetts.

Negnevitsky, M., Johnson, P. and Santoso, S. (2007). *Short term wind power forecasting using hybrid intelligent systems. 2007* IEEE Power Engineering Society General Meeting, pp. 1-4. 10.1109/PES.2007.385453

Rafique, S. F. and Jianhua, Z. (2018). Energy management system, generation and demand predictors: a review. *IET Generation, Transmission, Distribution, 12*(3), 519-530. 10.1049/iet-gtd.2017.0354

Thongam, J. S. and Ouhrouche, M. (2011). MPPT control methods in wind energy conversion systems. In R. Carriveau (ed.), Fundamental and Advanced Topics in Wind Power, IntechOpen, Rijeka, chapter 15. 10.5772/21657

Toro, V., Baron, E. D., and Mojica-Nava, E. (2019). Control jerárquico optimizado para una microrred de ca bajo ataque. *Ingeniería, 24*(1), 64-82. 10.14483/23448393.13760

Wais, P. (2017). Two and three-parameter weibull distribution in available wind power analysis. *Renewable Energy, 103*, 15-29. 10.1016/j.renene.2016.10.041

Yan, X., Abbes, D., Francois, B., and Bevrani, H. (2016). *Day-ahead optimal operational and reserve power dispatching in a pv-based urban microgrid*. Paper presented at the 2016 18th European Conference on Power Electronics and Applications (EPE'16 ECCE Europe), Karlsruhe, Germany. 10.1109/EPE.2016.7695614

Research Article / Systems-Computer Engineering http://dx.doi.org/10.15446/ing.investig.v40n3.82906

EF1-NSGA-III: An Evolutionary Algorithm Based on the First Front to Obtain Non-Negative and Non-Repeated Extreme Points

EF1-NSGA-III: Un Algoritmo Evolutivo Basado en el Primer Frente para Obtener Puntos Extremos no Negativos y no Repetidos

Luis Felipe Ariza Vesga[1], Johan Sebastián Eslava Garzón[2], and Rafael Puerta[3]

ABSTRACT

Multi-Objective and Many-objective Optimization problems have been extensively solved through evolutionary algorithms over a few decades. Despite the fact that NSGA-II and NSGA-III are frequently employed as a reference for a comparative evaluation of new evolutionary algorithms, the latter is proprietary. In this paper, we used the basic framework of the NSGA-II, which is very similar to the NSGA-III, with significant changes in its selection operator. We took the first front generated at the non-dominating sort procedure to obtain non-negative and non-repeated extreme points. This open-source version of the NSGA-III is called EF1-NSGA-III, and its implementation does not start from scratch; that would be reinventing the wheel. Instead, we took the NSGA-II code from the authors in the repository of the Kanpur Genetic Algorithms Laboratory to extend the EF1-NSGA-III. We then adjusted its selection operator from diversity, based on the crowding distance, to the one found on reference points and preserved its parameters. After that, we continued with the adaptive EF1-NSGA-III (A-EF1-NSGA-III), and the efficient adaptive EF1-NSGA-III (A2-EF1-NSGA-III), while also contributing to explain how to generate different types of reference points. The proposed algorithms resolve optimization problems with constraints of up to 10 objective functions. We tested them on a wide range of benchmark problems, and they showed notable improvements in terms of convergence and diversity by using the Inverted Generational Distance (IGD) and the HyperVolume (HV) performance metrics. The EF1-NSGA-III aims to resolve the power consumption for Centralized Radio Access Networks and the Bi-Objective Minimum Diameter-Cost Spanning Tree problems.

Keywords: evolutionary algorithm, many-objective optimization problem

RESUMEN

Los problemas de optimización de varios objetivos se han resuelto ampliamente usando algoritmos evolutivos durante algunas décadas. A pesar de que los algoritmos NSGA-II y NSGA-III se emplean con frecuencia como referencia para evaluar nuevos algoritmos evolutivos, este último es propietario. En este artículo, utilizamos el marco NSGA-II, similar al NSGA-III, con cambios en su operador de selección. Tomamos el primer frente generado por ordenamiento no dominante para obtener puntos extremos no negativos y no repetidos. Esta versión del NSGA-III se llama EF1-NSGA-III, y su implementación no comienza desde cero; eso sería reinventar la rueda. En lugar de eso, tomamos el código NSGA-II de los autores en el repositorio del Laboratorio de Algoritmos Genéticos Kanpur para extender el EF1-NSGA-III. Luego ajustamos su operador de selección de la diversidad en función de la distancia de hacinamiento al que se encuentra usando los puntos de referencia y preservamos sus parámetros. Después continuamos con el EF1-NSGA-III adaptativo (A-EF1-NSGA-III), y el eficiente adaptativo EF1-NSGA-III (A2-EF1-NSGA-III) contribuyendo en la explicación de cómo generar diferentes tipos de puntos de referencia. Los algoritmos propuestos resuelven problemas de optimización con restricciones de hasta 10 funciones objetivos. Los probamos en una amplia gama de problemas de referencia, y mostraron mejoras notables en términos de convergencia y diversidad utilizando las métricas de rendimiento de Distancia Generacional Invertida (IGD) e Hipervolumen (HV). El EF1-NSGA-III tiene como objetivo resolver el consumo de energía para las redes de acceso de radio centralizado y los problemas del árbol de expansión de diámetro mínimo bi-objetivo.

Palabras clave: algoritmo evolutivo, problema de optimización de muchos objetivos

Received: November 18th, 2019
Accepted: September 2nd, 2020

Introduction

Genetic algorithms (GAs) are random-based evolutionary methods. They are preferred over classical optimization

[1]Electrical Engineer, Universidad Nacional de Colombia, Colombia. M.Sc. Electronic and computers Engineering, Universidad de los Andes, Colombia. Affiliation: Ph.D. Student, Universidad Nacional de Colombia, Colombia. E-mail: lfarizav@unal.edu.co
[2]Electrical Engineer, Universidad Nacional de Colombia, Colombia. M.Sc. and Ph.D. Electrical Engineering, Universidade de Sao Paulo (USP), Brazil. Affiliation: Associate Professor, Universidad Nacional de Colombia, Colombia. E-mail: jseslavag@unal.edu.co
[3]M.Sc. Electronics Engineer, Pontificia Universidad Javeriana, Colombia. Ph.D. Photonics Engineering, Technical University of Denmark, Denmark. Affiliation: Senior Researcher, Ericsson, Sweden and Lecturer, Pontificia Universidad Javeriana, Colombia. E-mail: rafael.puerta@ericsson.com

How to cite: Ariza-Vesga, L. F., Eslava-Garzón, J. S., and Puerta, R. (2020). EF1-NSGA-III: An Evolutionary Algorithm Based on the First Front to Obtain Non-Negative and Non-Repeated Extreme Points. *Ingeniería e Investigación*, *40*(3), 55-69. 10.15446/ing.investig.v40n3.82906

due to their versatility in solving complex problems and finding multiple solutions without **a priori** information (Deb, 1999). They are inspired by fundamental genetic laws such as natural selection, first introduced by Fraser (1957), and further popularized by Holland (1975). When we study GAs, we contemplate an initial population of pseudo-random solutions that pass through genetic functions such as selection, crossover, and mutation to recombine and perturb solutions. Then, we evaluate these solutions with a fitness function in the hope of creating the fittest ones that will survive and evolve to the next generation. Finally, this process ends when we use a predefined termination criterion.

GAs then evolved to Multi-Objective and Many-objective Evolutionary Algorithms (MOEAs and MaOEAs) to optimize Multi-Objective or Many-Objective Optimization Problems (MOOPs and MaOPs) for fields such as engineering, business, mathematics, and physics (Li, Wang, Zhang, and Ishibuchi, 2018). They search for multiple solutions simultaneously on different non-convex and discontinuous regions next to the approximated Pareto front. Initially, research focused on solving MOOPs, but recently, there is an increasing interest in solving MaOPS. However, for MaOPs, we have a significant number of non-dominated solutions that exponentially increase with the number of objective functions. This is due to the selection operator and the dominant resistance caused by the dimensionality curse (Purshouse and Fleming, 2007).

The practical motivation of this paper is the implementation of an open-source version of the proprietary NSGA-III (Deb and Jain, 2014; Jain and Deb, 2014) called the EF1-NSGA-III (Ariza, 2019) that alleviates the above-mentioned issue of dimensionality. This algorithm solves problems with more than two objective functions, checks the feasibility of the population to fill the non-dominated sort procedure, and then uses the first front to generate non-negative and non-repeated extreme points during the normalization procedure. It has already been employed to reduce the power consumption for Cloud Radio Access Networks (Ariza, 2020), and resolve the Bi-Objective Minimum Diameter-Cost Spanning Tree problem (Prakash, Patvardhan, and Srivastav, 2020). The authors who solved the latter used the EF1-NSGA-III as a basis to generate a new algorithm called the Permutation-code NSGA-III (P-NSGA-III).

The EF1-NSGA-III uses a non-parametric method for diversity and executes faster when we take the renowned and efficient NSGA-II code, with prominent features such as simplicity, and an elitist approach. This algorithm, found at the repository of the Kanpur Genetic Algorithms Laboratory (KanGAL, 2011), is the core of the EF1-NSGA-III, but with significant changes to the selection operator. Our strategy for this paper is borrowed from professor Kalyanmoy Deb and researchers at Michigan State University. It helped them create and unify the NSGA-III to solve any mono-objective, multi-objective, and many-objective problem (Seada and Deb, 2015).

After we finished the extension of the EF1-NSGA-III, we continued with the adaptive EF1-NSGA-III (A-EF1-NSGA-III), and the efficient adaptive EF1-NSGA-III (A^2-EF1-NSGA-III), which use different schemes of adaptive reference points to increase their associated number of population members and accomplish a better distribution of solutions. The last two new algorithms are inspired by the A-NSGA-III and A^2-NSGA-III. These algorithms have already been implemented by referenced authors (Jain and Deb, 2013). Also, we contribute to explain how to generate reference points using the Das and Dennis (1998), two-layer, and adaptive methods.

The above contributions are the smoothest way we found to create a robust algorithm, rather than starting from scratch, as did Yarpiz (Matlab) (2018), jMetal (Java) (2018), nsga3cpp (C++) (Chiang, 2014), nsga3 (Python) (Marti, 2016), and PlatEMO (Matlab) (Tian, Cheng, Zhang, and Jin, 2017). Recently, a Multi-objective Optimization framework in python called pymoo was created, and its NSGA-III implementation is available (Blank and Deb, 2020). This implementation is employed to compare the NSGA-III and the EF1-NSGA-III in terms of some performance metrics.

In the remainder of this paper, we present a revision of related works. Then, we describe our extensions of the non-dominated sorting genetic algorithms EF1-NSGA-III, A-EF1-NSGA-III, and A^2-NSGA-III. After that, we present a detailed performance evaluation using statistical analysis. Finally, we draw conclusions.

Related work

Many real-life problems are MaOPs, and a whole army of evolutionary algorithms (Seada, Abouhawwash, and Deb, 2017, 2018) is waiting to be utilized to solve them. For example, some representative algorithms are the MOEA/D (Zhang and Li, 2007), the knee point-based algorithm (KnEA) (Zhang, Tian, and Jin, 2015), the HypE (Bader and Zitzler, 2011), and the NSGA-III. They are based on decomposition, convergence enhancements, indicators, and reference points, respectively. Any evolutionary algorithm mentioned before or even another extension of them has the chance to be improved considering different stages such as the generation of the parents and offspring populations, as well as the use of novel genetic functions that recombine and perturb solutions, the information feedback of individuals from previous iterations (Wang and Tan, 2019), or the evaluation of population members (the fitness allocation method) that allow solutions passing to the next generation.

Some of the already implemented learning methods to improve evolutionary algorithms are also inspired by nature, but, in this case, they mimic a herd or colony's behavior. We can mention some swarm intelligence methods as examples, such as chaotic krill herds (Wang, G., Guo, Gandomi, Hao, and Wang, H., 2014), elephant herds (Li,, Wang, and Alavi,, 2020), bee colonies (Wang and Ji, 2017), monarch butterflies (Wang et al., 2019), and moth colonies (Wang, 2016). These methods simulate the clustering behavior of chaotic krills, elephants, and also other insect behaviors such as migration and phototaxis. These algorithms imitate how groups of animals cooperate and learn by themselves or use information from other members to do a specific

task such as finding food. Some NP-hard problems that benefited from those improvements are scheduling, image, feature selection, detection, path planning, cyber-physical social system, texture discrimination, saliency detection, classification, object extraction, economic load dispatch, global numerical optimization, multi-objective optimization, knapsack problem, and fault diagnosis problems.

As can be seen, there are many options to improve evolutionary algorithms including the NSGA-III algorithm. Some new extensions of the NSGA-III algorithm have been developed since 2013. These are the A-NSGA-III and A^2-NSGA-III, which allocate adaptive reference points to improve the distribution of solutions (Jain and Deb, 2013). Another example is the U-NSGA-III, which uses the NSGA-III as the basis for implementing a unified algorithm to solve problems with up to 15 objective functions (Seada and Deb, 2015). The θ-NSGA-III trades off the diversity and the convergence of problems with more than four objective functions focused on the improvement of the non-dominated sort procedure (Yuan, Xu, and Wang, 2014).

Other works are the EliteNSGA-III, which increases the accuracy and diversity of the NSGA-III by maintaining an elite population archive to preserve previously generated elite solutions that would probably be eliminated by the original NSGA-III (Ibrahim et al., 2016). The E-NSGA-III utilizes extreme solutions in the population generation module to improve the overall quality of solutions (Wangsom, Bouvry, and Lavangnananda, 2018). The IFM-NSGAIII solves large-scale problems, instead of the small-scale problems, as the NSGA-III does, and introduces information feedback to influence the offspring population (Gu and Wang, 2020). The P-NSGA-III that modifies the elitist framework of the NSGA-III uses preferred vectors to improve local search (Shu, Wang, W., and Wang, R., 2018). Also, the Permutation-coded NSGA-III encodes chromosomes as permutations of graph vertices (Prakash et al., 2020).

Two more developments are the set of NSGA-III SBXAM, NSGA-III SIAM, and NSGA-III UCAM algorithms, which use an adaptive mutation operator and evaluate simulated binary crossover (SBX), uniform crossover (UC), and single-point crossover (SI) operators in large-scale and online problems (Yi, Deb, Dong, Alavi, and Wang, 2018; Yi et al., 2020). Finally, there is the B-NSGA-III, which converts continuous NSGA-III into binary NSGA-III modifying the initialization, crossover, and mutation operators for band selection in cloud contaminated hyper-spectral images (Gupta and Nanda, 2019).

The EF1-NSGA-III, A-EF1-NSGA-III, and A^2-EF1-NSGA-III algorithms

The basic framework of the proposed EF1-NSGA-III is similar to the KanGAL NSGA-II and inherits its parameters (Jain and Deb, 2013, 2014; Deb and Jain, 2014; Blank and Deb, 2020). It resolves MOOPs or MaOPs with conflicting objectives and inequality constraints focused on finding non-dominated solutions. Mathematically, the definition of the optimization problem is:

$$\text{Minimize} \quad f_i(\boldsymbol{x}) \quad i = 1, 2, \ldots, M \tag{1}$$

$$\text{Subject to} \quad g_j(\boldsymbol{x}) \geq 0 \quad j = 1, 2, \ldots, J \tag{2}$$

$$h_k(\boldsymbol{x}) = 0 \quad k = 1, 2, \ldots, K \tag{3}$$

$$x_l^{(L)} \leq x_l \leq x_l^{(H)}, \quad l = 1, 2, \ldots, n \tag{4}$$

Algorithm 1 EF1-NSGA-III procedure

Input: H structured reference points Z^s, initial random parent population P_t, $\mid P_t \mid$= the population size N
Output : $P_t + 1$

if (The remainder when N is divided by 4) **then**
  Start algorithm
else
  Abort algorithm
end if
for 1 to a number of generations defined by the user **do**
  $S_t = \emptyset, i = 1$
  Q_t = Tournament selection+Mutation+crossover (P_t)
  $R_t = P_t \cup Q_t$
  $(F_1, F_2, \ldots)$ = Non-dominated-sort(R_t)
  repeat
    $S_t = S_t \cup F_i$ and $i = i + 1$
  until $\mid S_t \mid \geq N$
  Last front to be included: $F_l = F_i$
  if ($\mid S_t \mid = N$) **then**
    $P_{t+1} = S_t$, continue
  else
    $P_{t+1} = \cup_{j=1}^{l-1} F_j$
    Points to be chosen from $F_l : K = N - \mid P_{t+1} \mid$
    Normalize objectives and create reference set Z^r : Normalize ($\boldsymbol{f^n}$, S_t, F_1, Z^r, Z^s)
    Associate each member $\boldsymbol{s}$ of S_t with a reference point: $[\pi(\boldsymbol{s}, d(\boldsymbol{s})]$ = Associate (S_t, Z^r): closest reference point, d: distance between $\boldsymbol{s}$ and $\pi(\boldsymbol{s})$
    Compute niche counts of reference point $j \in Z^r$: $\rho_{j_{S_t/F_l}} = \sum_{s \in S_t/F_l}((\pi(\boldsymbol{s}) = j)?1 : 0)$, and $\rho_{j_{F_l}} = \sum_{s \in F_l}((\pi(\boldsymbol{s}) = j)?1 : 0)$
    Choose K members one at a time from F_l to construct P_{t+1}: Niching ($K, \rho_{j_{S_t/F_l}}, \rho_{j_{F_l}}, \pi, d, Z^r, F_l, P_{t+1}$)
  end if
end for

where $\boldsymbol{x}$ is a vector with p decision variables, M is the number of conflicting objectives, J is the number of inequality constraints, K is the number of equality constraints, and l is the number of bounds from low (L) to high (H).

We propose the EF1-NSGA-III, where individuals in the first front (F_1) are contemplated to obtain non-negative and non-repeated extreme points and guide the evolution of the algorithm. Algorithm 1 resumes the procedure.

The EF1-NSGA-III first checks if the remainder of N is divided by four to make systematic application pair-wise selection and pair-wise recombination operations (Seada and Deb, 2015). Then, it runs a finite number of generations defined in a loop by the user. In the beginning, the population P_t is generated randomly, and its size is N (afterwards, the parent population comes from the last generation). Its offspring Q_t is generated using tournament selection, mutation, and crossover operations, and its size is N. The combined parent and offspring population $R_t = P_t + Q_t$ has a size of $2N$. Population R_t is sorted according to different non-dominated sorting levels or fronts $(F_1, F_2, \ldots, F_l)$, where F_l is the last

front. All members of each front are selected one at a time to build Population S_t. If front members reach the size of the population P_t, there is nothing more to do for that generation and $S_t = P_{t+1}$. If S_t population plus the last front exceeds the size of P_t, then the last front F_l is accepted partially, and $l + 1$ onward fronts are rejected. The last front F_l selects members for population S_t that maximize its diversity through the niching operator. This operator requires the determination of reference points on a hyper-plane and adaptive normalization of population members, association operator until reaching the niching preservation operation. Finally, the next generation parent population P_{t+1} is S_t, plus members from the niching operator, improve better the diversity until reaching the size N. There is no difference between the EF1-NSGA-III and the NSGA-III relay on normalization, niching procedures, and selection.

In the following section, we explain our implementations to generate reference points using simple algebraic principles. They are the well-known Das and Dennis, two-layer, k-layer, adaptive, and efficient adaptive methods (1998).

Reference Point Generators

Das and Dennis approach

We used the Das and Dennis method (1998) for EF1-NSGA-III, A-EF1-NSGA-III, and A^2-EF1-NSGA-III algorithms to generate well-distributed reference points and ensure diversity in obtained solutions. This method is the basis to generate different types of reference points such as two-layer, k-layer, adaptive, and efficient adaptive.

The predefined Das and Dennis approach places reference points on a normalized hyper-plane equally to all axis and has an intercept of one on each axis. If *p* divisions are considered along the *M* axis, the total number of reference points H is determined by

$$H = \binom{M + p - 1}{p} \tag{5}$$

This reference point generator is used for problems with up to five objective functions. An example can be seen in Figure 1a, for $M = 3$ and $p = 12$. High-dimensional problems with more objective functions require a different approach because the number of reference points increases exponentially with the number of objective functions.

Two-layer of Reference Points

This alternative reduces the number of reference points for high-dimensional problems compared to the Das and Dennis method. This approach is divided into the following steps:

1. *Generate reference points of the boundary layer:* Use the Das and Dennis approach. The boundary layer must satisfy the plane equation $x_1 + x_2 + \cdots + x_{nobj} = 1$, where $nobj = M$.

2. *Generate reference points of the inside layer:* Use the Das and Dennis approach again, but reference points of the inside layer should be on the plane $x_1 + x_2 + \cdots + x_{nobj} = 1 - \frac{1}{ndiv-1}$. The number of divisions of the inside layer is $ndiv - 1$, where $ndiv = p$.

3. *Move reference points from the inside layer to the boundary layer:* Reference points of the inside layer are moved to the boundary layer by adding the value $d = \frac{1}{(ndiv-1)*nobj}$ in all dimensions. Figure 1b, right side, depicts the distribution of reference points.

K-layer of Reference Points

It follows the two-layer reference points procedure, but with more than one inside layer. Additional information can be revised in Jiang and Yang (2017). Figure 1b, on the left side, shows the reference point distribution.

Adaptive Reference Points

First of all, this method requires the identification of ***useful*** or ***crowded*** reference points in cases where the niche count $\rho_{j_{S_t/F_l}} > 1 (j \in \{0, 1, \ldots, H-1\})$; and second, the generation of adaptive reference points around j_{th}. Before a new adaptive reference point is accepted, it must be non-repeated and lie on the positive orthant. After the adding task, the niche value of all reference points is updated, and adaptive reference points with niche count $\rho_{j_{S_t/F_l}} = 0$ are deleted. In the next generations, adaptive reference points are added and deleted as long as ***useful*** reference points exist. The procedure for each ***crowded*** reference point is described below:

1. *Generate initial adaptive reference points for one division:* Initial adaptive reference points $\boldsymbol{x_1} = (\frac{1}{ndiv}, 0, \ldots, 0)$, $\boldsymbol{x_2} = (0, \frac{1}{ndiv}, \ldots, 0)$ and $\boldsymbol{x_{nobj}} = (0, 0, \ldots, \frac{1}{ndiv})$ are generated using the Das and Dennis approach to satisfy the plane equation $x_1 + x_2 + \cdots + x_{nobj} = \frac{1}{ndiv}$.

2. *Move adaptive reference points to the plane* $x_1 + x_2 + \cdots + x_{nobj} = 0$: it is necessary to project the point $\boldsymbol{p} = (0, 0, \ldots, 0)$ onto the plane $x_1 + x_2 + \cdots + x_{nobj} = \frac{1}{ndiv}$. Any reference point of the plane can be chosen, for example, $\boldsymbol{r} = (\frac{1}{ndiv}, 0, \ldots, 0)$. With this vector in mind, vector $\boldsymbol{h} = \boldsymbol{p} - \boldsymbol{r} = (0, 0, \ldots, 0) - (\frac{1}{ndiv}, 0, \ldots, 0) = (-\frac{1}{ndiv}, 0, \ldots, 0)$ is obtained. Immediately, the unit normal vector is $\vec{\boldsymbol{u}} = \frac{1}{\sqrt[2]{nobj}}(1, 1, \ldots, 1)$. The vector to add initial adaptive reference points is:

$$\vec{\boldsymbol{d}} = |\, \boldsymbol{h}.\boldsymbol{u} \,| * \vec{\boldsymbol{u}} = \frac{1}{ndiv * nobj} * (1, 1, \ldots, 1) \tag{6}$$

3. *Move adaptive reference points to the* ***crowded*** *reference point jth*: In this step, after niching values $\rho_{j_{S_t/F_l}}$ are updated, we identify ***crowded*** reference points with niche values $\rho_{j_{S_t/F_l}} > 1$. After that, we add to the adaptive reference point near the ***crowded*** reference point. In the end, for a three-dimensional case, the last three reference points create the inner triangle around each crowded reference point, as shown in Figure 1c on the left side.

4. *Repeat previous tasks for all crowded reference points.*

Efficient Adaptive Reference Points

This method has the following steps:

1. *Generate efficient adaptive reference points for one division*: Initial efficient adaptive reference points $\boldsymbol{x_1} = (\frac{1}{\lambda * ndiv}, 0, \ldots, 0)$, $\boldsymbol{x_2} = (0, \frac{1}{\lambda * ndiv}, \ldots, 0)$, and $\boldsymbol{x_{nobj}} = (0, 0, \ldots, \frac{1}{\lambda * ndiv})$ are generated using the Das and Dennis approach (λ is a positive number called the scaling factor; its value is 1 for this work). They satisfy the plane equation $x_1 + x_2 + \cdots + x_{nobj} = \frac{1}{\lambda * ndiv}$

2. *Move efficient adaptive reference points to the plane* $x_1 + x_2 + \cdots + x_{nobj} = 0$: Initial efficient adaptive reference points are moved a distance $d = \frac{1}{\lambda * ndiv * nobj}$ in all dimensions.

3. *Subtract efficient adaptive reference points*: One of the previous efficient adaptive reference points is subtracted. In a three-dimensional case, this task is executed three-fold to obtain nine points (three of them are the same vector (0,0,0)).

4. *Move efficient adaptive reference points to the* ***crowded*** *reference point jth*: The final structure of reference points is moved next to the ***crowded*** reference point at a distance $d = \frac{1}{\lambda * ndiv * nobj}$ in all axes. Considering a three-dimensional example, the efficient adaptive reference points create three triangles connected to the crowded reference point as depicted in Figure 1c on the right side.

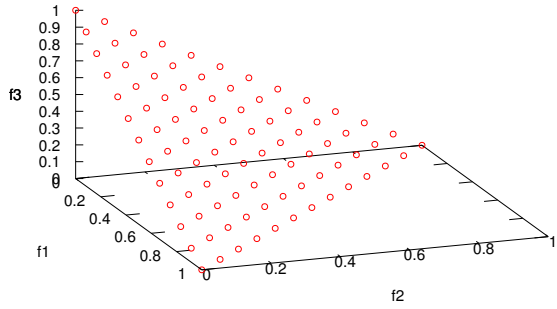
5. *Repeat for all crowded reference points.*

In next subsections, we show the procedures for the EF1-NSGA-III, except the associate one that is the same as the one from the NSGA-III authors.

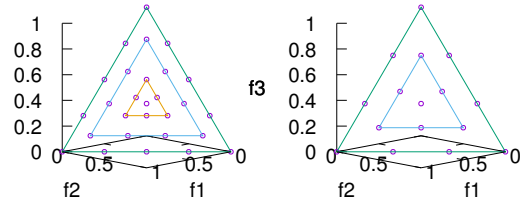
Normalization of Population Members

We normalized the objective values of population members to have the same range of the reference points as presented in Algorithm 2. The normalization procedure has the following steps:

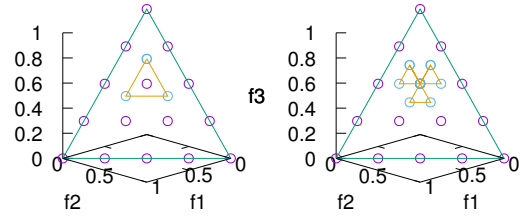
First, we construct the ideal vector $\bar{z} = \left\{z_1^{\min}, z_2^{\min}, z_j^{\min}, \ldots, z_M^{\min}\right\}$ of the population S_t which is determined by identifying the minimum value from each objective function. Second, we translate each objective $f_j(\boldsymbol{x})$ by subtracting $z_j^{\min}$, denoted as $f_j'(\boldsymbol{x}) = f_j(\boldsymbol{x}) - z_j^{\min}$. Third, we determine M extreme points to constitute a M-dimensional hyperplane that makes the Achievement Scalarization Function (***ASF***) minimum (Blank, Deb, and Roy, 2019), expressed as $z^{j,\max} = f'^n(\boldsymbol{x}) \mid \min_{n \in F_1}\{\boldsymbol{ASF}^j\}$ where $j \in \{1, 2, \ldots, M\}$, and n is the population size of the first front F_1. The $\boldsymbol{ASF}^j$ function finds the maximum translated objective value divided by the scalarization value in all dimensions, denoted as $\boldsymbol{ASF}^j = \max_{\min M}\{f_m'(\boldsymbol{x})/s_m^j\}$. For example, in a three-objective problem, we have the scalarization vector $\boldsymbol{s^1} = (1,\epsilon,\epsilon)$ for $j = 1$, $\boldsymbol{s^2} = (\epsilon, 1, \epsilon)$ for $j = 2$, $\boldsymbol{s^3} = (\epsilon, \epsilon, 1)$ for $j = 3$, where ϵ is a small value such as 0,0000000001.



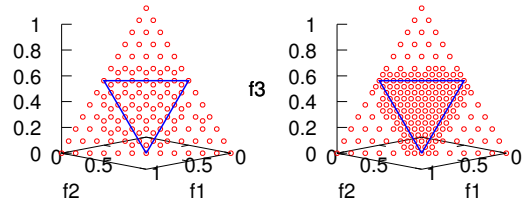
(a) Das and Dennis reference points distribution for $M = 3$ and $p = 12$.



(b) Left side, k-layer of reference points. Right side, two-layer of reference points.



(c) Left side, adaptive reference points. Right side, efficient adaptive reference points.



(d) Left-side, adaptive reference points. Right-side, efficient reference points.

Figure 1. Reference points approaches for NSGA-III, A-NSGA-III, and A^2-NSGA-III.
Source: Authors

Fourth, we find the matrix $\boldsymbol{z^{max-1}} = \left[[z^{1,\max}]; [z^{2,\max}]; [z^{3,\max}]; [\cdots]; [z^{M,\max}]\right]^{-1}$.

Finally, the objectives $f_j^n(\boldsymbol{x}) = \frac{f_j'(\boldsymbol{x})}{a_j}$ are normalized using $a_j = \sum_{i=1}^{M}$ as the intercept (elements of $\boldsymbol{z^{max-1}}$ j column). Eventually, repeated or negative intercepts can appear. Therefore, we use the $\boldsymbol{z^{max-1}} = \left[[z_1^{\max}]; [z_2^{\max}]; [z_3^{\max}]; [\cdots]; [z_M^{\max}]\right]^{-1}$ vector composed of the maximum

value of each function to generate, intercept, and then normalize the functions (the last generation positive intercept can be utilized as well). Recently, authors have revealed how to handle degenerate cases and negative intercepts (Blank et al., 2019) to avoid the algorithm getting stuck, but that solution remains as future research.

Algorithm 2 Normalize (f^n, S_t, F_1) procedure

Input: S_t, F_1
Output : f^n, Z^r (Das and Dennis reference points)
for $j = 1$ to M **do**
  Compute ideal point: $Z_j^{min} = \min_{S \in S_t} f_j(\boldsymbol{s})$
  Translate objectives: $f'_j(\boldsymbol{s}) = f_j(\boldsymbol{s}) - z_j^{min}\ \forall \boldsymbol{s} \in S_t$
  Compute extreme points: ($z^{j,max}$, $j = 1, \ldots, M$) of F_1
end for
Compute intercepts a_j for $j = 1, \ldots, M$
Normalize objectives f^n: $f_j^n(\boldsymbol{x}) = \frac{f'_i(\boldsymbol{x})}{a_i}$

Niche-preservation Operation

The niching procedure is presented in Algorithm 3. It identifies the reference point set $J_{\min_{S_t/F_l}} = j : argmin_{j \in Z^r} \rho_{S_t/F_l}$ having minimum ρ_{S_t/F_l} (Deb and Jain, 2014). If there are multiple associated reference points, one member of the set called $\bar{j} \in J_{\min_{S_t/F_l}}$ is chosen at random (the EF1-NSGA-III takes the first associated reference point, but any of them at random is valid).

For each generation, we neglect the reference point j which has population members associated with neither S_t/F_l nor F_l. Also, j is excluded if $\rho_{j_{F_l}} = 0$ and $\rho_{j_{S_t/F_l}} > 0$. The above conditions guarantee that the niching function runs K times to fill all P_{t+1} vacancies.

Algorithm 3 Niching ($K, \rho_{j_{S_t/F_l}}, \rho_{j_{F_l}}, \pi, d, Z^r, F_l$) procedure

Input: $K, \rho_{j_{S_t/F_l}}, \rho_{j_{F_l}}, \pi(\boldsymbol{s} \in S_T), d(\boldsymbol{s} \in S_T), Z^r, F_l, P_{t+1}$
Output : P_{t+1}
$k = 1$
while $k \leq K$ **do**
  $J_{\min_{\rho_{j_{S_t/F_l}}}} = \{j : argmin_{j \in Z^r} \rho_{j_{S_t/F_l}}\}$
  $\bar{j}$=Select the first member of ($J_{\min_{\rho_{j_{S_t/F_l}}}}$)
  $I_{\bar{j}} = \{\boldsymbol{s} : \pi = \bar{j}, \boldsymbol{s} \in F_l\}$
  if $I_{\bar{j}} \neq 0$ **then**
    if $\rho_{j_{S_t/F_l}} = 0 \&\& \rho_{j_{F_l}} = 0$ **then**
      $P_{t+1} = P_{t+1} \cup (\boldsymbol{s} : argmin_{\boldsymbol{s} \in I_{\bar{j}}} d(\boldsymbol{s}))$
    else
      $P_{t+1} = P_{t+1} \cup$ first member of ($I_{\bar{j}}$)
    end if
    $\rho_{\bar{j}} = \rho_{\bar{j}} + 1, F_l = F_l \backslash \boldsymbol{s}$
    $k = k + 1$
  else
    $Z^r = Z^r \backslash \{\bar{j}\}$
  end if
end while

After excluding some *useless* reference points, no matter how many S_t/F_l members are associated with the reference point $\bar{j}$, the one F_l member having the shortest perpendicular distance from the reference line $\bar{j}$ is added to P_{t+1}. Finally, the niche count $\rho_{\bar{j}_{S_t/F_l}}$ for reference point $\bar{j}$ is incremented by one and $\rho_{\bar{j}_{F_l}}$ is subtracted by one. Remember that, after niching, $\sum_{j=1}^{|H|} \rho_{j_{St/Fl}} = N$.

The next subsection explains the A-EF1-NSGA-III, and A^2-EF1-NSGA-III algorithms and their procedures to add and delete adaptive reference points.

Adaptive EF1-NSGA-III and Efficient Adaptive EF1-NSGA-III

Several MOOPs and MaOPs that are solved using the NSGA-III might have some reference points that will never be associated with the population. We address this fact through adaptive and efficient adaptive reference points that enhance the distribution of solutions.

We extended the A-EF1-NSGA-III and the A^2-EF1-NSGA-III - more details about their algorithms in Jain and Deb (2013). The last one has a higher adaptive reference point density, modulated by the scaling factor λ. It goes close to the Pareto-optimal front in smaller regions, allows adding reference points in the hyperplane corner, improves the deletion routine, and reduces the premature introduction of reference points in undesirable regions. The two adaptive EF1-NSGA-III algorithms define reference points with niche counts $\rho_{\bar{j}_{S_t/F_l}} > 0$ as ***useful***, and reference points with $\rho_{\bar{j}_{S_t/F_l}} = 0$ as ***useless***. The following procedures are added after the niche count of all reference points are updated:

Add adaptive or efficient adaptive reference points

This task generates new reference points based on the adaptive and efficient adaptive reference points studied before, which are plotted in Figure 1d at the end of the IDTLZ1 optimization process (the left-side corresponds to adaptive reference points and the right-side to efficient adaptive reference points). One of the NSGA-III termination criteria is to have all reference points associated with each population member ($\rho_{\bar{j}_{S_t/F_l}} = 1$), thus indicating a well-distributed Pareto front. Notwithstanding, sometimes that is impossible because the amount of reference points is higher than the population size, and some of them have niche counts $\rho_{\bar{j}_{S_t/F_l}} = 0$. They are relocated next to ***crowded*** reference points, and their capacity to capture a well-widespread Pareto front depends on the scaling factor λ and the settled down condition.

Delete adaptive or efficient adaptive reference points

This task deletes existing reference points. It is required to associate the added reference points to the population and update the niche value $\rho_{\bar{j}_{S_t/F_l}}$ (do not forget that $\sum_{j=1}^{|H|} \rho_{\bar{j}_{S_t/F_l}} = N$ for P_{t+1}). The deletion function is computationally more expensive than the adding one. In this case, before deleting *useless* reference points, we need to repeat the association and niching procedures and steady-state new reference points insertion routines.

The subsections below depict the constraint handling and the complexity discussion of EF1-NSGA-III algorithms.

Constraint Handling

We extended the constraint domination principle adopted in the NSGA-II algorithm for EF1-NSGA-III, A-EF1-NSGA-III, and A^2-EF1-NSGA-III. It is handled automatically by the initialization process, where all population members satisfy the bounds. Also, the creation of offspring solutions must be within lower and upper bounds. The evaluation of each individual considers inequality constraints, which are configured in the problem definition function at the beginning of the simulation. When the non-dominated sorting procedure starts, all individuals within the combined population $R_t = P_t \cup Q_t$ are categorized as feasible and infeasible. $\boldsymbol{x}^{(1)}$ is constraint-dominate $\boldsymbol{x}^{(2)}$ if one of the following conditions is true. First, if $\boldsymbol{x}^{(1)}$ is feasible and $\boldsymbol{x}^{(2)}$ is infeasible. Second, if $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$ are infeasible and $\boldsymbol{x}^{(1)}$ has the smaller constraint violation. Third, if $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$ are feasible and $\boldsymbol{x}^{(1)}$ dominates $\boldsymbol{x}^{(2)}$ with the usual domination principle.

There are four cases for the non-domination sort function in the tournament procedure to select individuals for the crossover function:

- First, all solutions are infeasible; solutions are sorted by minimum violation values and copied to the next generation.
- Second, the number of feasible individuals is lower than the population size; all feasible individuals are copied to the next generation.
- Third, the number of infeasible individuals equals twice the population size; the unconstrained non-domination sorting procedure is followed.
- Finally, the number of feasible individuals is higher or equal to the population size; the unconstrained non-domination sorting procedure is carried out with feasible individuals.

Complexity discussion

The NSGA-III and EF1-NSGA-III have a similar complexity because the EF1-NSGA-III inherits some of the NSGA-III procedures. The niche complexity requires $O(N)$ computations, and the rest is bounded by the maximum between $O(N^2 \log^{M-2} N)$ and $O(MN^2)$. However, if $M > N^2 \log^{M-2} N$, then he generation-wise complexity of the EF1-NSGA-III is $O(MN^2)$. The A-EF1-NSGA-II and A^2-EF1-NSGA-III have a higher complexity, thus requiring more computations for each generation. In those cases, the complexity of the niche function is $2 * O(N)$.

Performance evaluation

In this section, we provide the performance results of EF1-NSGA-III, A-EF1-NSGA-III, and A^2-EF1-NSGA-III to solve the DTLZ test (Deb, Thiele, and Laumanns, 2005) and real (water and car-side impact) problems handling constraints on three to ten objective functions. The employed hardware is a 3rd generation Intel core i7-3700x4 with 8 GBs of RAM, and the utilized software in Ubuntu 16.04, kernel 4.13.0-45-generic, and GCC version 5.4.0. We ran all the simulations using 20 realizations with the same random seeds for each problem.

Additionally, we took the population and the number of reference points for each DTLZ test problem reported in Table I of the work by Deb and Jain (2014) and in Table 1 of this work. For tuning purposes, we used the NSGA-III parameters employed by authors. They are shown in Table II of Deb and Jain (2014), as well as here in Table 2.

Table 1. Reference points and population sizes used for EF1-NSGA-III, A-EF1-NSGA-III, and A^2-EF1-NSGA-III algorithms. Nobj is the number of objective functions, and Boundary ndiv is the number of divisions of the boundary plane built with reference points. Within ndiv is the number of division of the inside plane built with reference points, and Popsize is the population size.

Nobj	Boundary ndiv	Inside ndiv	H	Popsize
3	12	0	91	92
5	6	0	210	212
8	3	2	156	156
10	3	2	275	276

Source: Authors

Table 2. Parameter values used in EF1-NSGA-III, A-EF1-NSGA-III, and A^2-EF1-NSGA-III algorithms.

Parameters	A/A^2/EF1-NSGA-III
Simulated binary crossover probability ρ_c	1
Polynomial mutation probability ρ_m	1/n
Distribution index for crossover η_c	30
Distribution index for mutation η_m	20

Source: Authors

Traditionally, the ZDT (Zitzler, Deb, and Thiele, 2000), DTLZ (Deb et al., 2005), WFG (Huband, Hingston, Barone, and While, 2006), and LZ/UF (Li and Zhang, 2009)(Zhang and Li, 2007) test suites are commonly-used benchmark multi-objective test problems. They are employed to validate MOEAs and MaOEAs under some difficulties, such as objective scalability, complicated Pareto sets, bias, disconnection, or degeneracy (Li et al., 2018). There are even more tests and real problems to evaluate evolutionary algorithms (Surjanovic and Bingham, 2013). However, we concentrated our effort to benchmark DTLZ test problems with different versions of the NSGA-III. Along with the proposed algorithms, we compared our results with those reported by authors and some publicly available open-source versions in terms of Inverted Generational Distance (IGD) and HiperVolume (HV) performance metrics (Chiang, 2014; Tian et al., 2017; Bi and Wang, 2017; Blank and Deb, 2020).

Normalized DTLZ1-4 Test Problems

The original NSGA-III implementation (Deb and Jain, 2014; Jain and Deb, 2014), and others found in references of this work (Ariza, 2019; Chiang, 2014; Tian et al., 2017; Blank and Deb, 2020) find well-distributed solutions when

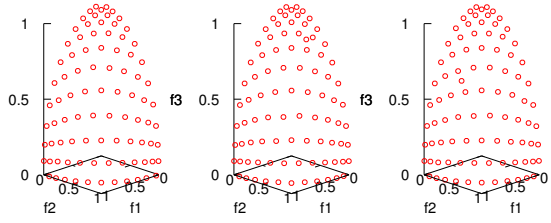
solving MaOPs. We noticed that there is a small difference between them regarding the IGD performance metric on three-objective to ten-objective normalized DTLZ1-4 test problems, as summarized in Table 3. It shows that the EF1-NSGA-III outperforms most scenarios, followed by the proposals by Deb and Jain (2014), Blank and Deb (2020), Chiang (2014), and Tian et al. (2017). Nevertheless, it has poor performance when solving normalized eight-objective DTLZ1 test problems and above.

Table 3. Best, Worst and Median IGD values for NSGA-III

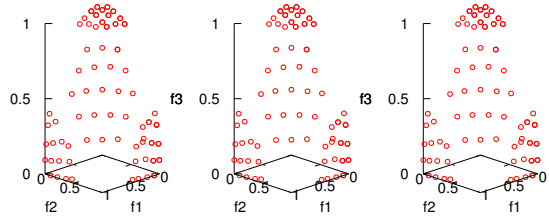
Problem	M	MaxGen	EF1-NSGA-III	Deb and J, 2014	Chiang, 2014	Tian, 2017	Blank, 2020
			$1{,}419x10^{-4}$	$4{,}880x10^{-4}$	$7{,}593x10^{-4}$	$2{,}0567x10^{-1}$	$4{,}86x10^{-4}$
DTLZ1	3	400	**$1{,}527x10^{-3}$**	$4{,}880x10^{-3}$	$2{,}053x10^{-1}$	$2{,}1505x10^{-1}$	$5{,}519x10^{-3}$
			$3{,}686x10^{-4}$	$1{,}308x10^{-3}$	$2{,}845x10^{-3}$	$2{,}0643x10^{-1}$	$1{,}429x10^{-3}$
			$1{,}077x10^{-4}$	$5{,}116x10^{-4}$	$7{,}718x10^{-4}$	$5{,}27x10^{-1}$	$5{,}23x10^{-4}$
DTLZ1	5	600	**$3{,}437x10^{-4}$**	$1{,}979x10^{-3}$	$4{,}733x10^{-1}$	$5{,}35x10^{-1}$	$2{,}513x10^{-3}$
			$1{,}361x10^{-4}$	$9{,}799x10^{-4}$	$2{,}05x10^{-3}$	$5{,}27x10^{-1}$	$1{,}121x10^{-3}$
			$5{,}223x10^{1}$	**$2{,}044x10^{-3}$**	$3{,}454x10^{-3}$	$9{,}6773x10^{-1}$	$2{,}34x10^{-3}$
DTLZ1	8	750	$5{,}876x10^{1}$	**$8{,}721x10^{-3}$**	$4{,}290x10^{-1}$	$2{,}5829x10^{-1}$	$1{,}0268x10^{-1}$
			$5{,}554x10^{1}$	**$3{,}979x10^{-3}$**	$5{,}279x10^{-3}$	$9{,}7509x10^{-1}$	$3{,}804x10^{-3}$
			$8{,}033x10^{1}$	**$2{,}215x10^{-3}$**	$2{,}257x10^{-3}$	$0{,}998x10^{-1}$	$2{,}419x10^{-3}$
DTLZ1	10	1 000	$9{,}744x10^{1}$	**$6{,}869x10^{-3}$**	$2{,}744x10^{-1}$	$1{,}541x10^{-1}$	$5{,}443x10^{-3}$
			$8{,}256x10^{1}$	**$3{,}462x10^{-3}$**	$4{,}235x10^{-3}$	$1{,}101x10^{-1}$	$3{,}098x10^{-3}$
			$9{,}993x10^{-4}$	$1{,}262x10^{-3}$	$1{,}585x10^{-3}$	$5{,}4468x10^{-1}$	$1{,}137x10^{-3}$
DTLZ2	3	250	$3{,}451x10^{-3}$	**$2{,}114x10^{-3}$**	$6{,}911x10^{-3}$	$5{,}4524x10^{-1}$	$2{,}212x10^{-3}$
			$1{,}591x10^{-3}$	**$1{,}357x10^{-3}$**	$2{,}731x10^{-3}$	$5{,}4485x10^{-1}$	$1{,}385x10^{-3}$
			$2{,}610x10^{-3}$	$4{,}254x10^{-3}$	$4{,}701x10^{-3}$	$1{,}612x10^{-1}$	$4{,}372x10^{-3}$
DTLZ2	5	350	$6{,}582x10^{-3}$	**$5{,}862x10^{-3}$**	$1{,}547x10^{-1}$	$1{,}6523x10^{-1}$	$5{,}948x10^{-3}$
			$2{,}890x10^{-3}$	$4.982x10^{-3}$	$6{,}299x10^{-3}$	$1{,}6517x10^{-1}$	$4{,}897x10^{-3}$
			$5{,}582x10^{-3}$	$1{,}371x10^{-1}$	$1{,}54x10^{-1}$	$3{,}0123x10^{-1}$	$1{,}368x10^{-1}$
DTLZ2	8	500	**$9{,}013x10^{-3}$**	$1{,}811x10^{-1}$	$2{,}002x10^{-1}$	$5{,}9472x10^{-1}$	$1{,}754x10^{-1}$
			$7{,}001x10^{-3}$	$1{,}571x10^{-1}$	$1{,}740x10^{-1}$	$3{,}1599x10^{-1}$	$1{,}502x10^{-1}$
			$5{,}297x10^{-3}$	$1{,}350x10^{-1}$	$1{,}384x10^{-1}$	$2{,}123x10^{-1}$	$1{,}194x10^{-1}$
DTLZ2	10	750	**$6{,}560x10^{-3}$**	$1{,}697x10^{-1}$	$2{,}004x10^{-1}$	$7{,}011x10^{-1}$	$1{,}813x10^{-1}$
			$5{,}444x10^{-3}$	$1{,}528x10^{-1}$	$1{,}727x10^{-1}$	$5{,}319x10^{-1}$	$1{,}528x10^{-1}$
			$5{,}189x10^{-4}$	$9.751x10^{-4}$	$1{,}374x10^{-3}$	$5{,}4495x10^{-1}$	$3{,}295x10^{-3}$
DTLZ3	3	1 000	**$8{,}251x10^{-3}$**	$6{,}665x10^{-3}$	$1{,}129x10^{-1}$	$5{,}4955x10^{-1}$	$1{,}379x10^{-1}$
			$1{,}927x10^{-3}$	$4{,}007x10^{-3}$	$4{,}757x10^{-3}$	$5{,}4627x10^{-1}$	$5{,}647x10^{-3}$
			$5{,}342x10^{-4}$	$3{,}086x10^{-3}$	$3{,}462x10^{-3}$	$1{,}659x10^{-1}$	$1{,}839x10^{-3}$
DTLZ3	5	1 000	**$7{,}759x10^{-3}$**	$1{,}196x10^{-1}$	$1{,}04x10^{-1}$	$3{,}656x10^{-1}$	$9{,}898x10^{-3}$
			$1{,}103x10^{-3}$	$5{,}960x10^{-3}$	$9{,}447x10^{-3}$	$1{,}722x10^{-1}$	$6{,}487x10^{-3}$
			$2{,}200x10^{-3}$	$1{,}244x10^{-1}$	$1{,}564x10^{-1}$	$3{,}1355x10^{-1}$	$1{,}668x10^{-1}$
DTLZ3	8	1 000	**$1{,}044x10^{-1}$**	$9{,}649x10^{-1}$	$1{,}386x10^{-1}$	$3{,}346x10^{-1}$	$8{,}799x10^{-1}$
			$3{,}243x10^{-3}$	$2{,}375x10^{-1}$	$3{,}014x10^{-1}$	$3{,}1965x10^{-1}$	$3{,}429x10^{-1}$
			$1{,}383x10^{-3}$	$8{,}849x10^{-3}$	$2{,}014x10^{-1}$	$2{,}224x10^{-1}$	$9{,}186x10^{-3}$
DTLZ3	10	1 500	**$2{,}255x10^{-3}$**	$2{,}083x10^{-1}$	$4{,}817x10^{-1}$	$7{,}897x10^{-1}$	$6{,}517x10^{-1}$
			$1{,}764x10^{-3}$	$1{,}188x10^{-1}$	$3{,}092x10^{-1}$	$5{,}3427x10^{-1}$	$1{,}221x10^{-1}$
			$4{,}059x10^{-4}$	$2{,}915x10^{-4}$	$2{,}447x10^{-4}$	$5{,}3578x10^{-1}$	**$2{,}01x10^{-4}$**
DTLZ4	3	600	$5{,}319x10^{-1}$	**$4{,}286x10^{-1}$**	$5{,}33x10^{-1}$	$5{,}279x10^{-1}$	$9{,}50x10^{-1}$
			$5{,}317x10^{-4}$	$5{,}970x10^{-4}$	$5.778x10^{-4}$	$5{,}452x10^{-1}$	**$3{,}30x10^{-4}$**
			$3{,}632x10^{-4}$	$9{,}849x10^{-4}$	$4{,}305x10^{-4}$	$1{,}64x10^{-1}$	$4{,}08x10^{-4}$
DTLZ4	5	1 000	$5{,}319x10^{-1}$	$1{,}721x10^{-3}$	$1{,}108x10^{-3}$	$1{,}659x10^{-1}$	**$1{,}595x10^{-3}$**
			$6{,}182x10^{-4}$	$1{,}255x10^{-3}$	$7{,}294x10^{-4}$	$1{,}651x10^{-1}$	**$5{,}65x10^{-4}$**
			$2{,}200x10^{-3}$	$5{,}079x10^{-3}$	$3{,}406x10^{-3}$	$3{,}151x10^{-1}$	$3{,}334x10^{-3}$
DTLZ4	8	1 250	$1{,}044x10^{-1}$	$6{,}051x10^{-1}$	$5{,}496x10^{-3}$	$6{,}5476x10^{-1}$	**$5{,}476x10^{-3}$**
			$3{,}243x10^{-3}$	$7{,}054x10^{-3}$	$4{,}345x10^{-3}$	$3{,}1532x10^{-1}$	$4{,}046x10^{-3}$
			$1{,}642x10^{-3}$	$5{,}694x10^{-3}$	$3{,}633x10^{-3}$	$4{,}19x10^{-1}$	$3{,}626x10^{-3}$
DTLZ4	10	2 000	**$2{,}254x10^{-3}$**	$1{,}076x10^{-1}$	$5{,}301x10^{-3}$	$4{,}761x10^{-1}$	$5.601x10^{-3}$
			$1{,}764x10^{-3}$	$6{,}337x10^{-3}$	$4{,}601x10^{-3}$	$4{,}401x10^{-1}$	$4{,}413x10^{-3}$

Source: Authors

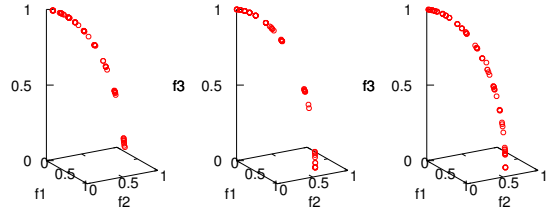
We visualized some three-objective problems solved with EF1-NSGA-III, A-EF1-NSGA-III, and A^2-EF1-NSGA-III in Figure 2. They are DTLZ2 (Figure 2a), C2-DTLZ2 (Figure 2b), DTLZ5 (Figure 2c), DTLZ7 (Figure 2d), DTLZ1 (Figure 2e), DTLZ4 (Figure 2f), IDTLZ1 (Figure 2g), and Car-Side Impact (Figure 2h) problems, respectively. At first glance, The EF1-NSGA-III seems to achieve well-distributed and well-spread solutions for those problems. Also, the A^2-EF1-NSGA-III performs better than the A-EF1-NSGA-III, and the latter is better than the EF1-NSGA-III for some DTLZ test problems with difficulties, for instance, in normalized DTLZ5, DTLZ7, and IDTLZ1 test problems.



(a) 3D DTLZ2 test problem.

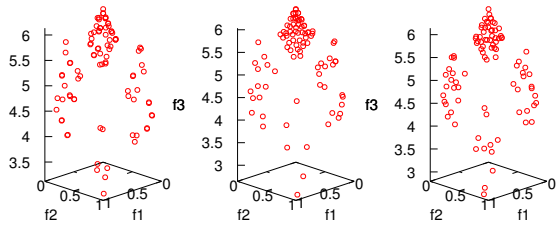


(b) 3D C2-DTLZ2 test problem.

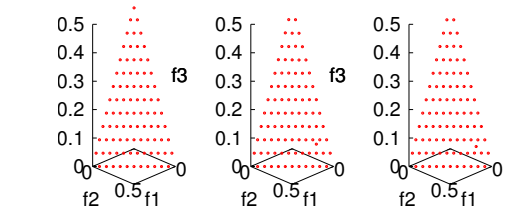


(c) 3D DTLZ5 test problem.

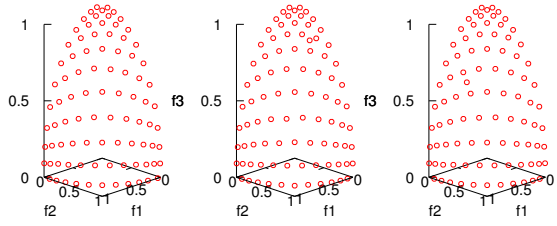
Figure 2. From left to right: Using the EF1-NSGA-III, EF1-A-NSGA-III and EF1-A^2-NSGA-III to solve three dimensional problems.
Source: Authors



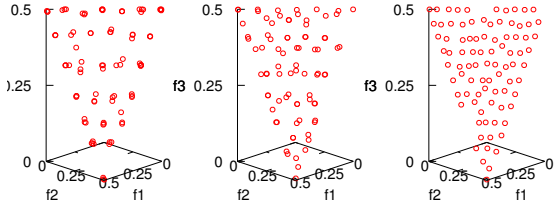
(d) 3D DTLZ7 test problem.



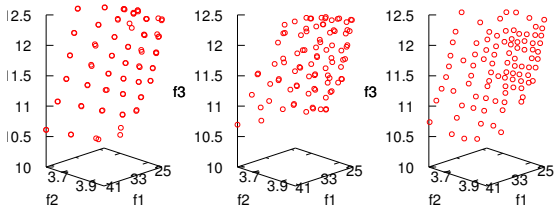
(e) 3D DTLZ1 test problem.



(f) 3D DTLZ4 test problem.



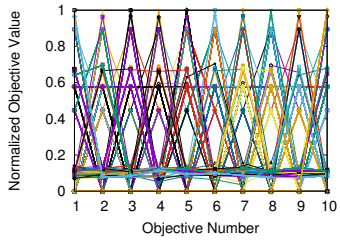
(g) 3D IDTLZ1 test problem.



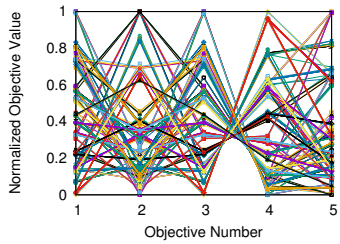
(h) 3D Car-Side Impact problem.

Figure 2. Continuation

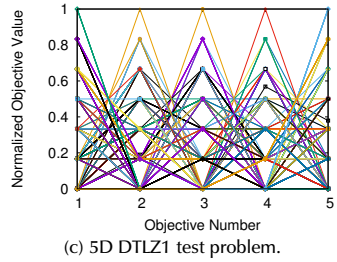
High dimensional problems are depicted in Figure 3 as well. In this case, we employed parallel coordinates plots to visualize the solutions for ten-objective DTLZ4 (Figure 3a), five-objective Water (Figure 3b), five-objective DTLZ1 (Figure 3c), and five-objective DTLZ2 (Figure 3d) problems, respectively.



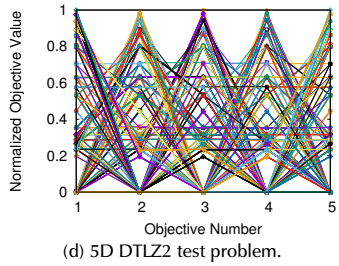
(a) 10D DTLZ4 test problem.



(b) 5D Water problem.



(c) 5D DTLZ1 test problem.



(d) 5D DTLZ2 test problem.

Figure 3. EF1-NSGA-III solving High dimensional problems.
Source: Authors

Normalized DTLZ5 and DTLZ7 Test Problems

DTLZ5 and DTLZ7 test problems have irregular Pareto fronts that reveal the diversity maintenance of reference point adaption.

The DTLZ5 is a degenerate test problem whose Pareto Front lies on a dimension curve independent of the number of objectives. On the contrary, the DTLZ7 problem examines the algorithm's ability to sustain sub-population in different Pareto-optimal regions. Figures 2c and 2d depict the Pareto fronts of three-objective DTLZ5 and DTLZ7 problems. Other algorithms, such as Adaw (Li and Yao, 2020), have a consistently better distribution of solutions when solving both problems mentioned before. We observed that the EF1-NSGA-II is slightly better than other proposals in terms of the IGD metric, as shown in Table 4.

Table 4. Average and Standard Deviation IGD values for NSGA-III

Problem	M	MaxGen	EF1-NSGA-III	Tian, 2017	Bi, 2017	Blank, 2020
DTLZ5	3	1 000	6,6588x10^{-2}	1,195x10^{-2}	1,438x10^{-2}	**1,468x10^{-2}**
			4,114x10^{-2}	3,07x10^{-3}	2,19x10^{-3}	**1,589x10^{-3}**
DTLZ5	5	1 000	1,585x10^{-1}	1,515x10^{-1}	**9,654x10^{-2}**	9,912x10^{-2}
			4,67x10^{-2}	6,582x10^{-2}	3,60x10^{-2}	**2,35x10^{-2}**
DTLZ5	8	1 000	**3,007x10^{-1}**	4,1928x10^{-1}	3,473x10^{-1}	3,712x10^{-1}
			4,376x10^{-2}	5,07x10^{-1}	9,93x10^{-2}	9,042x10^{-2}
DTLZ5	10	1 500	**1,902x10^{-1}**	2,487x10^{-1}	4,548x10^{-1}	5,114x10^{-1}
			3,932x10^{-2}	5,96x10^{-2}	8,49x10^{-2}	7,712x10^{-2}
DTLZ7	3	1 000	1,108x10^{-1}	9,063x10^{-2}	**7,794x10^{-2}**	1,308x10^{-1}
			6,186x10^{-2}	6,64x10^{-2}	**1,59x10^{-3}**	1,151x10^{-1}
DTLZ7	5	1 000	2,955x10^{-1}	**2,782x10^{-1}**	2,853x10^{-1}	2,955x10^{-1}
			1,227x10^{-2}	1,64x10^{-2}	5,83x10^{-3}	**5,642x10^{-3}**
DTLZ7	8	1 000	**5,679x10^{-1}**	7,746x10^{-1}	7,441x10^{-1}	6,242x10^{-1}
			3,182x10^{-2}	5,36x10^{-2}	**1,68x10^{-2}**	2,994x10^{-2}
DTLZ7	10	1 500	**6,525x10^{-1}**	9,250x10^{-1}	9,850x10^{-1}	7,579x10^{-1}
			3,218x10^{-2}	**2,40x10^{-2}**	1,00x10^{-1}	4,293x10^{-2}

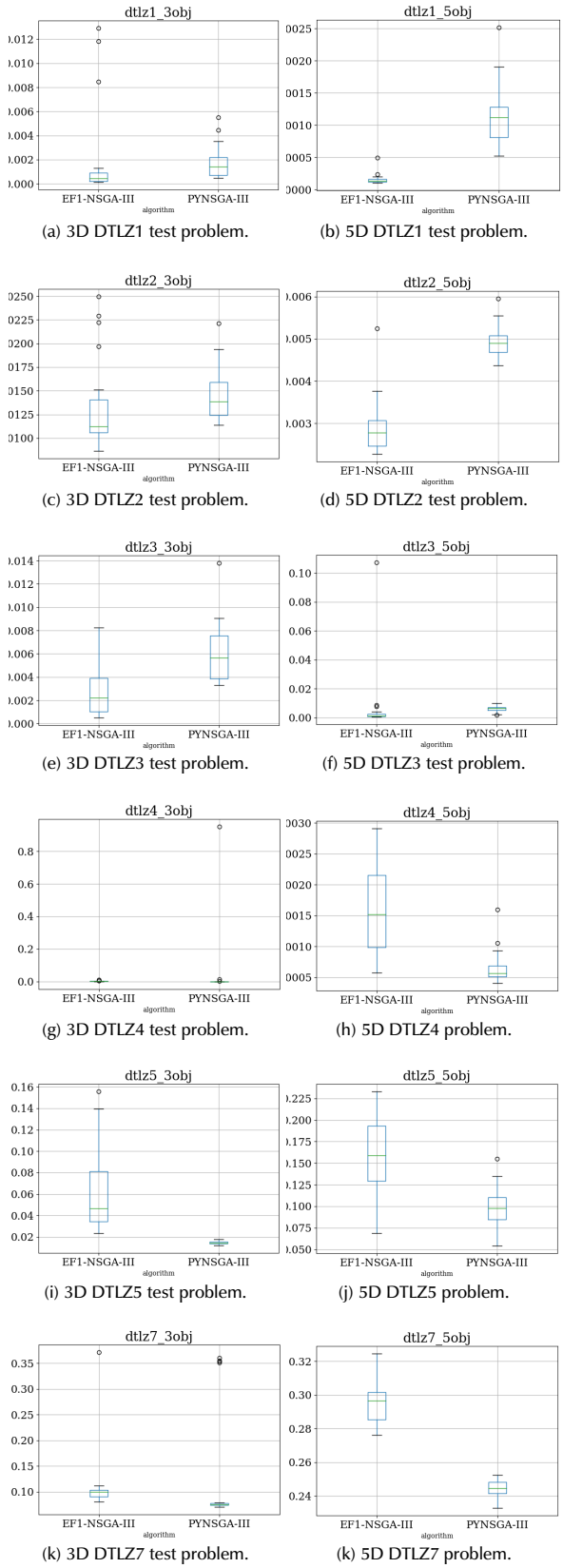
Source: Authors

DTLZ1-5 and DTLZ7 Statistical benchmark

We made a statistical benchmark between EF1-NSGA-III and PYNSGA-III (Pymoo NSGA-III algorithm) in terms of IGD and HyperVolume (HV) performance metrics for DTLZ1-5 and DTLZ7 problems. Their statistical boxplots are shown in Figures 4 and 5 for three-objective and five-objective DTLZ problems running 20 realizations with the same random seeds. First, from an IGD perspective, we see in Figures 4a, 4b, 4c, 4d, 4e, and 4f, that the EF1-NSGA-III performs better than the PYNSGA-III and provides lower IGD values and dispersion. Notwithstanding, Figures 4g, 4i, 4j, 4k, 4l, and 4h depict a different scenario, where the PYNSGA-III performs better than the EF1-NSGA-III.

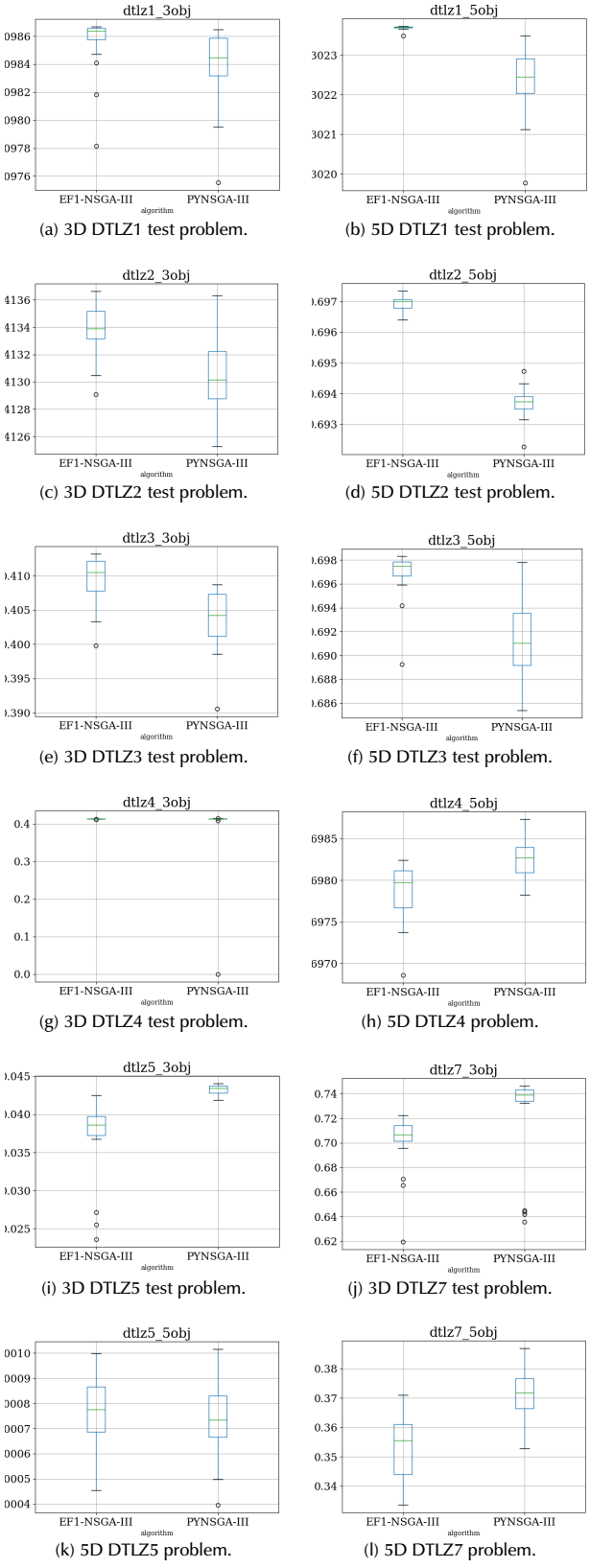
Then, when we analyzed the boxplots of Figure 5 related to the HV performance metric, we notice the same behavior obtained concerning the IGD performance metric. In this scenario, high HV values mean a better performance of convergence and distribution of solutions. That validation helps us be confident of the performance of the EF1-NSGA-III with respect to the PYNSGA-III. For example, Figures 5a, 5b, 5c, 5d, 5e, and 5f, accomplish a better performance in terms of HV for the EF1-NSGA-III rather than the PYNSGA-III. Hovewer, the PYNSGA-III performs better than the EF1-NSGA-III for DTLZ4, DTLZ5 and DTLZ7, as seen in Figures 5g, 5h, 5i, 5j, 5k and 5l.

Exceptionally, the EF1-NSGA-III strives to solve the DTLZ4 test problem for five objectives and above, and those drawbacks are represented with a significant data dispersion, as shown in Figures 4h and 5h. This kind of convergence difficulty occurs for some seeds, so the idea is to increase the number of realizations to reduce this effect.



(a) 3D DTLZ1 test problem. (b) 5D DTLZ1 test problem.
(c) 3D DTLZ2 test problem. (d) 5D DTLZ2 test problem.
(e) 3D DTLZ3 test problem. (f) 5D DTLZ3 test problem.
(g) 3D DTLZ4 test problem. (h) 5D DTLZ4 problem.
(i) 3D DTLZ5 test problem. (j) 5D DTLZ5 problem.
(k) 3D DTLZ7 test problem. (k) 5D DTLZ7 problem.

Figure 4. Boxplots in terms of IGD.
Source: Authors



(a) 3D DTLZ1 test problem. (b) 5D DTLZ1 test problem.
(c) 3D DTLZ2 test problem. (d) 5D DTLZ2 test problem.
(e) 3D DTLZ3 test problem. (f) 5D DTLZ3 test problem.
(g) 3D DTLZ4 test problem. (h) 5D DTLZ4 problem.
(i) 3D DTLZ5 test problem. (j) 3D DTLZ7 test problem.
(k) 5D DTLZ5 problem. (l) 5D DTLZ7 problem.

Figure 5. Boxplots in terms of HV.
Source: Authors

Inverted-DTLZ1 Test Problem

The inverted-DTLZ1 problem is an unconstrained test problem that modifies the DTLZ1 problem. The DTLZ1 transformation requires $f_i(\boldsymbol{x}) = 0,5(1 + g(\boldsymbol{x})) - f_i(\boldsymbol{x})$, for $i = 1,\ldots,M$ objectives. We can see that the A^2-EF1-NSGA-III distributes the population better than the A-EF1-NSGA-III and the EF1-NSGA-III. The Hyper-volume values are shown in Table 5 and visually confirmed in Figure 2g. We employed the same algorithm from reference (Blank and Deb, 2020) to evaluate the HV. It takes the obtained Pareto front from EF1-NSGA-III and PYNSGA-III and a reference point. We used the point [0,6, 0,6, 0,6] for the three-objective IDTLZ1 problem, the point [39,4, 12,5] for the Car-side impact problem, and the point [75000, 1400, 2900000, 6700000, 26000] for the water problem to obtain HV values. Sometimes, for instance, the Car-side impact problem solved with the EF1-NSGA-III has better HV values than the A^2-EF1-NSGA-III and the A-EF1-NSGA-III. It remains as future work where the number of realizations should be increased.

Table 5. Best, Worst and Median Hyper-volume values obtained for A^2-NSGA-III, A-NSGA-III and NSGA-III

Problem	M	MaxGen	A^2-NSGA-III	A-NSGA-III	NSGA-III
Car-Side Impact (EF1-NSGA-III)	3	500	3,7771	2,6168	**4,4349**
			4,8367	4,6992	4,6179
			4,1767	3,9892	**4,5316**
Water (EF1-NSGA-III)	5	500	**2,9195x10^{24}**	2,7388x10^{24}	2,7656x10^{24}
			3,0128x10^{24}	2,895x10^{24}	2,901x10^{24}
			2,967x10^{24}	2,847x10^{24}	2,8444x10^{24}
IDTLZ1 (EF1-NSGA-III)	3	400	**5,905x10^{-2}**	5,799x10^{-2}	5,782x10^{-2}
			6,191x10^{-2}	6,182x10^{-2}	6,189x10^{-2}
			6,032x10^{-2}	6,037x10^{-2}	5,987x10^{-2}
IDTLZ1(Deb and Jain, 2014)	3	400	**6,374x10^{-2}**	6,094x10^{-2}	6,111x10^{-2}
			6,621x10^{-2}	6,54x10^{-2}	6,305x10^{-2}
			6,569x10^{-2}	6,540x10^{-2}	6,229x10^{-2}
IDTLZ1 (Tian et al., 2017)	3	400	N/A	2,191x10^{-1}	2,063x10^{-1}
			N/A	2,127x10^{-1}	1,978x10^{-1}
			N/A	2,153x10^{-1}	2,018x10^{-1}
Water (Jain and Deb, 2013)	5	500	**0,545**	0,540	0,534
			0,540	0,5349	0,528
			0,543	0,5365	0,531

Source: Authors

Normalized constrained DTLZ test problems

We considered three cases of constrained DTLZ test problems. The first is the C1-DTLZ1 problem that uses the constraint $c(\boldsymbol{x}) = 1 - \frac{f_M(\boldsymbol{x})}{0,6} - \sum_{i=1}^{M-1} \frac{f_i(\boldsymbol{x})}{0,5} \geq 0$. The second is the C1-DTLZ3 problem that employs the constraint $c(\boldsymbol{x}) = (\sum_{i=1}^{M} f_i(\boldsymbol{x})^2 - 16)(\sum_{i=1}^{M} f_i(\boldsymbol{x})^2 - r^2)) \geq 0$, where $r = \{9, 12, 5, 12, 5, 15, 15\}$ is the radius of the hyper-sphere for $M = \{3, 5, 8, 10, 15\}$. Finally, the C2-DTLZ2 problem that utilizes the constraint $c(\boldsymbol{x}) = -\min\{\min_{i=1}^{M}[(f_i(\boldsymbol{x}) - 1)^2 + \sum_{j=1, j\neq i}^{M} f_j^2 - r^2], [\sum_{i=1}^{M}(f_i(\boldsymbol{x}) - \frac{1}{\sqrt{M}})^2 - r^2)]\} \geq 0$, where $r = 0,4$ for $M = 3$ and $0,5$ otherwise.

Table 6 shows a consistently better performance of the EF1-NSGA-III in terms of IGD for the studied constrained problem, except for C1-DTLZ1 on eight-objective to ten-objective problems, where the EF1-NSGA-III gets stuck. The approximated Pareto-front of C2-DTLZ2 is depicted in Figure 2b that shows a good distribution of solutions.

Table 6. Best, Worst and Median IGD values obtained for NSGA-III. Constrained DTLZ problems

Problem	M	MaxGen	U/H	EF1-NSGA-III	Deb and J, 2014	Tian, 2017
C1-DTLZ1	3	500	91/91	**2,996x10^{-4}**	1,229x10^{-3}	2,002x10^{-2}
				1,389x10^{-2}	2,256x10^{-2}	2,139x10^{-2}
				2,948x10^{-3}	4,932x10^{-3}	2,026x10^{-2}
C1-DTLZ1	5	600	210/210	**1,393x10^{-4}**	2,38x10^{-3}	5,092x10^{-2}
				2,215x10^{-3}	1,024x10^{-2}	5,259x10^{-2}
				1,165x10^{-4}	4,347x10^{-3}	5,203x10^{-2}
C1-DTLZ1	8	800	156/156	7,010x10^{-1}	**4,843x10^{-3}**	9,006x10^{-2}
				7,072x10^{-1}	**4,140x10^{-2}**	9,684x10^{-2}
				7,037x10^{-1}	**1,361x10^{-2}**	9,548x10^{-2}
C1-DTLZ1	10	1 000	275/275	7,300x10^{-1}	**3,042x10^{-3}**	1,078x10^{-1}
				7,3323x10^{-1}	**2,762x10^{-2}**	1,095x10^{-1}
				7,315x10^{-1}	**6,358x10^{-3}**	1,086x10^{-1}
C2-DTLZ2	3	250	58/91	**4,158x10^{-4}**	1,581x10^{-3}	4,703x10^{-2}
				1,780x10^{-3}	6,733x10^{-3}	4,866x10^{-2}
				6,631x10^{-4}	2,578x10^{-3}	4,823x10^{-2}
C2-DTLZ2	5	350	80/210	**6,955x10^{-4}**	2,762x10^{-3}	1,382x10^{-1}
				1,0801x10^{-3}	7,596x10^{-3}	1,397x10^{-1}
				9,173x10^{-4}	3,873x10^{-3}	1,386x10^{-1}
C2-DTLZ2	8	500	72/156	**1,066x10^{-3}**	1,404x10^{-2}	2,363x10^{-1}
				2,591x10^{-3}	8,662x10^{-1}	8,596x10^{-1}
				1,410x10^{-3}	2,352x10^{-2}	2,406x10^{-1}
C2-DTLZ2	10	750	110/275	**4,929x10^{-4}**	1,978x10^{-2}	2,648x10^{-1}
				8,234x10^{-4}	3,491x10^{-2}	5,118x10^{-1}
				5,894x10^{-4}	2,694x10^{-2}	2,716x10^{-1}
C1-DTLZ3	3	1 000	91/91	**5,189x10^{-4}**	8,649x10^{-4}	N/A
				8,251x10^{-3}	(13)	N/A
				2,184x10^{-3}	8,139x10^{-3}	N/A
C1-DTLZ3	5	1 500	210/210	**2,189x10^{-4}**	1,028x10^{-3}	N/A
				2,384x10^{-3}	(15)	N/A
				5,260x10^{-4}	5,101x10^{-2}	N/A
C1-DTLZ3	8	2 500	156/156	**3,002x10^{-4}**	1,656x10^{-3}	N/A
				2,257x10^{-4}	(14)	N/A
				4,538x10^{-4}	0,196x10^{-2}	N/A
C1-DTLZ3	10	3 500	275/275	**3,347x10^{-4}**	2,437x10^{-3}	N/A
				7,117x10^{-4}	(18)	N/A
				3,747x10^{-4}	1,445x10^{-2}	N/A

Source: Authors

Car-Side Impact and Water Problems

The Car-Side Impact problem optimizes the frontal structure of the vehicle for crash-worthiness. It minimizes the weight, the pubic force experienced by a passenger, and the V-Pillar's average velocity to resist the impact load. Its mathematical formulation can be consulted in Jain and Deb (2014). We can see, in Figure 2h, that the A^2-EF1-NSGA-III has a better distribution of solutions compared to the A-EF1-NSGA-III and EF1-NSGA-III. Nevertheless, in Table 5, we notice that the EF1-NSGA-III achieves better HV values than the A^2-EF1-NSGA-III or the A-EF1-NSGA-III. It is necessary to use statistical analysis with more realizations to better understand the distribution of solutions.

Conversely, as shown in Table 5 and Figure 3b, the Water problem has better performance when it is solved with the A^2-EF1-NSGA-III, A-EF1-NSGA-III, and the EF1-NSGA-III, which is expected. This problem has five objective functions, three design variables, and seven constraints. Its mathematical formulation is found in the work by (Jain and Deb, 2014).

Variation of convergence metrics

We used IGD and Υ values in each generation to visualize the convergence. IGD convergence is depicted in Figure 6 and the Υ convergence in Figure 7. The IGD convergence plot is divided into two cases. The first (Figure 6 on the left side) contains better IGD decreasing performance for DTLZ1, DTLZ3, DTLZ4, DTLZ5, and DTLZ7 problems. However, the DTLZ1 is the fastest. In contrast, the second (Figure 6, on the right side) shows bad decreasing IGD performance for IDTLZ1, DTLZ5, and DTLZ7 problems. They will not achieve good performance even though we increase the number of generations. High dimensional DTLZ3 and DTLZ4 problems probably will achieve a good IGD value, but they require significantly more generations. We also divided the Υ convergence plot in two cases. The first (Figure 7, top side) for low dimensional cases, and the second, for high dimensional cases (Figure 7, bottom side). We noticed that this metric is representative for problems up to eight objectives. We also saw that the DTLZ1, DTLZ2, DTLZ3, and DTLZ4 problems exhibit better Υ convergence, but DTLZ3 and DTLZ4 problems require more generations to accomplish similar Υ values. Finally, high dimensional test problems did not provide good Υ values, as shown in Figure 7, bottom part.

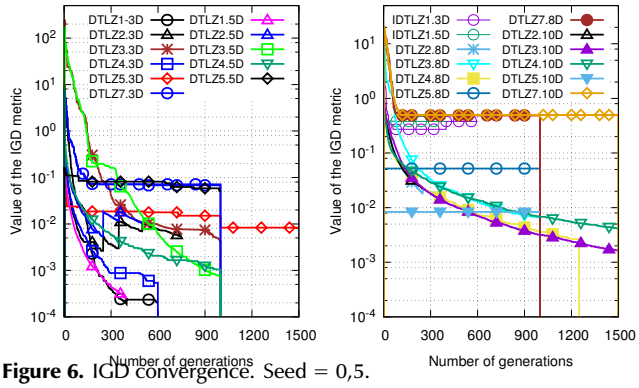


Figure 6. IGD convergence. Seed = 0,5.
Source: Authors

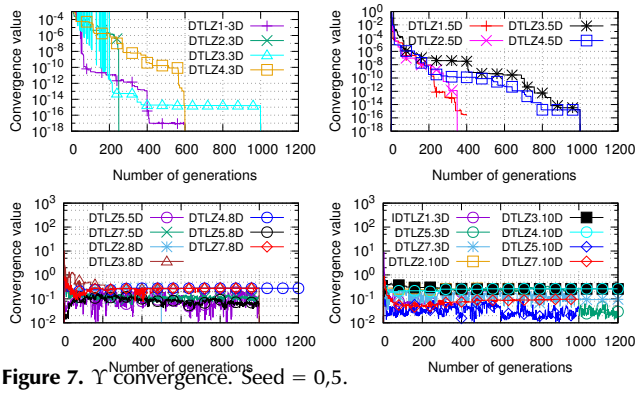


Figure 7. Υ convergence. Seed = 0,5.
Source: Authors

Conclusions

We successfully extended the NSGA-II algorithm from the KanGAL's Code to the EF1-NSGA-III, the A-EF1-NSGA-III, and the A^2-EF1-NSGA-III. These algorithms have been applied to solve constrained and unconstrained Many-objective optimization problems (up to ten-objective problems). Our proposals have shown their niche in finding a set of well-converged and well-diversified solutions repeatedly over multiple runs. They are susceptible to be employed as a reference for a comparative evaluation of new evolutionary algorithms because they yield consistently better results than those reported in references (Deb and Jain, 2014; Jain and Deb, 2014; Bi and Wang, 2017; Tian et al., 2017; Chiang, 2014; Blank and Deb, 2020).

References

Ariza Vesga, L. F. (2019). *A fast non-dominated sorting genetic algorithm extension to solve many-objective problems*. https://github.com/lfarizav/NSGA-III

Ariza Vesga, L. F. (2020). *Many-objective problems optimization focused on energy efficiency applied to 5G heterogeneous cellular networks using the small cell switch off framework* (Doctoral thesis, Universidad Nacional de Colombia, Bogotá, Colombia). https://repositorio.unal.edu.co/handle/unal/77582

Bader, J. and Zitzler, E. (2011). HypE: An Algorithm for Fast Hypervolume-Based Many-Objective Optimization. *Evolutionary Computation, 19*(1), 45-76. 10.1162/evco_a_00009

Bi, X. and Wang, C. (2017). An improved NSGA-III algorithm based on elimination operator for many-objective optimization. *Memetic Computing, 9*(4), 361-383. 10.1007/s12293-017-0240-7

Blank, J. and Deb, K. (2020). Pymoo: Multi-Objective Optimization in Python. *IEEE Access, 8*, 89497-89509. 10.1109/access.2020.2990567

Blank, J., Deb, K., and Roy, P. C. (2019). Investigating the Normalization Procedure of NSGA-III. In Deb, K., Goodman, E., Coello-Coello, C. A., Klamroth, K., Miettinen, K., Mostaghim, S., and Reed, P. (Eds.) *Evolutionary Multi-Criterion Optimization* (vol. 11411, pp. 229-240). New York, NY: Springer, Cham 10.1007/978-3-030-12598-1_19

Chiang, T.-C. (2014). *nsga3cpp: A c++ implementation of nsga-iii*. https://web.ntnu.edu.tw/~tcchiang/publications/nsga3cpp/nsga3cpp.htm

Das, I., and Dennis, J. E. (1998). Normal-Boundary Intersection: A New Method for Generating the Pareto Surface in Nonlinear Multicriteria Optimization Problems. *SIAM Journal on Optimization, 8*(3), 631-657. 10.1137/s1052623496307510

Deb, K. (1999). An introduction to genetic algorithms. *Sadhana, 24*, 293315. 10.1007/BF02823145

Deb, K. and Jain, H. (2014). An Evolutionary Many-Objective Optimization Algorithm Using Reference-Point-Based Nondominated Sorting Approach, Part I: Solving Problems With Box Constraints. *IEEE Transactions on Evolutionary Computation, 18*(4), 577-601. 10.1109/tevc.2013.2281535

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation, 6*(2), 182-197. 10.1109/4235.996017

Deb, K., Thiele, L., Laumanns, M., and Zitzler, E. (2005). Scalable Test Problems for Evolutionary Multiobjective Optimization. In Abraham, A., Jain, L., and Goldberg, R. (Eds.) *Evolutionary Multiobjective Optimization* (pp. 105-145) New York, NY: Springer, Cham. 10.1007/1-84628-137-7_6

Fonseca, C., Paquete, L., and Lopez-Ibanez, M. (2006). *An Improved Dimension-Sweep Algorithm for the Hypervolume Indicator*. Paper presented at the IEEE International Conference on Evolutionary Computation, Vancouver, BC, Canada. 10.1109/cec.2006.1688440

Fraser, A. (1957). Simulation of Genetic Systems by Automatic Digital Computers II. Effects of Linkage on Rates of Advance Under Selection. *Australian Journal of Biological Sciences, 10*(4), 492. 10.1071/bi9570492

Gu, Z. and Wang, G. (2020). Improving NSGA-III algorithms with information feedback models for large-scale manyobjective optimization. *Future Generation Computer Systems, 107*, 49-69. 10.1016/j.future.2020.01.048

Gupta, R., Nanda, S. J. (2019). *Many-Objective B/NSGA-III for Band Selection in Cloud Contaminated Hyper-Spectral Images*. Paper presented at the 2019 International Conference on Information Technology (ICIT), Bhubaneswar, India. 10.1109/icit48102.2019.00068

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. 10.7551/mitpress/1090.001.0001

Huband, S., Hingston, P., Barone, L., and While, L. (2006). A review of multiobjective test problems and a scalable test problem toolkit. *IEEE Transactions on Evolutionary Computation, 10*(5), 477-506. 10.1109/tevc.2005.861417

Ibrahim, A., Rahnamayan, S., Martin, M. V., and Deb, K. (2016). *EliteNSGA-III: An improved evolutionary many-objective optimization algorithm*. Paeper presented at the 2016 IEEE Congress on Evolutionary Computation (CEC). 10.1109/cec.2016.7743895

Ishibuchi, H., Imada, R., Setoguchi, Y., and Nojima, Y. (2016). *Performance comparison of NSGA-II and NSGA- III on various many-objective test problems*. Paper presented at 2016 IEEE Congress on Evolutionary Computation (CEC). 10.1109/cec.2016.7744174

Jain, H. and Deb, K. (2013). An Improved Adaptive Approach for Elitist Nondominated Sorting Genetic Algorithm for Many-Objective Optimization. In Purshouse, R. C., Fleming, P. J., Fonseca, C. M., Greco S., and Shaw, J. (Eds.) *Evolutionary Multi-Criterion Optimization* (vol. 7811, pp. 307-321). New York, NY: Springer, Cham. 10.1007/978-3-642-37140-0_25

Jain, H. and Deb, K. (2014). An Evolutionary Many-Objective Optimization Algorithm Using Reference-Point Based Nondominated Sorting Approach, Part II: Handling Constraints and Extending to an Adaptive Approach. *IEEE Transactions on Evolutionary Computation, 18*(4), 602-622. 10.1109/tevc.2013.2281534

Jiang, S. and Yang, S. (2017). A Strength Pareto Evolutionary Algorithm Based on Reference Direction for Multiobjective and Many-Objective Optimization. *IEEE Transactions on Evolutionary Computation, 21*(3), 329-346. 10.1109/tevc.2016.2592479

jMetal (2015). *Metaheuristic algorithms in java*. http://jmetal.sourceforge.net

KanGAL (2011). *Software Developed at KanGAL*. http://www.iitk.ac.in/kangal/codes.shtml

Li, K., Wang, R., Zhang, T., and Ishibuchi, H. (2018). Evolutionary Many-Objective Optimization: A Comparative Study of the State-of-the-Art. *IEEE Access, 6*, 26194-26214. 10.1109/access.2018.2832181

Li, H. and Zhang, Q. (2009). Multiobjective Optimization Problems With Complicated Pareto Sets, MOEA/D and NSGA-II. *IEEE Transactions on Evolutionary Computation, 13*(2), 284-302. 10.1109/tevc.2008.925798

Li, M. and Yao, X. (2020). What Weights Work for You? Adapting Weights for Any Pareto Front Shape in Decomposition-Based Evolutionary Multiobjective Optimisation. *Evolutionary Computation, 28*(2), 227-253. 10.1162/evco_a_00269

Li, W., Wang, G., and Alavi, A. H. (2020). Learning-based elephant herding optimization algorithm for solving numerical optimization problems. *Knowledge-Based Systems, 195*, 105675. 10.1016/j.knosys.2020.105675

Li, H., Deb, K., Zhang, Q., Suganthan, P., and Chen, L. (2019). Comparison between MOEA/D and NSGA-III on a set of novel many and multi-objective benchmark problems with challenging difficulties. *Swarm and Evolutionary Computation, 46*, 104-117. 10.1016/j.swevo.2019.02.003

Marti, L. (2016). *An implementation of nsga-iii in python*. https://github.com/lmarti/nsgaiii

Prakash, V. P., Patvardhan, C., and Srivastav, A. (2020). A novel Hybrid Multi-objective Evolutionary Algorithm for the bi-Objective Minimum Diameter-Cost Spanning Tree (bi-MDCST) problem. *Engineering Applications of Artificial Intelligence, 87*, 103237. 10.1016/j.engappai.2019.103237

Purshouse, R. C. and Fleming, P. J. (2007). On the Evolutionary Optimization of Many Conflicting Objectives. *IEEE Transactions on Evolutionary Computation, 11*(6), 770-784. 10.1109/tevc.2007.910138

Redcedartech (2015). *Heeds smashes barriers on multi-objective design studies*. https://redcedartech.com/newsletters/HEEDSNews-Mar15.htm

Seada, H. and Deb, K. (2015). U-NSGA-III: A Unified Evolutionary Optimization Procedure for Single, Multiple, and Many Objectives: Proof-of-Principle Results. In António Gaspar-Cunha, A., Henggeler-Antunes, C., and Coello-Coello, C. (Eds.) *Evolutionary Multi-Criterion Optimization* (vol. 9019, pp. 34-49). New York, NY: Springer, Cham. 10.1007/978-3-319-15892-1_3

Seada, H., Abouhawwash, M., Deb, K. (2017). Towards a Better Balance of Diversity and Convergence in NSGA-III First Results. In Trautmann, H., Rudolph, G., Klamroth, K., Schütze, O., Wiecek, M., Jin, Y., and Grimme C. (Eds.) *Evolutionary Multi-Criterion Optimization* (vol. 10173, pp. 545-559). New York, NY: Springer, Cham. 10.1007/978-3-319-54157-0_37

Seada, H., Abouhawwash, M., and Deb, K. (2018). Multiphase Balance of Diversity and Convergence in Multiobjective Optimization. IEEE *Transactions on Evolutionary Computation, 23*(3), 503-513. 10.1109/tevc.2018.2871362

Shu, Z., Wang, W., and Wang, R. (2018). Design of an Optimized Architecture for Manned and Unmanned Combat System-of-Systems: Formulation and Coevolutionary Optimization. *IEEE Access, 6*, 52725-52740. 10.1109/access.2018.2870969

Surjanovic, S. and Bingham, D. (2013). *Virtual Library of Simulation Experiments: Test Functions and Datasets*. http://www.sfu.ca/ ssurjano

Tian, Y., Cheng, R., Zhang, X., and Jin, Y. (2017). PlatEMO: A MATLAB Platform for Evolutionary Multi-Objective Optimization [Educational Forum]. *IEEE Computational Intelligence Magazine, 12*(4), 73-87. 10.1109/mci.2017.2742868

Wangsom, P., Bouvry, P., and Lavangnananda, K. (2018). *Extreme Solutions NSGA-III* (*E-NSGA-III*) *for Scien-tific Workflow Scheduling on Cloud*. Paper presented at the 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, USA. 10.1109/icmla.2018.00184

Wang, G., Guo, L., Gandomi, A. H., Hao, G., and Wang, H. (2014). Chaotic Krill Herd algorithm. Information Sciences, 274, 17-34. 10.1016/j.ins.2014.02.123

Wang, G. and Tan, Y. (2019). Improving Metaheuristic Algorithms with Information Feedback Models. *IEEE Transactions on Cybernetics, 49*(2), 542-555. 10.1109/tcyb.2017.2780274

Wang, H. and Yi, J. (2017). An improved optimization method based on krill herd and artificial bee colony with information exchange. *Memetic Computing, 10*(2), 177-198. 10.1007/s12293-017-0241-6

Wang, G., Deb, S., and Cui, Z. (2019). Monarch butterfly optimization. *Neural Computing and Applications, 31*(7), 1995-2014. 10.1007/s00521-015-1923-y

Wang, G. (2016). Moth search algorithm: A bio-inspired metaheuristic algorithm for global optimization problems. *Memetic Computing, 10*(2), 151-164. 10.1007/s12293-016-0212-3

Yarpiz (2018). *Nsga-iii: Non-dominated sorting genetic algorithm, the third version*. http://yarpiz.com/456/ypea126-nsga3

Yi, J., Deb, S., Dong, J., Alavi, A. H., and Wang, G. (2018). An improved NSGA-III algorithm with adaptive mutation operator for Big Data optimization problems. *Future Generation Computer Systems, 88*, 571-585. 10.1016/j.future.2018.06.008

Yi, J., Xing, L., Wang, G., Dong, J., Vasilakos, A. V., Alavi, A. H., and Wang, L. (2020). Behavior of crossover operators in NSGA-III for large-scale optimization problems. *Information Sciences, 509*, 470-487. 10.1016/j.ins.2018.10.005

Yuan, Y., Xu, H., and Wang, B. (2014). An improved NSGA-III procedure for evolutionary many-objective optimization. In Igel, C. (Ed.) *GECCO '14: Proceedings of the 2014 Annual Conference on Genetic and Evolutionary Computation* (pp. 661-668). New York, NY: ACM. 10.1145/2576768.2598342

Zhang, Q., Liu, W., and Li, H. (2009). *The performance of a new version of MOEA/D on CEC09 unconstrained MOP test instances*. Paper presented at the 2009 IEEE Congress on Evolutionary Computation. 10.1109/cec.2009.4982949

Zhang, Y., Wang, G., Li, K., Yeh, W., Jian, M., and Dong, J. (2020). Enhancing MOEA/D with information feedback models for large-scale many-objective optimization. *Information Sciences, 522*, 1-16. 10.1016/j.ins.2020.02.066

Zhang, Q. and Li, H. (2007). MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition. *IEEE Transactions on Evolutionary Computation, 11*(6), 712-731. 10.1109/tevc.2007.892759

Zhang, X., Tian, Y., and Jin, Y. (2015). A Knee Point-Driven Evolutionary Algorithm for Many-Objective Optimization. *IEEE Transactions on Evolutionary Computation, 19*(6), 761-776. 10.1109/tevc.2014.2378512

Zhang, Q., Zhou, A., Zhao S., Suganthan, P. N., Liu, W., and Tiwari S. (2009). *Multiobjective optimization test instances for the cec-2009 special session and competition*. https://www.al-roomi.org/multimedia/CEC_Database/CEC2009/MultiObjectiveEA/CEC2009_MultiObjectiveEA_TechnicalReport.pdf

Zitzler, E., Deb, K., and Thiele, L. (2000). Comparison of Multiobjective Evolutionary Algorithms: Empirical Results. *Evolutionary Computation, 8*(2), 173-195. 10.1162/106365600568202

Research Article / Telecommunications Engineering http://dx.doi.org/10.15446/ing.investig.v40n3.81421

Entropy-Based Image Encryption Using Orthogonal Variable Spreading Factor (OVSF)

Cifrado de Imágenes Basado en la Entropía Utilizando el Factor de Propagación Variable Ortogonal (OVSF)

Dora Maria Ballesteros[1], Diego Renza[2], and Jimmy Peña[3]

ABSTRACT

The purpose of image encryption is to provide data privacy and security. The former ensures that only authorized personnel can access the original content, while the latter implies that there is no evident relationship between the encrypted and the original content, and that the key space is equally likely and large enough. In the current state of the field, there are several proposals of image encryption techniques with very high privacy (in terms of entropy) but weak in terms of security (i.e., small key space). Recently, a new encoding-based method that provides a long key space (namely $8{,}57 \times 10^{506}$) with a middle value of entropy (87%) was proposed. Our proposal preserves the strength of the image encryption methods based on encoding, but with a higher value placed on security than the preliminary works. Every pixel of an image is mapped into an orthogonal code based on 256 bits. The 8-OVSF codes are selected to encode the image, given that the entropy of the inter-symbol is near the possible maximum. Numerous test results verify that our ciphered data have a very high value of entropy (98,5%) with an equally likely and long key space ($8{,}57 \times 10^{506}$), thus providing an adequate balance between privacy and security.

Keywords: entropy, OVSF, image encryption, security, privacy, ciphered data agriculture

RESUMEN

El objetivo del cifrado de imágenes es proporcionar privacidad y seguridad a los datos. La primera garantiza que solo el personal autorizado pueda acceder al contenido original, mientras que la otra implica que no exista relación evidente entre el contenido cifrado y el original, y que el conjunto de claves tenga igualdad de probabilidad y sea lo suficientemente grande. En el estado del arte existen numerosas propuestas de técnicas de cifrado de imágenes con alta privacidad (en términos de entropía), pero con deficiencia en términos de seguridad (es decir, un conjunto de claves pequeño). Recientemente, se propuso un método basado en codificación que proporciona un espacio de clave grande (específicamente, $8{,}57 \times 10^{506}$) con un valor intermedio de entropía (87 %). Nuestra propuesta conserva la fortaleza de los métodos de cifrado de imágenes basados en codificación, pero con mayor valor de seguridad que los trabajos anteriores. Cada píxel de una imagen es mapeado a un código ortogonal de 256 bits. Los códigos 8-OVSF se seleccionan para codificar la imagen, debido a que la entropía del inter-símbolo es cercana al máximo posible. Los resultados de numerosas pruebas demuestran que nuestros datos cifrados tienen un valor de entropía muy alto (98,5 %) con un conjunto de claves grande ($8{,}57 \times 10^{506}$) e igualmente probable, lo que proporciona un equilibrio adecuado entre privacidad y seguridad.

Palabras clave: entropía, OVSF, cifrado de imágenes, seguridad, privacidad, datos cifrados

Received: August, 1st 2019
Accepted: July, 13th 2020

Introduction

Nowadays, the number of images that are published on social networks, chats, public web pages, and other digital mediums is rapidly increasing. In some cases, these images contain confidential content, which is why their originator wants only authorized personnel to have access rights (privacy) through robust key (security) encryption schemes. Accordingly, the security and privacy standards of multimedia content have aroused the interest of the scientific community in the field of information technologies. Therefore, the quantity and quality of state-of-the-art digital privacy and security methods have increased. One way to provide privacy and security to digital image content is through encryption, by applying the properties of diffusion and confusion (Shannon, 1949). Diffusion implies that, if the image is slightly changed, the encrypted data must change significantly; regarding confusion, the direct relationship between the key and the encrypted data must be null. Typically, pixel permutation guarantees confusion, whereas pixel substitution is used for the diffusion property. The ideal result is a ciphered image

[1]Ph.D. in Electronic Engineering, Universitat Politecnica de Catalunya, España. Affiliation: Full professor. Faculty of Engineering, Universidad Militar Nueva Granada, Colombia. Email: dora.ballesteros@unimilitar.edu.co
[2]Ph.D. in Computer Science, Universidad Politecnica de Madrid, España. Affiliation: Full professor. Faculty of Engineering, Universidad Militar Nueva Granada, Colombia. Email: diego.renza@unimilitar.edu.co
[3]Telecommunications Engineering, Universidad Militar Nueva Granada, Colombia. Affiliation: Young research. Faculty of Engineering, Universidad Militar Nueva Granada, Colombia. Email: u1401120@unimilitar.edu.co

How to cite: Ballesteros, D. M., Renza, D., and Peña, J. (2020). Entropy-Based Image Encryption Using Orthogonal Variable Spreading Factor (OVSF). *Ingeniería e Investigación*, *40*(3), 70-80. 10.15446/ing.investig.v40n3.81421

with a uniform distribution (histogram), regardless of the original image's pixel characteristics.

For two decades, one of the most popular methods for image encryption was based on chaotic mapping. Proposed initially by Fridrich (1998), this approach divides the image into several sub-blocks of non-fixed size. Next, after a quantized map is formed, pixel permutation (place) and bit permutation (grayscale level) are applied. The simpler chaotic 1D-map is known as a 'logistic map' (Ye and Huang, 2017; Telem, Segning, Kenne, and Fotsin, 2014). More recent proposals include 2D and 3D sequences (Saljoughi, and Mirvaziri, 2019; Hua, Jin, Xu, and Huang, 2018), as well as dynamic functions at runtime (Asgari-Chenaghlu, Balafar, and Feizi-Derakhshi, 2019). Although the uncertainty (entropy) of the ciphered data in these methods has reached a very high value, security remains a significant challenge, given that these methods have been cryptanalyzed due to weaknesses in the diffusion rounds (Li, Xie, Liu, and Cheng, 2014; Feng, He, Li, and Li, 2019). Currently, researchers focus their efforts on the following aspects: (1) improving the confusion property, (2) improving the diffusion property, and (3) changing the relationship between confusion and diffusion.

To enhance the diffusion property, chaotic sequences are mixed with DNA-based techniques, which consist of changing the value of the shuffled pixel by an encoding process into nucleotides (Zhang, Fang, and Ren, 2014; Chai, Gan, Yuan, Chen, and Liu, 2019). Nevertheless, DNA encoding has also been cryptanalyzed (Wen, Yu, and Lü, 2019; Akhavan, Samsudin, and Akhshani, 2019). Another solution consists of elliptic curves for obtaining pseudo-random sequences (Hayat, and Azam, 2019). This method was also cryptanalyzed (Khoirom, Laiphrakpam, and Themrichon, 2018). Unlike the afore-mentioned methods, in literature, there are proposals that differ from the traditional schemes which have yet to be broken. For instance, Kumar and Quan (2019) analyze images using polar decomposition and the Shearlet transform, whereas, in Ballesteros, Peña, and Renza, (2018) diffusion and confusion tasks are performed through an encoding process using scrambled Collatz conjecture-based codes. However, the entropy value of the ciphered data fails to reach the possible maximum.

In summary, many proposals of methods concerning image encryption have very high entropy values, but suffer from security weaknesses. However, there are some non-traditional proposals that have yet to be cryptanalyzed. The lingering issue with these proposals is that their entropy values are not high enough. Therefore, it can be generally stated that systems that focus on privacy do not possess optimal security, and vice-versa. As a result, the trade-off between security and privacy is still a challenge to be overcome. It is thus necessary to increase the key space and equalize the likelihood in encryption keys for methods based on chaotic mapping while also improving the degree of ciphered data uncertainty in non-traditional schemes.

The highlights of the proposed method in this study are the following:

- Image content privacy is provided through the Orthogonal Variable Spreading Factor (OVSF) coding process. Each pixel (8-bits) of the input image is encoded into an orthogonal code (256-bits) using a specific map obtained from a key. Due to the quasi-perfect symmetry between ones and zeros of the entirety of orthogonal codes, a very high entropy value is theoretically expected.
- The maximum number of possible mappings between the 8-bit pixel values and the 256-bit orthogonal codes provides security to the ciphered content. It corresponds to the number of permutations used to scramble the 8-OVSF codes; a very high value of 256!
- Since the number of bits in the encrypted data is 32 times greater than the input image, it may be more convenient to group them into 16-bit words and write them as an audio file (samples). In other words, our proposed method is image-to-audio encryption.

Background concepts

The aim of this section is to provide some basic concepts which are necessary to understand the proposed method.

Entropy

In information theory, a well-known parameter that measures uncertainty and data distribution is entropy (Robinson, 2008). The greater the homogeneity in data distribution, the greater the value of the entropy. The most famous entropy formula is Shannon's entropy equation, calculated in terms of the probability of each available data value. It is shown as Equation (1):

$$H(I) = -\sum_{k=0}^{L-1} p(k) \log_2 (p(k)) \, (1), \tag{1}$$

where I is the input image (or audio), $p(k)$ is the probability of occurrence of the value k in the image (or audio), and L is the total number of levels of the image (or audio), with $L = 2^q$ for q bits of quantization. For example, in grayscale images, $L = 2^8$. If all intensity values have equal occurrence, then entropy is equal to q (8 in the example). However, if occurrence is not homogenous, the entropy value will be lower because uncertainty decreases.

Orthogonal Variable Spreading Factor (OVSF)

OVSF codes are originally used for channelization in Wideband Code Division Multiple Access (WCDMA). These codes are characterized by being orthogonal to one another, as well as their length being determined in terms of the level (L). An L-OVSF has 2^L codes of 2^L bits each. A way to obtain OVSF codes is through the application of an iterative tree that uses the value of the predecessor code as the root of the current code (Saini and Bhooshan, 2006). Each root has two descendants that double its length. The first

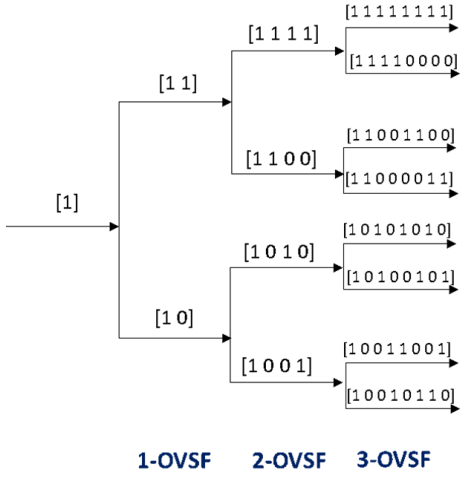


Figure 1. OVSF code tree.
Source: Authors

descendant repeats the predecessor code, and the second one complements it. Figure 1 shows the 3-OVSF code tree.

The first root is code "1", which does not have a counterpart. In the 1-OVSF, there are 2^1 codes of 2^1 bits, obtained as follows: the first code doubles its root and the result is "11". The second one uses the root in the first part and complements it in the second part, thus obtaining "10". In the second level (2-OVSF), there are 4-codes with four bits each. Each pair of codes has the same root. For the first descendant, the root is repeated; for the second, the root is complemented. This process is repeated until all codes are obtained.

The main advantage of OVSF codes is orthogonality, which results in 50% of the bits in a pair of codes being different for every available pair. For example, the codes '1100' and '1001' are equal in the first and third bit, but the second and fourth bits are different.

In terms of entropy, OVSF is expected to provide a high value of uncertainty for the encrypted image due to the quantity of zeros being similar to the quantity of ones. For example, for 3-OVSF, there are seven codes where the number of zeros is equal to the number of ones; only the first code has all bits equal to "1". In total, there are 28 zeros (0,44%) and 36 ones (56%). For 8-OVSF, the number of bits equal to 0 is 32640 (0,498%), while the number of bits equal to 1 is 32 896 (0,502%). For higher values of L, the quantity of zeros and ones is more symmetrical. Thus, the orthogonality property of OVSF codes is used in our proposal for obtaining high entropy values in encrypted images.

The proposed scheme

Recently, a scheme for image encryption based on the Collatz Conjecture was proposed (Ballesteros et al., 2018). Its proposal differs from classical schemes in some aspects: (i) an encoding block with a non-fixed length map is used to replace the permutation and diffusion processes; (ii) the output is not a ciphered image, but a ciphered audio; (iii) the number of available keys related to the security of the scheme is significantly higher than state-of-the-art methods. Since its encoding process uses binary codes obtained from the Collatz Conjecture, which does not have symmetry between zeros and ones, the entropy of the ciphered audio is not close to the highest possible value. The authors reported entropy values of 14 for audio files quantized to 16-bits. In the current proposal, the aim is to preserve the strengths of (Ballesteros et al., 2018) and improve upon its weaknesses by replacing the encoding block with 8-OVSF codes which should theoretically provide greater entropy given their orthogonality. As a result, the uncertainty about which image corresponds to the encrypted audio is greater compared to the scheme in Ballesteros et al. (2018).

Figure 2 shows the proposed general diagram. The inputs of the image coding are the image (I_{2D}) and the *seed*, the outputs are the ciphered audio (CA) and the *key*. In the image recovering module, the inputs are CA and the *key*, the output is the recovered image (RI_{2D}).

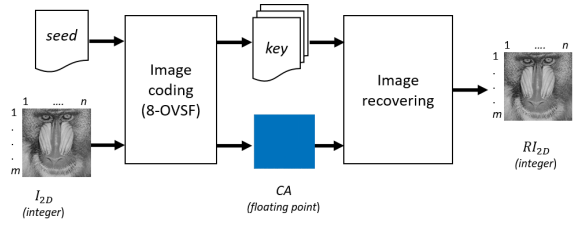


Figure 2. General block diagram of the proposed method.
Source: Authors

Each module in Figure 2 is explained as follows:

Image coding

This module is used at the transmitter stage in order to send the information with unintelligible content. Figure 3 presents the block diagram of this module, including the following blocks: generation of 8-OVSF codes, scrambling the 8-OVSF codes, mapping block, splitting the binary code, and creating the audio file. These blocks are detailed further below:

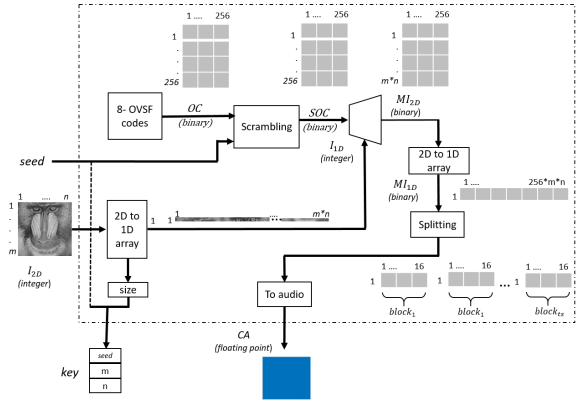


Figure 3. Specific block diagram of the image coding module.
Source: Authors

Generation of 8-OVSF codes: In this block, 256 orthogonal codes of 256 bits are obtained. The creation of these codes follows the theory presented in the *Orthogonal Variable Spreading Factor* section. The first code of the 8-OVSF has all bits equal to one, while the remaining binary codes have an equal number of zeros and ones. These 256 codes form a matrix with dimensions 256 x 256, with each row being an orthogonal code. Unlike the method proposed by Ballesteros et al. (2018), the length of each code is fixed. The output of the block is given the name "OC" (OVSF Codes).

Scrambling the 8-OVSF codes: The aim of this block is to disorder the matrix obtained in the previous block. Using a seed value, the order of the rows in the OC matrix is reorganized. The output is named "SOC" (Scrambled OVSF Codes) with dimensions being 256 x 256. This block provides security to the scheme (analyzed in detail in Section of Security analysis).

Mapping: The 2D image 'I_{2D}' is converted into a row vector, from left to right and top to bottom. The output is named 'I_{1D}'. Each value of I_{1D} is mapped to an orthogonal code. Therefore, every 8-bit pixel is represented by an orthogonal code of 28 bits (i.e., 256 bits). In this regard, I_{1D} is used as the multiplexor selector, where the inputs correspond to the rows in the SOC matrix. The output is a matrix of 256 columns (bits) with m x n rows, named 'MI_{2D}'.

Finally, the total number of bits of encrypted data will be 32 times greater than the original grayscale image. Once each pixel of the image has been mapped, values are arranged into a 1D vector. The output is named '$MI_{ID'}$ (Mapped Image).

Splitting the binary code and creating the audio file: Next, it is necessary to split the binary code into 16 bit words. Each orthogonal code has 16 sub-blocks of 16 bits by block. According to Equation 2, the total number of sub-blocks, ts, is equal to:

$$ts = 16 \times m \times n \tag{2}$$

The result is a matrix of $ts \times 16$, named 'S'. Finally, every sequence of 16 bits of S is transformed into a floating point value, in the range of [−1 1], as a sample of the ciphered audio. It is saved as a WAV file with a specific value for frequency sampling f_s (e.g., f_s = 8 kHz). Time (in seconds) of the ciphered audio is defined by Equation 3:

$$T = \frac{ts}{f_s} = \frac{16 \times m \times n}{f_s} \tag{3}$$

For example, if the original image is 128 × 128 and f_s = 8 kHz, then T = 32,768 (s). You can note that *ts* is the number of samples of ciphered audio.

Key: According to Shannon's theory, security of an encryption system must rely solely on the key. In our proposal, the key is composed of the seed, the number of rows (m), and the number of columns (n) in the image. However, security analysis is performed only on the seed.

The image encoding procedure is illustrated with the following example. Suppose that the system works with 2-OVSF codes (to simplify the example), and the orthogonal codes are '1111', '1100', '1010', and '1001'. The value of the OC matrix will be:

$$OC = \begin{pmatrix} 1111 \\ 1100 \\ 1010 \\ 1001 \end{pmatrix} \tag{4}$$

And now, suppose that from the seed value, the OC matrix is scrambled as follows:

$$SOC = \begin{pmatrix} 1001 \\ 1100 \\ 1111 \\ 1010 \end{pmatrix} \tag{5}$$

Suppose also that the 2D image is 2 x 2, with the following values:

$$I_{2D} = \begin{pmatrix} 1 & 3 \\ 2 & 0 \end{pmatrix} \tag{6}$$

Note that the highest value of I_{2D} for the current example is 3; thus, the system can work with 2-OVSF. For our proposed method, the highest intensity of a pixel is 255, and therefore, it is necessary to work with 8-OVSF.

Continuing the example, I_{2D} is rearranged into a 1D array, resulting in:

$$I_{1D} = \begin{pmatrix} 1 & 3 & 2 & 0 \end{pmatrix} \tag{7}$$

Then, each value of OI_{1D} is used as the multiplexor selector in which the input is SOC and the output is MOI_{2D}. For the current example, row 1 of SOC is selected, followed by row 3, row 2, and row 0. The following matrix is obtained:

$$MI_{2D} = \begin{pmatrix} 1100 \\ 1010 \\ 1111 \\ 1010 \end{pmatrix} \tag{8}$$

Next, MOI_{2D} is transformed to 1D array:

$$MI_{1D} = (1100101011111010) \tag{9}$$

With this 1D array, one sample of the ciphered audio is obtained. However, for the real method proposed in this study, 16 samples are obtained for each pixel of the original image.

Image recovery

The proposed scheme is intended to provide covert communication between a transmitter and receiver. The ciphered audio and the key are transmitted in separate channels, e.g., email, social networks, public webpages, and others. Once the receiver acquires both the ciphered audio and the key, the image can be recovered (see Figure 4).

Each block is explained as follows:

Generation of 8-OVSF codes: This block works in the same way as the corresponding block in the image coding module. The output is named 'OC', with a size of 256 x 256.

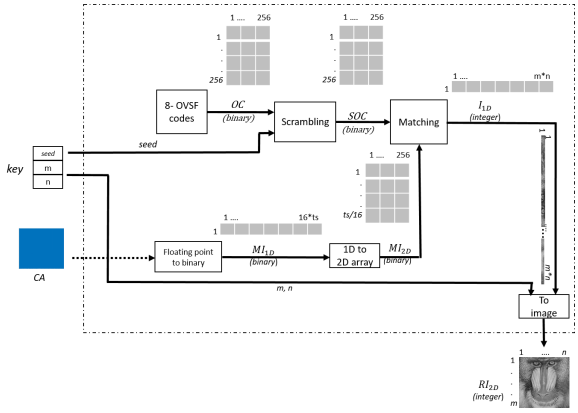


Figure 4. Specific block diagram of the image recovering module.
Source: Authors

Scrambling the 8-OVSF codes: Using the same seed of the image coding module, the OC matrix is reorganized in terms of its rows. The number of available options of the scrambled matrix was discussed in detail in the *Security analysis* section. The output is SOC.

Deciphering the audio: The inputs of this block are the unintelligible, ciphered audio files and the SOC matrix. First, every sample of the ciphered audio, CA, is represented by 16 bits. Second, a 1D array sequence is obtained by the concatenation of all binary representations of CA; this is called MI_{1D}. The total number of bits is calculated using

$$bits = 16 \times ts, (10), \tag{10}$$

where ts is the total number of samples of CA.

Next, MI_{1D} is split into sub-blocks of 256 bits. The number of sub-blocks is obtained by using the ratio between bits and 256; then, the result is $ts/16$. If the process has been performed successfully, $ts/16$ must be equal to $m \times n$. Next, the sub-blocks are arranged into a matrix of $ts/16$ rows with 256 columns named MI_{2D}. This means that every row of MI_{1D} is composed of orthogonal code. Finally, the MI_{2D} code is compared against any SOC code, and subsequently restores the row position of the corresponding match, named I_{1D}.

Creating the grayscale image: Inputs of this block include the key (specifically the values of m and n) and I_{1D}, which is reorganized into a matrix of $m \times n$. The result is the recovered image, I_{2D}.

The above steps are illustrated through an example:

In a similar way to the previously outlined example regarding the image coding module, suppose the system works with 2-OVSF codes and the orthogonal codes are '1111', '1100', '1010', and '1001'. The value of the OC matrix is:

$$OC = \begin{pmatrix} 1111 \\ 1100 \\ 1010 \\ 1001 \end{pmatrix} \tag{11}$$

From the same seed value of the image coding module, the OC matrix is scrambled as follows:

$$SOC = \begin{pmatrix} 1001 \\ 1100 \\ 1111 \\ 1010 \end{pmatrix} \tag{12}$$

In addition, suppose the binary sequence value obtained from the CA is $MI_{1D} = [1100101011111001]$. Since the system works with 2-OVSF, the length of every orthogonal code is 4 bits, and MOI_{1D} is split into 4-bit sub-blocks. The result is the following MI_{2D} value:

$$MI_{2D} = \begin{pmatrix} 1100 \\ 1010 \\ 1111 \\ 1010 \end{pmatrix} \tag{13}$$

The first MI_{2D} code, '1100', is compared against each SOC code as the algorithm searches for a match. It is found that if row zero of MI_{2D} is equal to the first row of SOC, then the returned value is 1. If the first row of MI_{2D}, '1010', is matched with the third row of SOC, then the returned value is 3. Next, the second row of MI_{2D}, '1111', is matched with the 2^{nd} row of SOC, and the returned value is 2. Finally, the 3^{rd} row of MI_{2D}, "1001", is matched with row zero of OC, and the returned value is 0. At the end, the returned value is $I_{1D} = [1320]$. With $m = 2$ and $n = 2$, the recovered image is:

$$RI_{2D} = \begin{pmatrix} 1 & 3 \\ 2 & 0 \end{pmatrix} \tag{14}$$

It is easy to verify that RI_{2D} is equal to I_{2D}.

Performance assessment

Certain metrics have been selected in order to evaluate the performance of the proposed method in terms of similarity between the input image and the recovered image, as well as the quality of the CA.

Metrics to evaluate image similarity

Among the metrics commonly used to measure the image similarity are the Structural Similarity Index (SSIM) and Peak Signal to Noise Ratio (PSNR). These metrics are explained below.

Structural Similarity Index (SSIM): Consider the comparison of two images (A and B), considering luminance (*l*), contrast (c), and structural (s) terms, as follows:

$$\begin{aligned} l(A,B) &= \frac{2\mu_A\mu_B+c_1}{\mu_A^2+\mu_B^2+c_1} \\ c(A,B) &= \frac{2\sigma_A\sigma_B+c_2}{\sigma_A^2+\sigma_B^2+c_2} \\ s(A,B) &= \frac{\sigma_{AB}+c_3}{\sigma_A\sigma_B+c_3} \end{aligned} \tag{15},$$

where μ is the mean, σ is the standard deviation, and CC_1, C_2, and C_3 are constants. With the above results, SSIM is obtained via the following multiplication:

$$SSIM(A,B) = l(A,B) \times c(A,B) \times s(A,B) \tag{15}$$

Using $C3 = 0.5 \times C2$, the Equation above is re-written as:

$$SSIM\,(A,B) = \frac{(2\mu_A\mu_B + c_1)\,(2\sigma_{AB} + c_2)}{(\mu_A^2 + \mu_B^2 + c_1)\,(\sigma_A^2 + \sigma_B^2 + c_2)} \tag{16}$$

SSIM ranges between 0 and 1. The lowest value implies that the structural similarity is null; otherwise, similarity would be high.

***Peak Signal to Noise Ratio* (*PSNR*)**: It is commonly used to compare two images. It is obtained as follows:

$$PSNR = 10 \times \log_{10}\left(\frac{255^2}{MSE}\right) \tag{17}$$

With

$$MSE = \frac{1}{m \times n}\sum_{i=1}^{m \times n}(A_i - B_i)^2 \tag{18}$$

where MSE is the Mean Squared Error, A and B are the images, and the index i represents the absolute position of the pixel (e.g., left to right and up to down). If A and B are equal, then MSE is 0 and PSNR is ∞. Higher values of PSNR are preferable.

Metrics to evaluate quality of the ciphered audio

A good CA file has unintelligible content, meaning that it should appear and sound like noise. Mathematically, this implies that neither neighboring samples are correlated, nor is its entropy low. One way to measure intercorrelation is by using DS metric through Equation (20):

$$DS = \frac{\sum_{i=2}^{m-1}\sqrt{|CA_i - CA_{i-1}| + |CA_i - CA_{i+1}|}}{m-2}, \tag{19}$$

where CA_i is the current sample of CA, CA_{i+1} is the right sample of CA_i, CA_{i-1} is the left sample of CA_i, and m is the total number of samples. Taking into account that natural audio signals are highly intercorrelated, the current sample should be very similar to its neighbors, with the resulting DS being very low. On the other hand, in CA files, the difference between paired neighbor samples is high. The highest value of DS is illustrated in Figure 5 as an example.

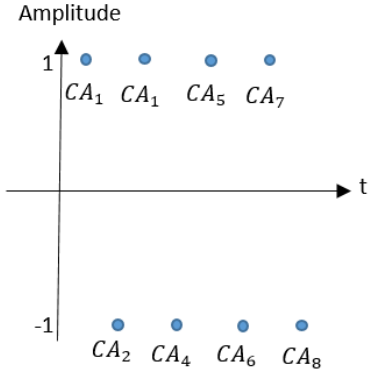


Figure 5. SHighest distance between neighbors.
Source: Authors

Suppose the odd samples are equal to 1; and the even samples, equal to −1. The dynamic range of that signal is 2. Then, by using Equation (20), DS is obtained:

$$DS = \frac{\sum_{i=2}^{m-1}\sqrt{|2| + |2|}}{m-2} = \frac{\sum_{i=2}^{7}(2)}{m-6} = 2 \tag{20}$$

The result above is the maximum DS value for audio signals with a dynamic range of 2. However, in a real-world scenario, if the audio signal is quantized to 16 bits, the total number of different values is 216, distributed in the range [−1 1], with a mean close to 0. Then, a maximum DS value would be:

$$DS_{\max} = \sqrt{|CA^+ - \mu| + |CA^- - \mu|} \tag{21}$$

Where $CA+$ is the highest value of CA, CA^- is the lowest value of CA, and μ is the mean of the audio signal. For $CA^+ = 1$, $CA^- = -1$, and $\mu = 0$, $DS_{\max}$ is equal to $\sqrt{2}$. This means that the ciphered data has unintelligible content.

On the other hand, entropy is a well-known metric to measure the uncertainty of data and quality of CA. For unintelligible audio content with a uniform distribution (i.e., all values being likely), entropy is equal to the number of quantization bits. For example, if the audio is represented with 16 bits, the audio content will be highly unintelligible, due to an entropy value of 16. In other words, the lower the entropy value, the higher the intelligibility of the audio. The formula for entropy was presented in the *Entropy* Section.

Validation

The aim of this section is to validate the performance of the proposed system in terms of the quality of the CA and the recovered image as well as the security analysis. A total of 20 grayscale images (128 x 128 pixels) were used as input for the image coding module; each image is ciphered with 200 keys. At the end, 4000 CA signals were obtained. Figure 6 shows the selected grayscale images.

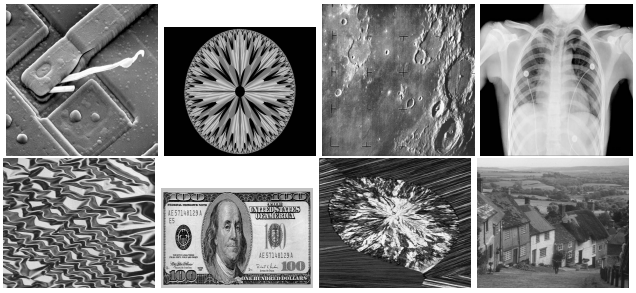


Figure 6. Twenty grayscale images for the validation stage.
Source: Open Source

Preliminary results

The performance of the proposed method is illustrated for three images. Figure 7 shows the original image, the CA, and the recovered image; Figure 8 shows the data distribution (image and CA); and Figure 9 shows the data correlation. According to the results shown in Figure 7, each recovered image is highly similar to the input image. In all three cases, the SSIM is higher than 0,999 and PSNR is ∞. Additionally, the ciphered signals look like noise, with a maximum value of 1 and minimum value of −1. Histograms of the CA signals (Figure 8(b), 8(d), 8(f)) are quasi-uniform, even though the histograms of the images are not uniform (Figure 8(a), 8(c), 8(e)). The entropy of the images is around 6 for 8 bits of quantization (75%), while the entropy of the CA files is 15,75 for 16 bits of quantization (98,5%). These results suggest that the system performed as expected. Finally, Figure 9 shows the behavior of adjacent (horizontal) pixels and neighboring samples. It is clear that the original image is highly correlated because the behavior of the adjacent pixels is around the main diagonal. However, in the case of neighboring samples (audio signal), the data is uncorrelated because the results are scattered in all directions.

Quality of the ciphered audio and the recovered image

This section tests the performance of the proposed method in terms of the quality of the recovered images and their CA files. For the first group of tests, SSIM (Figure 10) and PSNR between the input image and the recovered image were calculated. For the second group, the DS of the CA, entropy of the input image, and entropy of the CA were measured (Figure 11).

Figure 10 shows a very high structural similarity between the input image and the recovered images for all 4000 tests. SSIM values are around 1, and higher than 0,9999990. The PSNR values of 15 images were ∞, while the others were higher than 90 dB. Considering the results above, it means that the proposed method is reversible.

Figure 11(a) shows the entropy results for 4000 CA files. Most of the results (95% confidence) are between 15,75 and 15,78. Therefore, the CA signals are very close to perfectly demonstrating the behavior of unintelligible data (i.e., 16 for data quantized with 16-bits). Regarding DS (Figure 11(b)),

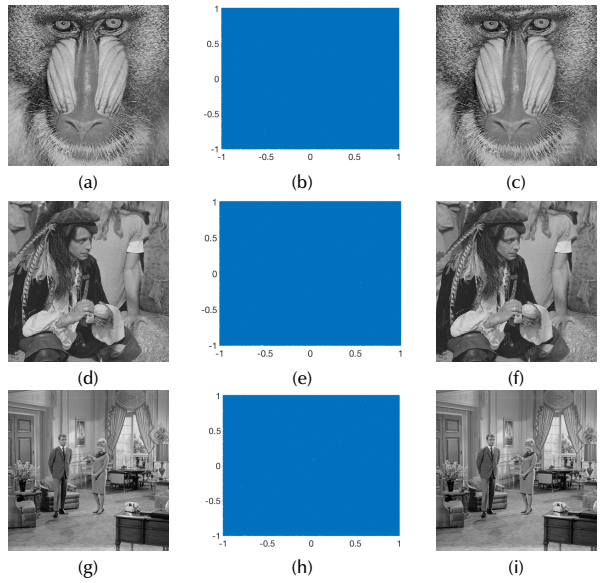


Figure 7. Quality analysis: (a-c) Show input images, (d-f) Show their ciphered audio signals, and (g-i) Show the recovered images.
Source: Open Source and Authors

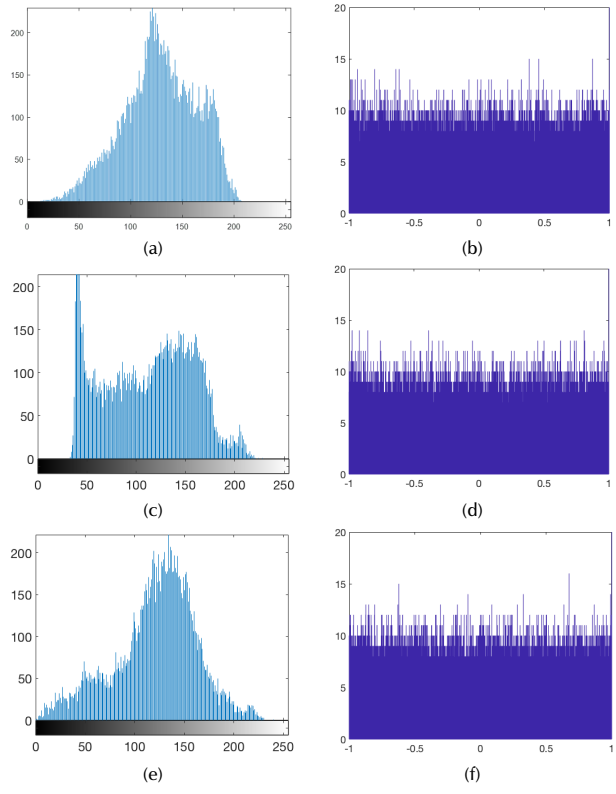


Figure 8. Data distribution analysis: Graphs (a, c, e) show the histogram of the input images, (b, d, f) show histograms of the ciphered data.
Source: Authors

most of the data is in the range [1,1 1,13], which is very close to the expected value (i.e., 1,41). Therefore, the high-quality standard of the ciphered data is verified.

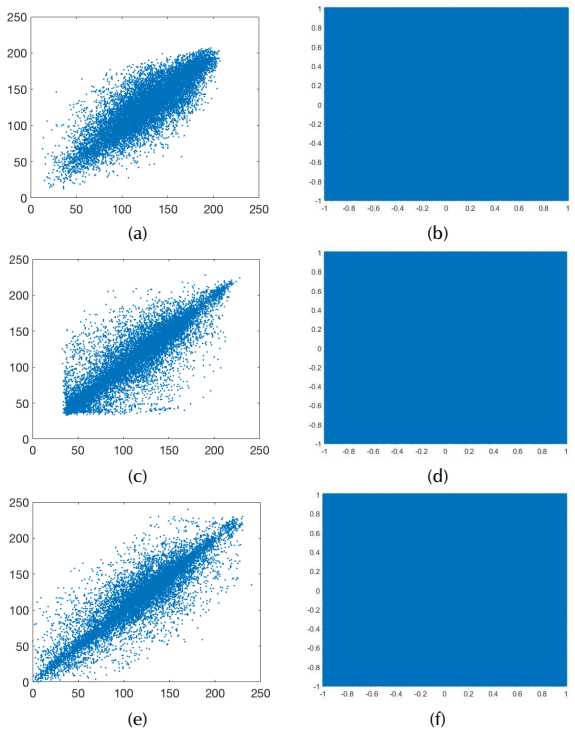


Figure 9. Data correlation: (a, c, e) Show adjacent pixels, graphs (b, d, f) show adjacent samples of their ciphered data.
Source: Authors

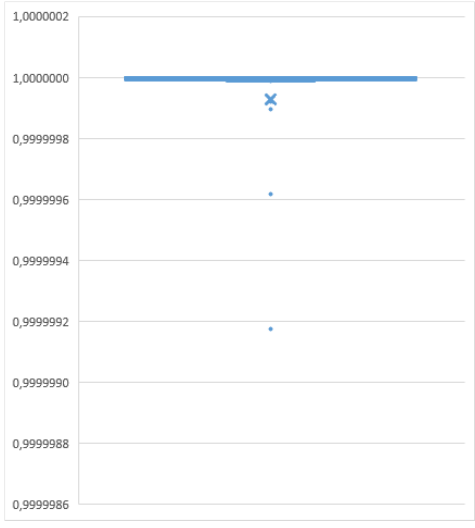


Figure 10. Quality of the recovered image: SSIM.
Source: Authors

Security analysis

The following step in the validation process consists of analyzing and testing the security of the proposed method. First, a theoretical analysis is performed on the key space. Secondly, key sensitivity is measured through several tests.

Key space: According to Shannon's theory, the security of a system must rely solely on the key. It is assumed that the details of the method (e.g., image coding and decoding

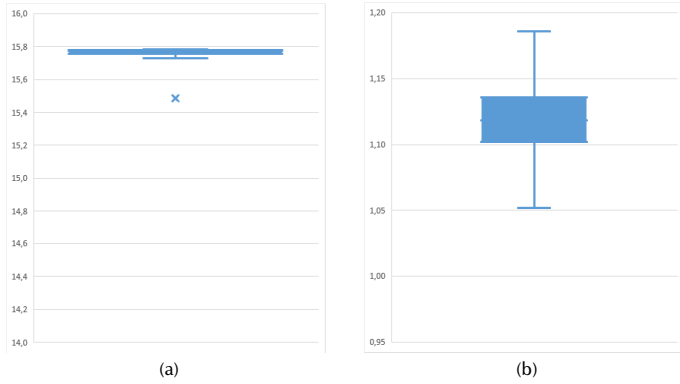


Figure 11. Quality of the ciphered data: (a) Entropy, (b) DS.
Source: Authors

modules) are known by a third person (e.g., eavesdropper). Then, the keys must satisfy the following conditions:

- The number of keys must be large enough to resist a force attack, at least for a considerably long time.
- All keys must be of equal likelihood, which means that the uncertainty of the keys (entropy) must be as high as possible.
- Using a different key must provide a different output.

Albeit the third parties have enough time and hardware resources for testing all the keys, they have no certainty of which one is correct.

To satisfy the first condition, our proposed system uses a seed value as an input for a pseudo-random number generator, which reorganizes the 8-OVSF matrix. Then, there are many available scrambled matrices as the factorial of the number of orthogonal codes. That is, the key space is $256! = 8{,}57 \times 10^{506}$. The reader can then verify that the above-mentioned value is the same as in Ballesteros et al. (2018). For the second condition, the method works with seed values of different lengths and characteristics (e.g., only numbers, only letters, hybrid, uppercase and/or lowercase). The third condition is analyzed in the following section.

Key sensitivity analysis: Although the first two conditions of Shannon's theory are satisfied, the system can still be insecure if two slightly different keys provide the same results. This means that the key sensitivity must be verified as well. At the receiver, the original key is slightly changed (e.g., an upper case instead of a lower case of the same character), and next, the new key is used in the image recovering module. Thus, the dissimilarity between the input image and the recovered image is expected to be high. SSIM is selected to compare the images.

Figure 12 shows an example of this test, using the key "Shannon" in the transmitter, whereas the key "Shannon" is used in the receptor. Figure 13 shows the results (confidence range) of SSIM between the original image and the recovered image for 200 tests.

Figure 12. Example of recovered image using a slight change in the key in the recovering module.
Source: Authors

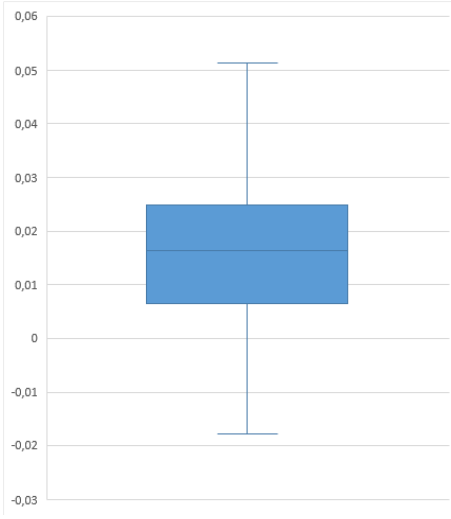


Figure 13. Key sensitivity analysis (confidence range of 95%): SSIM between the input image and the recovered image with a slight change in the key.
Source: Authors

According to Figure 12 and Figure 13, the structural similarity is very low. The ciphered data is very sensitive to the key. That is, if the non-authorized user has access to the ciphered data and knows the method but not the exact key, the original content will not be revealed.

Comparison with state-of-the-art methods

Most of the image encryption methods provide a ciphered image in which the size of the original image is preserved. In our proposal, the output is an audio signal instead of an image. For this reason, comparison to the state-of-the-art methods is divided into three parts: a) image-to-image encryption (Table 1), b) audio-to-audio encryption (Table 2), and c) image-to-audio encryption (Table 3). In all cases, the entropy of the ciphered data (quality of the process) and key space (security of the method) were taken into account.

According to Table 1, the image encryption methods reached a very high entropy value for the encrypted data. However, in audio encryption methods (Table 2), there is still a need for improvement in terms of entropy. In terms of the key space, audio encryption methods are better than image encryption ones. In conclusion, the strength of the image encryption methods is the weakness of the audio encryption methods and vice-versa.

Table 1. Image to image encryption: comparison with some related works

Reference	Core of the method	Entropy	Key space	Cryptanalyzed
Hayat et al., 2019	Elliptic curve	7,97 (8 bits) = 99,6%	10^{38}	Li et al., 2014; Feng et al., 2019
Ye et al., 2017	Logistic map (i.e., 1D chaotic map)	7,99 (8 bits) = 99,8%	10^{42}	
Asgari et al., 2019	Chaotic map (i.e., 2D or 3D)	7,99 (8 bits) = 99,8%	10^{48}	
Chai et al., 2019	DNA sequence	7,99 (8 bits)= 99,8%	10^{98}	

Source: Authors

Table 2. Audio to audio encryption: comparison with some related works

Reference	Core of the method	Entropy	Key space	Cryptanalyzed
Hameed et al., 2018	Logistic map	15,96 (16 bits) = 99,7%	10^{38}	No yet
Belmeguenai et al., 2017	Pseudo-random generator	Not reported	10^{154}	
Renza et al., 2019	Collatz coding	7,50 (8 bits) = 93,7%	10^{506}	
Kalpana et al., 2019	Bidirectional associative memory	Not reported	10^{540}	

Source: Authors

Table 3. Image to audio encryption: comparison with our predecessor

Reference	Core of the method	Entropy	Key space	Cryptanalyzed
Ballesteros et al., 2018	Collatz coding	14 (16 bits) = 87,5%	10^{506}	No yet
ours	8-OVSF coding	15,77 (16 bits) = 98,5%	10^{506}	

Source: Authors

Conclusions

The purpose of this research was to increase the entropy of encrypted data obtained by encryption methods (i.e. 87,5%), while preserving the key space (10^{506}). the mapping process between the pixels of an image and the resulting orthogonal codes provided ciphered data with a very high entropy percentage (98,5%) in a similar way to that of image encryption methods based on chaotic mapping (99,6%).

Taking into account that the image encryption methods based on this encoding method have yet to be been broken and the results of several tests demonstrated high key sensitivity (i.e., SSIM = 0,015 between the original image and the recovered image with a slight change in the key), it is concluded that the aforementioned challenge has been overcome, i.e., the transmitted content remains private and can only be revealed by authorized personnel.

Acknowledgements

This research was funded by "Vicerrectoría de Investigaciones, Universidad Militar Nueva Granada" under grant IMP-ING-2936 of 2019

References

Akhavan, A., Samsudin, A., and Akhshani, A. (2017). Cryptanalysis of an image encryption algorithm based on DNA encoding. *Optics and Laser Technology, 95*, 94-99. 10.1016/j.optlastec.2017.04.022

Asgari-Chenaghlu, M., Balafar, M. A., and Feizi-Derakhshi, M. R. (2019). A novel image encryption algorithm based on polynomial combination of chaotic maps and dynamic function generation. *Signal Processing, 157*, 1-13. 10.1016/j.sigpro.2018.11.010

Ballesteros, D. M., Peña, J., and Renza, D. (2018). A Novel Image Encryption Scheme Based on Collatz Conjecture. *Entropy, 20*(21), 901. 10.3390/e20120901

Belmeguenai, A., Ahmida, Z., Ouchtati, S., and Djemii, R. (2017). A novel approach based on stream cipher for selective speech encryption. *International Journal of Speech Technology, 20*(3), 685-698. 10.1007/s10772-017-9439-8

Broumandnia, A. (2019). Designing digital image encryption using 2D and 3D reversible modular chaotic maps. *Journal of Information Security and Applications, 47*, 188-198. 10.1016/j.jisa.2019.05.004

Chai, X., Gan, Z., Yuan, K., Chen, Y., and Liu, X. (2019). A novel image encryption scheme based on DNA sequence operations and chaotic systems. *Neural Computing and Applications, 31*(1), 219-237. 10.1007/s00521-017-2993-9

Feng, W., He, Y., Li, H., and Li, C. (2019). Cryptanalysis and improvement of the image encryption scheme based on 2D logistic-adjusted-sine map. *IEEE Access, 7*, 12584-12597. 10.1109/ACCESS.2019.2893760

Fridrich, J. (1998). Symmetric ciphers based on two-dimensional chaotic maps. *International Journal of Bifurcation and chaos, 8*(06), 1259-1284. 10.1142/S021812749800098X

Hameed, Y. and Ali, N. (2018). An efficient audio encryption based on chaotic logistic map with 3D matrix. *Journal of Theoretical and Applied Information Technology, 96*, 5142-5152.

Hayat, U. and Azam, N. A. (2019). A novel image encryption scheme based on an elliptic curve. *Signal Processing, 155*, 391-402. 10.1016/j.sigpro.2018.10.011

Hua, Z., Jin, F., Xu, B., and Huang, H. (2018). 2D Logistic-Sine-coupling map for image encryption. *Signal Processing, 149*, 148-161. 10.1016/j.sigpro.2018.03.010

Kalpana, M., Ratnavelu, K., and Balasubramaniam, P. (2019). An audio encryption based on synchronization of robust BAM FCNNs with time delays. *Multimedia Tools and Applications, 78*(5), 5969-5988. 10.1007/s11042-018-6373-y

Khoirom, M. S., Laiphrakpam, D. S., and Themrichon, T. (2018). Cryptanalysis of multimedia encryption using elliptic curve cryptography. *Optik, 168*, 370-375. 10.1016/j.ijleo.2018.04.068

Kumar, R. and Quan, C. (2019). Asymmetric multi-user optical cryptosystem based on polar decomposition and Shearlet transform. *Optics and Lasers in Engineering, 120*, 118-126. 10.1016/j.optlaseng.2019.03.024

Li, C., Xie, T., Liu, Q., and Cheng, G. (2014). Cryptanalyzing image encryption using chaotic logistic map. *Nonlinear Dynamics, 78*(2), 1545-1551. 10.1007/s11071-014-1533-8

Liu, J. M. and Qu, Q. (2010, October). Cryptanalysis of a substitution-diffusion based image cipher using chaotic standard and logistic map. In Qingling, l., Fei, Y., and Yun L. (Eds.) *2010 IEEE Third International Symposium on Information Processing* (pp. 67-69), Qingdao, China: IEEE. 10.1109/ISIP.2010.33

Patidar, V., Pareek, N. K., and Sud, K. K. (2009). A new substitution–diffusion based image cipher using chaotic standard and logistic maps. *Communications in Nonlinear Science and Numerical Simulation, 14*(7), 3056-3075. 10.1016/j.cnsns.2008.11.005

Renza, D., Mendoza, S., and Ballesteros L, D.M. (2019). High-uncertainty audio signal encryption based on the Collatz conjecture. *Journal of Information Security and Applications, 46*, 62-69. 10.1016/j.jisa.2019.02.010

Robinson, D. W. (2008). Entropy and uncertainty. *Entropy, 10*(4), 493-506. 10.3390/e10040493

Saini, D. S. and Bhooshan, S. V. (2006, July). Adaptive assignment scheme for OVSF codes in WCDMA. In Dini, P., Christer, Å., Dini, C., and Borcoci, E. (Eds.) *2006 IEEE International Conference on Wireless and Mobile Communications (ICWMC'06)* (pp. 65-65), Bucharest, Romania: IEEE. 10.1109/ICWMC.2006.15

Saljoughi, A. S. and Mirvaziri, H. (2019). A new method for image encryption by 3D chaotic map. *Pattern Analysis and Applications, 22*(1), 243-257. 10.1007/s10044-018-0765-5

Shannon, C. E. (1949). Communication theory of secrecy systems. *The Bell system technical journal, 28*(4), 656-715. 10.1002/j.1538-7305.1949.tb00928.x

Telem, A. N. K., Segning, C. M., Kenne, G., and Fotsin, H. B. (2014). A simple and robust gray image encryption scheme using chaotic logistic map and artificial neural network. *Advances in Multimedia, 2014*. 10.1155/2014/602921

Wang, B., Wei, X., and Zhang, Q. (2013). Cryptanalysis of an image cryptosystem based on logistic map. *Optik, 124*(14), 1773-1776. 10.1016/j.ijleo.2012.06.020

Wen, H., Yu, S., and Lü, J. (2019). Breaking an image encryption algorithm based on DNA encoding and spatiotemporal chaos. *Entropy, 21*(3), 246. 10.3390/e21030246

Ye, G. and Huang, X. (2017). An efficient symmetric image encryption algorithm based on an intertwining logistic map. *Neurocomputing, 251*, 45-53. 10.1016/j.neucom.2017.04.016

Zhang, J., Fang, D., and Ren, H. (2014). Image encryption algorithm based on DNA encoding and chaotic maps. *Mathematical Problems in Engineering, 2014*. 10.1155/2014/917147

Zhu, H., Zhao, Y., and Song, Y. (2019). 2D logistic-modulated-sine-coupling-logistic chaotic map for image encryption. *IEEE Access, 7*, 14081-14098. 10.1109/ACCESS.2019.2893538

Research Article / Agricultural Engineering http://dx.doi.org/10.15446/ing.investig.v40n3.82877

Composting of Byproducts from the Orange (*Citrus sinensis* (L.) Osbeck) and Sugarcane (*Saccharum* spp. hybrids) Agroindustries

Compostaje de Subproductos de las Agroindustrias de Naranja (*Citrus sinensis* (L.) Osbeck) y Caña de Azúcar (*Saccharum* spp. hybrids)

Teresita de Jesús Debernardi-Vázquez[1], Noé Aguilar-Rivera[2], and Rosalía Núñez-Pastrana[3]

ABSTRACT

There is a global problem involving the generation, management, disposal, and treatment of agro-industrial waste, since it requires technologies for its transformation into the sustainable production of food, fodder, biofuels, fibers, bioproducts, and biofertilizers. The citrus and sugarcane agroindustries generate waste or byproducts that require a special treatment to be reused. Sugarcane byproducts already have an established but unsustainable use. In the case of citrus fruits, 50-60% of the organic waste has an acidic pH (3-4), 95% organic matter, and 80-90% moisture. The objective of this research was to evaluate and model the pile composting method applied to byproducts of the orange and sugarcane agroindustries to obtain compost with competitive physicochemical parameters for use as a soil improver. The pile was designed with a pyramidal-rectangular arrangement, and 9 points were established for temperature sampling in three different sections during composting. An average of 55 °C and a pH of 9,1 were obtained at the end of the process. Compared to the regional control, the C/N ratio (27,5), organic matter (65,5%), and macro- and microelements show improved characteristics for use in food production in agro-ecological agriculture.

Keywords: orange peel, compost, bagasse, temperature

RESUMEN

Hay una problemática global que implica la generación, manejo, disposición y tratamiento de residuos agroindustriales, dado que requieren tecnologías para su transformación en la producción sostenible de alimentos, forrajes, biocombustibles, fibras, bioproductos y biofertilizantes. Las agroindustrias de cítricos y caña de azúcar generan residuos o subproductos que requieren un tratamiento especial para ser reutilizados. Los subproductos de caña ya tienen un uso establecido, pero no sostenible. En el caso de los cítricos, entre 50-60 % de los residuos orgánicos tienen pH ácido (3-4), 95 % de materia orgánica y 80-90 % de humedad. El objetivo de la presente investigación fue evaluar y modelar el método de compostaje de pila en subproductos de la agroindustria de naranja y caña de azúcar para obtener composta con parámetros fisicoquímicos competitivos para uso como mejorador de suelos. La pila se diseñó con un arreglo piramidal–rectangular y se establecieron 9 puntos para el muestreo de temperatura en tres secciones diferentes durante el compostaje. Se obtuvieron un promedio de 55 °C y un pH de 9,1 al final del proceso. Comparado con el testigo regional, la relación C/N (27,5), la materia orgánica (65,5 %), y los macro y microelementos presentan características mejoradas para su empleo en la producción de alimentos en la agricultura agroecológica.

Palabras clave: cáscara de naranja, composta, bagazo, temperatura

Received: October 10th, 2019
Accepted: July 7th, 2020

Introduction

Currently, research aimed at the use of agronomic techniques that improve productivity in terms of quantity and quality of products has increased (Edgerton, 2009). At the same time, there are ongoing efforts to promote the conservation of natural resources, the protection of soil, and the reduction of environmental impact (Zhao et al., 2018). Recent studies that compare the properties of soil treated with organic and chemical fertilizers have shown the former to improve nitrogen and organic matter contents and increase productivity (Drinkwater and Snapp, 2007; Yang, Su, Wang, and Yang, 2016). The addition of organic supplements such as compost makes it possible to stabilize the soil structure, increase the density of the fertile layer, and maintain a healthy

[1]Agrochemical Engineer, Universidad Veracruzana, México. M.Sc. in Chemical Engineer Instituto Tecnológico de Orizaba Veracruz, México. Ph.D. Student. Ph.D. Program in agricultural sciences, Universidad Veracruzana, México.
Email: tjdebernardi@gmail.com
[2]Chemical Engineer, Instituto Tecnológico de Orizaba Veracruz, México. M.Sc. in Forest Products, Universidad de Guadalajara, México. Ph.D. in Environmental Sciences, Universidad Autónoma de San Luis Potosí, México. Research professor, Universidad Veracruzana, México. Email: naguilar@uv.mx
[3]Biochemical Engineer, Instituto Tecnológico de Morelia, México. M.Sc. in Food. Instituto Tecnológico de Tepic, México. Ph.D. in Plant Biotechnology, Centro de Investigación Científica de Yucatán, México. Research professor, Universidad Veracruzana, México. ronunez@uv.mx

How to cite: Debernardi-Vázquez, T. de J., Aguilar-Rivera, N., Núñez-Pastrana, R. (2020). Composting of Byproducts from the Orange (*Citrus sinensis* (L.) Osbeck) and Sugarcane (*Saccharum* spp. hybrids) Agroindustries. *Ingeniería e Investigación*, *40*(3), 81-88. 10.15446/ing.investig.v40n3.82877

environment (Carrizo, Alesso, Cosentino, and Imhoff, 2015). With the use of biological processes, agricultural residues can be transformed into biofertilizers (Muscolo, Papalia, Settineri, and Mallamaci, 2018) for the development of sustainable agriculture. The speed and degree of degradation of the waste depends on the quality of the materials used and the conditions of the process.

The composting process is a method that involves thermophilic and mesophilic organisms which consume organic matter as a substrate under controlled conditions, thus obtaining a stabilized, mature, deodorized, hygienic, pathogen-free, and plant-seed product, rich in humic substances with ideal properties to improve the soil (Lim, Lee, and Wu, 2016). Composting of lignocellulosic food waste is an alternative for sustainable conversion, as it generates a green product; it can be considered as an effective method to transform organic matter into a potentially safe, stable, and harmless material that can be used as a soil improver, organic biofertilizer, or peatmoss substitute in cultivated soils (Chen et al., 2014).

Several studies indicate that the addition of various materials during compost preparation process can accelerate the process and improve the quality of the product (Gabhane et al, 2012; Makan, 2015). In the composting of organic matter with livestock residues, bagasse is commonly used as a structuring agent. The main purpose is to maintain the appropriate water content in the mixture and promote the formation of porous spaces in the compost mass. This increases oxygen availability and reduces the loss of static pressure in piles with forced aeration (Teixeira, de Matos, and Melo, 2015). The addition of sufficient cane bagasse increases the pH from 3,5 to 5 and can improve the transformation of nutrients from organic waste by controlling pH (Cole, Roberts, Garside, de Nys, and Paul, N. A, 2016). In addition, bagasse can also be used as a carbon source to adjust the carbon/nitrogen (C/N) ratio and improve the availability of nutrients (nitrogen, phosphorus, and potassium) by increasing the content of organic matter and nitrogen in the compost, thus reducing the loss of gaseous nitrogen (Kumar, Verma, Singh, Kumar, U., and Shweta, 2010). In previous studies, the application of cane bagasse in the composting of municipal solid waste improved the quality and shortened the process time required for its stabilization (Mohee et al., 2015). Another study showed that adding 15% sugarcane bagasse and 20% exhausted grape marc during the two-stage method for composting green waste improved the quality of the final product and reduced the compost maturation time to 21 days, compared to the 90-270 days required by the traditional process (Zhang and Sun, 2016). There are several studies in which various agroindustrial wastes have been used for the production of compost: filter mud, vinasse, coffee pulp, citrus, fruit-horticultural waste, etc. (Caballero and Soto, 2019)

Citrus fruits are the most consumed fruits around the world, due to the nutritional value and health benefits resulting from the amount of secondary metabolites they possess (Panwar, Panesar, and Chopra, 2019). Brazil, India, Mexico, Spain, and the United States produce over two-thirds of the world's citrus fruits (Zema, et al., 2018). The *Citrus* genus includes several important fruits but is dominated by sweet oranges (*Citrus sinensis*), which can be classified into three groups: common oranges (White or Golden oranges), navel orange, and pigmented or blood orange (Ortiz-Sánchez, Solarte-Toro, Orrego-Alzate, Acosta-Medina, and Cardona-Alzate, 2020; Mamma and Christakopoulos, 2014).

Citrus processing involves the following stages: after the harvest, they are transported to the processing plant, where the waste is removed for later washing. The next steps consist of extracting the juice by squeezing and centrifuging, after which the juice is heated to activate the pectinase enzymes and transferred to the concentrators to add citric acid, enzymes, and vitamin C, in order to obtain the best quality. Finally, the juice is packaged in sterilized containers (Taghizadeh-Alisaraei, Hosseini, Ghobadian, and Motevali, 2017).

Approximately one-third of citrus fruits are used in the process, which produces between 50 and 60% organic waste. The processed citrus residues have an acidic pH (3-4) and are characterized by having a high quantity of organic matter (95%) and a large amount of water (between 80 and 90%) (Behzad and Keikhosro, 2018). Citrus byproducts include fats, free sugars (glucose, fructose and sucrose), organic acids, polymers (cellulose, hemicellulose and pectin), enzymes (pectinases, phosphatases and peroxidases), flavonoids, essential oils (limolene), and pigments (Boukroufa, Boutekedjiret, Petigny, Rakotomanomana, and Chemat, 2015). Organic acids from processed oranges (malic, malonic, and oxalic), together with other components, represent 1% of the dry weight (Raveh et al., 2020). In citrus fruits, the taste depends mainly on the amount of sugars and organic acids, while volatile components are associated with aroma. Therefore, the residues from citrus processes obtained during the extraction of the juice contain seeds that are characterized by their high content of oils, where C_{16} and C_{18} acids predominate (Bora, Kamle, Mahato, Tiwari, and Kumar, 2020; Joglekar, Pathak, Mandavgane, and Kulkarni, 2019). Veracruz is the main citrus producing state (Table 1 and Figure 1).

Table 1. Citrus production

Citrus	planted area (ha)	Harvested citrus (t)	Yield (t/ha)
Orange	167,884	2,331,660	11,13
Lemon	46,319	658,282	12.20
Tangerine	20,583	322,494	15,75
Grapefruit	7,896	242,409	29,40
Total	242,682	3,554,845	18,76

Source: SEDARPA, 2019

In Mexico, sugarcane production generates more than 2 million jobs, both directly and indirectly, and is carried out in 15 federal states and 227 municipalities (Figure 2). The 2018/19 harvest was conducted in 50 sugar mills. The harvested area was 804 060 ha, producing 57 036 700 t of cane to be ground, with a field yield of 70 936 t/ha, a production of 6 425 919 t, with a factory yield of 11 266%,

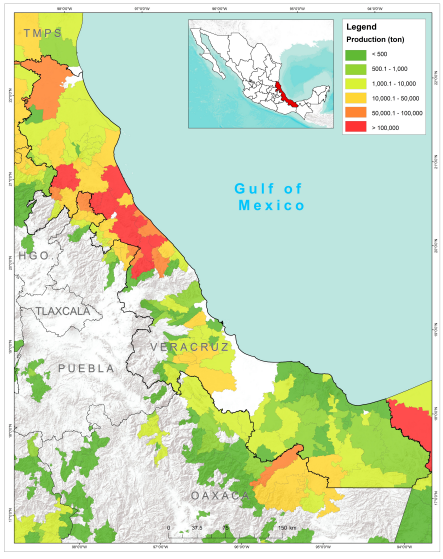


Figure 1. Citrus producing municipalities in Veracruz, Mexico.
Source: SEDARPA, 2019

and an agroindustrial yield of 7 992 t of sugar/ha. Veracruz produces 40% of the total domestic production of cane and sugar (CONADESUCA, 2019).

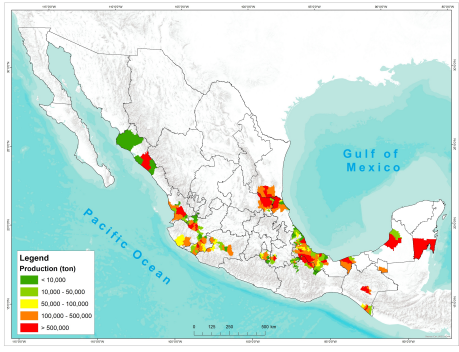


Figure 2. Sugarcane producing municipalities in Mexico.
Source: CONADESUCA, 2019

Among the main byproducts of the sugarcane industrialization process is filter mud, which is formed by the residues that are obtained in the sugarcane juice clarification process during raw sugar production. It is a dark material, consisting of a mixture of fiber, coagulated colloids, wax, albuminoid substances, calcium phosphates, and soil particles. On average, 30 kg are produced for each ton of sugarcane that is ground. It is generally applied on soils next to sugar mills without composting treatment. The filter mud components that stand out are organic matter, calcium, phosphorus, and nitrogen (PROCAÑA, 2019).

Bagasse is the residue of the industrial sugar manufacturing process, the remainder of the sugarcane stalks after the sugary juice they contain is extracted through the mills. Whole bagasse is divided into pith and fiber. Within this context, cane bagasse is a potential fibrous residue for the production of green products that is available in large quantities and has a chemical composition very similar to wood. However, the traditional and most widespread use today is the production of steam by combustion in the boilers of the sugar mill itself. This activity consumes from 50 to 100% of the bagasse that is generated in the process. Bagasse fibers are rigid with irregular, well-defined contours; they are the bearers of the structural elements necessary for the derivatives industry, whose composition is influenced by the agricultural processing conditions of the cane, the type of cut, the collection, and sugar technology (Aguilar, 2010).

The ash of sugarcane bagasse is an inorganic residue that remains after burning it in boilers, whose final temperature is between 700 and 750 °C. It is predominantly composed of silicon, aluminum, calcium, and iron compounds, along with minimal amounts of magnesium, titanium, sodium, and potassium. Although its constituents are oxides, they are represented mostly as a mixture of silicates, oxides, and sulfates with tiny amounts of other elements. Silicates have their origin in slate and clay minerals; iron oxide derives mainly from pyrites, which, when burned, produce ferric oxide and sulphur oxides; whereas calcium and magnesium oxides result from the decomposition of mineral carbonates. Sulfates are formed mostly by the interaction of pyrite mineral carbonates and oxygen.

Ash decomposition varies according to the variety of the cane and age, the type of soil, and the amount of fertilizers. The amount of ash in the bagasse depends largely on the amount of soil, which is a foreign material collected during harvest. The normal amounts of ash in low rainfall conditions are between 2 and 4% of the total bagasse (Andreão, Suleiman, Cordeiro, and Nehdi, 2019). The objective of this research was to model the process of composting byproducts from the orange and sugarcane agroindustries simultaneously in a compost pile, in order to obtain stabilized biofertilizer with suitable parameters for use as a soil improver.

Materials and methods

In this investigation, a univariate experimental design was used for composting orange and cane byproducts through the pile method. The experimental process was conducted in a greenhouse located in the facilities of the Faculty of Biological and Agricultural Sciences of the University of Veracruz, located in Peñuela, Amatlán de los Reyes, Veracruz, México. The experiment was carried out in a greenhouse so that outdoor climatic and environmental conditions would not interfere with the experimental results and variables such as temperature and moisture.

During the composting process, the following materials were used: orange peel, green sugarcane leaves, bagasse, ash, filter mud, sugar mill byproducts, and spent substrate from the production of *Pleurotus ostreatus*, a highly consumed mushroom in the region. However, the residue from its production has no use and is considered waste or garbage. The moisture of the substrates was determined by gravimetry (Table 2).

Table 2. Byproducts used

Material	Amount (kg)	Initial moisture (%)	Composition (%)
Orange peel	89	84	48,58
Green-sugarcane leaf	40	80	21,83
Sugarcane bagasse	23	15	12,55
Ash	6,0	12	3,28
Sugarcane filter mud	11,2	20	6,11
Edible Mushroom spent substrate	14	16	7,65
Total	183,2	–	100

Source: Authors

The materials were previously weighed and conditioned for the preparation of the compost pile, which was made in a pyramid arrangement with nine sampling points (Figure 3).

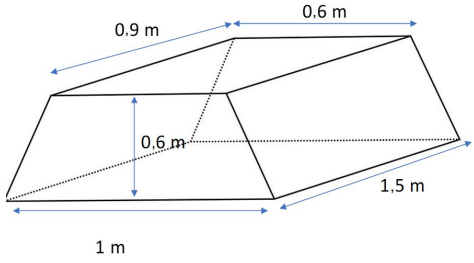


Figure 3. Layout of the compost pile.
Source: Authors

The materials were placed in interspersing layers to keep the moisture of the substrate and the pore space in balance, thus ensuring gas exchange inside the pile. During the experimental process, temperature was monitored manually using a Tel-Tru thermometer for composts at nine points: three on the surface, three in the center, and three at the base corresponding to the treatments. The pH in the substrate was determined once a week using a Hanna Instruments HI2209-01 pH meter. The pile was irrigated weekly during the turning process.

At the end of the process, nutrient and micronutrient compositions were analyzed and compared with regional biofertilizers: filter mud and compost (filter mud 75%, and bagasse 25%) from a sugar mill based on the Official Mexican Standards NMX-AA-025-1984 and NOM-021-SEMARNAT-2000.

Results and Discussion

During the composting process, various changes in the substrate structure were observed: fungi growth and yeasts, complete degradation of the components (which became humus), temperature increase, dehydration, etc. It is important to note that, in this research, the growth of microorganisms on the surface of the substrate was not quantified; it was only observed qualitatively and was correlated with the increase in temperature inside it. The evolution of temperature in the different sampling sections is presented in Figures 4, 5, and 6.

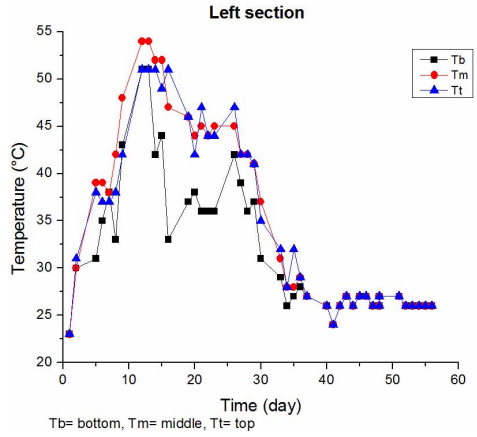


Figure 4. Temperature evolution in left section.
Source: Authors

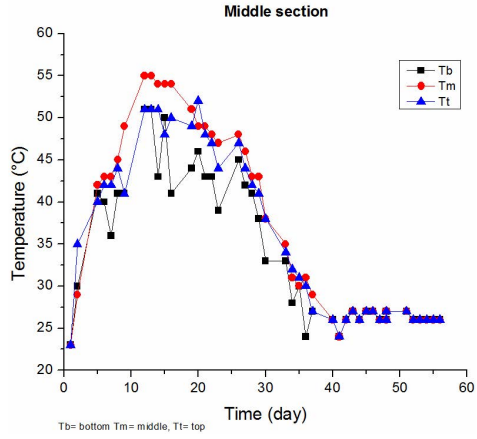


Figure 5. Temperature evolution in middle section.
Source: Authors

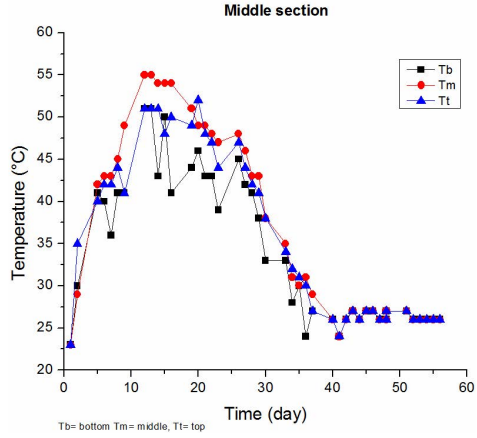


Figure 6. Temperature evolution in right section.
Source: Authors

Figures 4, 5, and 6 show that, at the beginning of the composting process, the substrate temperature was equal

to that of the environment. However, once the metabolic process of the mesophilic microorganisms had begun, the necessary conditions were created so that different groups of bacteria, yeasts, and fungi developed, producing an increased temperature in the substrate as a result of metabolic activity (*Cai et al., 2013*).

A technology commonly used to manage organic waste is biological composting, which stabilizes organic matter to a humus with almost no odor or pathogens and can be applied beneficially to the soil (Nwanze and Clark, 2019). During this process, bacteria, actinomycetes, fungi, molds, and yeasts oxidize long and short chain fatty acids, paper products, and other contaminants, thus producing heat, which leads to a reduction in waste due to microbial conversion and the evaporation of water (Tuomela, Vikman, Hatakka, and Itävaara, 2000). Protozoa help to consume bacteria and fungi, while rotifers control the growth of bacteria and protozoa, which minimizes the health risks associated with pathogens (Haug, 1993). The main aim of composting is to maximize conversion of organic material. Therefore, water is added to the process when the organic matrix reaches a certain dryness in order to preserve moisture for optimal microbial activity and maximum organic conversion. As a consequence, long residence times of around 50 days are required (Cai et al., 2013). The factors affecting the composting process are temperature, the initial C/N ratio, aeration, porosity, moisture content, and pH (Shafawati and Siddiquee, 2013). These parameters are regulated and controlled to provide an optimal environment for the development of microorganisms that degrade organic matter (López et al., 2015).

The composting process can be divided into three phases: the initial activation phase, a thermophilic phase, and a mesophilic or maturation phase (Chowdhury, Akratos, Vayenas, and Pavlou, 2013). Most of the degradation of organic waste occurs during the thermophilic phase. In this phase, microorganisms degrade readily available compounds in organic waste. In general, a high microbial activity translates into a high degradation of organic waste (Fourti, 2013). It is characterized by the high temperature in the composting pile due to the heat released by microbial catabolism (Singh and Kalamdhad, 2014). The high temperature reached at this stage is also crucial for pathogen reduction and disinfection. Temperatures above 55 °C are required to eliminate pathogens in organic waste (Tian et al., 2012).

The maximum temperature in the three sections of the pile was 55 °C on average, which was obtained on the 15th day of the process. In piles with a greater volume of material, the temperature could reach 70 °C during the degradation of animal manure (Tang, Yu, Liu, Xu, and Shen, 2011) and green waste (Cáceres, Coromina, Malińska, and Marfà, 2015). The end of the thermophilic phase and the beginning of the maturation phase were indicated by temperature decrease in the composting pile. As organic waste stabilizes, the temperature will continue to decrease to room temperature (Sánchez, Serramiá, Civantos, Fernández, and Roig, 2010), which also marks the depletion of the degradable organic fraction in the waste (Ravindran and Sekaran, 2010).

Figure 7 presents the evolution of pH in the compost pile. This parameter is an indicator of the aeration process, since, in anaerobic conditions, organic acids are produced which would cause the pH to decrease. The components present in the substrate of the tested pile determined the bioprocess that occurred, given that, when the content of lignocellulose-rich materials (cane bagasse) is predominant, the degradation process is lower in relation to the amount of organic matter and the C/N ratio obtained at the end of the process.

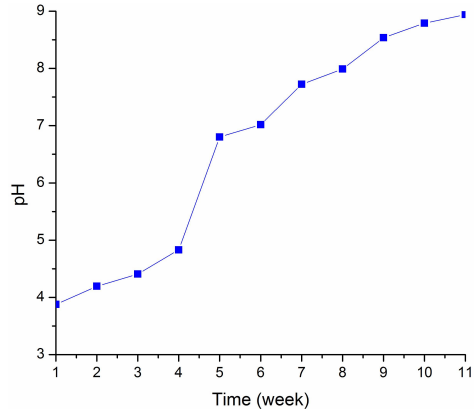


Figure 7. Evolution of pH in the compost pile.
Source: Authors

The composting process was performed under aerobic conditions (Figure 6). The initial pH of the employed substrate was 3,5 and, at the end of the process (11 weeks), it was 9,1. The aeration process contributed to raising the pH of the substrate, thus preventing the generation of organic acids due to the growth of anaerobic microorganisms, although ash and filter mud contributed to its alkalinization.

Table 3 presents the results obtained from the physicochemical analysis of the final product with regional composts used as a control.

The C/N ratio indicates that there is available nitrogen, and that the organic matter present in the humus will continue to decompose once it is incorporated as a substrate for plant growth. The results obtained show that the final product has the organic matter and macro and micronutrients necessary for its use as a soil improver. However, it is advisable to take into account the mechanism of its use and agronomic and economic evaluation in plantation, perennial, vegetable, and ornamental crops, as well as to evaluate the availability of byproducts throughout the agricultural year and the production costs for their industrial scale-up in sugar mills.

The importance of the valuation of citrus residues focuses on the potential of bioeconomy with conventional and sustainable technologies and public policies applied by stakeholders. Otherwise, the environmental impact caused by these byproducts of the citrus processing agroindustry makes it necessary to search for productive alternatives to establish value chains and bioproducts through integrated technological schemes such as citrus biorefining (Satari and Karimi, 2018). The main environmental issue with processed

Table 3. Physicochemical analysis of the compost obtained

Variable	Unity	Filter Mud (control)	Filter mud compost (control)	Cane-Orange waste
Moisture	(%)	47,565	44,468	50,067
pH		7,790	6,062	9,113
Electric conductivity	dS m^{-1}	1,624	1,878	5,340
Ash	(%)	50,775	42,303	34,540
Organic matter	(%)	49,225	55,457	65,460
Total carbon	(%)	28,553	33,953	37,970
Total nitrogen	(%)	0,969	1,890	1,407
C/N ratio		33	18,445	27,517
Calcium (CaO)	(%)	6,544	5,075	4,689
Magnesium (MgO)	(%)	0,898	0,696	0,842
Sodium (Na_2O)	(%)	0,051	0,026	0,087
Potassium (K_2O)	(%)	0,537	0,487	2,079
Phosphorus (P_2O_5)	(%)	4,713	2,920	1,786
Iron (Fe)	(%)	0,621	0,854	0,164
Copper (Cu)	(%)	0,006	0,028	0,002
Zinc (Zn)	(%)	0,026	0,073	0,010
Manganese (Mn)	(%)	0,259	0,285	0,051

Source: Authors

citrus residues is their high degree of fermentation due to their carbohydrate content, which can accelerate the degradation process (Lin et al., 2013). Traditionally, processed citrus residues are incinerated or deposited in landfills. However, this is insufficient and problematic in terms of environmental impact, the energy cost to eliminate moisture content (if they are subjected to incineration), and the economic cost involved in its transport to landfills (Wei et al., 2017). Only a part of the processed waste is used as a supplement for animal feed, although it has the disadvantage of containing a low level of protein, despite its organic matter content.

This situation prevails in a significant number of countries in the Americas and Asia, so it is possible to obtain value-added derivatives of processed citrus residues which can be used as an energy source, for example, by fermentation and biodrying. In addition, bioproducts can be obtained by means of physicochemical, biotechnological, or combined treatments.

Conclusions

The incorporation of citrus and green sugarcane byproducts favors the process of conventional composting byproducts such as bagasse, filter mud, and ash, which can take more than 90 days to show signs of degradation. Considering the current situation of the country, acid soils, highly degraded by monocultures and waste burning and the technological, social, and economic conditions of citrus and sugarcane producers, as well as processing plants, composting is a viable option to valorize the byproducts obtained from these agroindustries. The pile composting method can be effectively applied, and it lasts from 3 to 4 months depending on the type of structuring agent used as bagasse and the composition of the residues. On the other hand, the final product has ideal characteristics to be used as a soil improver in the fields or to be marketed as a peat moss or compost substitute of coffee pulp and filter mud. This is because the pH (9,1), C/N ratio (27,5), organic matter (65,5%), carbon (38%) and macro- and microelements show improved characteristics in relation to the regional compost or filter mud from sugar mill, given that the product converts residues into biofertilizer for use in food production and minimizes environmental pollution.

Acknowledgements

The authors are grateful for the support provided by the Doctorate Program in Agricultural Sciences of the University of Veracruz, México, and the National Council of Science and Technology (CONACyT).

References

Aguilar R. N. (2010). Azúcar, coproductos y subproductos en la diversificación de la agroindustria de la caña de azúcar. *Revista VIRTUALPRO*, 106.

Andreão, P. V., Suleiman, A. R., Cordeiro, G. C., and Nehdi, M. L. (2019). Sustainable use of sugarcane bagasse ash in cement-based materials. *Green Materials*, *7*(2), 61-70. 10.1680/jgrma.18.00016

Behzad S. and Keikhosro K. (2018). Citrus processing wastes: Environmental impacts, recent advances, and future perspectives in total valorization. *Resources, Conservation and Recycling, 129*, 153-167. 10.1016/j.resconrec.2017.10.032

Boukroufa, M., Boutekedjiret, C., Petigny, L., Rakotomanomana, N., and Chemat, F. (2015). Bio-refinery of orange peels waste: a new concept based on integrated green and solvent free extraction processes using ultrasound and microwave techniques to obtain essential oil, polyphenols and pectin. *Ultrasonics Sonochemistry, 24*, 72-79. 10.1016/j.ultsonch.2014.11.015

Bora, H., Kamle, M., Mahato, D. K., Tiwari, P., and Kumar, P. (2020). Citrus Essential Oils (CEOs) and Their Applications in Food: An Overview. *Plants, 9*(3), 357. 10.3390/plants9030357

Caballero, E. and Soto, C. (2019). Valorization of agro-industrial waste into bioactive compounds: techno-economic considerations. In Bastidas-Oyanedel, J-. R. and Schmidt, J. E. (Eds.) *Biorefinery* (pp. 235-252). Cham, Switzerland: Springer. 10.1007/978-3-030-10961-5_10

Cáceres, R., Coromina, N., Malińska, K., and Marfá, O. (2015). Evolution of process control parameters during extended co-composting of green waste and solid fraction of cattle slurry to obtain growing media. *Bioresource Technology 179*, 396-406. 0.1016/j.biortech.2014.12.051

Cai, L., Chen, T. B., Gao, D., Zheng, G. D., Liu, H. T., and Pan, T. H. (2013). Influence of forced air volume on water evaporation during sewage

sludge biodrying. *Water Research 47*(13), 4767-4773. 10.1016/j.watres.2013.03.048

Carrizo, M. E., Alesso, C.A., Cosentino, D., and Imhoff, S., (2015). Aggregation agents and structural stability in soils with different texture and organic carbon content. *Scientia Agricola, 72*, 75-82. 10.1590/0103-9016-2014-0026

Chen, Y. N., Zhou, W., Li, Y. P., Zhang, J. C., Zeng, G. M., Huang, A. Z., Huang, J. X. (2014). Nitrite reductase genes as functional markers to investigate diversity of denitrifying bacteria during agricultural waste composting. *Environmental Biotechnology, 98*, 4233-4243. 10.1007/s00253-014-5514-0

Chowdhury, A. K. M. M. B., Akratos, C. S., Vayenas, D. V., and Pavlou, S. (2013). Olive mill waste composting: a review. *International Biodeterioration and Biodegradation, 85*, 108-119. 10.1016/j.ibiod.2013.06.019

Cole, A. J., Roberts, D. A., Garside, A. L., de Nys, R., and Paul, N. A. (2016). Seaweed compost for agricultural crop production. *Journal of Applied Phycology, 28*, 629-642. 10.1007/s10811-015-0544-2

CONADESUCA (2019). *Datos de la Zafra 2018/2019*. http://www.conadesuca.gob.mx

Drinkwater, L. E. and Snapp, S. (2007). Nutrients in agroecosystems: rethinking the management paradigm. *Advances in Agronomy*, *92*, 163-186. 10.1016/S0065-2113(04)92003-2

Edgerton, M.D. (2009). Increasing Crop Productivity to Meet Global Needs for Feed, Food, and Fuel. *Plant Physiology, 149*(1), 7-13. 10.1104/pp.108.130195

Fourti, O. (2013). The maturity tests during composting of municipal solid wastes. *Resources, Conservation and Recycling 72*, 43-49. 10.1016/j.resconrec.2012.12.001

Gabhane, J., William, S.P., Bidyadhar, R., Bhilawe, P., Anand, D., Vaidya, A.N., and Wate, S. R. (2012). Additives aided composting of green waste: effects on organic matter degradation, compost maturity, and quality of the finished compost. *Bioresource Technology, 114*, 382-388. 10.1016/j.biortech.2012.02.040

Haug, T. R. (1993). The Practical Handbook of Compost Engineering. Boca Ratón, FL: Lewis Publisher, CRC Press LCC.

Kumar, R., Verma, D., Singh, B. L., Kumar, U., and Shweta. (2010). Composting of sugarcane waste by-products through treatment with microorganisms and subsequent vermicomposting. *Bioresource Technology 101*, 6707-6711. 10.1016/j.biortech.2010.03.111

Joglekar, S. N., Pathak, P. D., Mandavgane, S. A., and Kulkarni, B. D. (2019). Process of fruit peel waste biorefinery: a case study of citrus waste biorefinery, its environmental impacts and recommendations. *Environmental Science and Pollution Research*, *26*(34), 34713-34722 10.1007/s11356-019-04196-0

Lim, S. L., Lee, L. H., and Wu, T. Y. (2016). Sustainability of using composting and vermicomposting technologies for organic solid waste biotransformation: recent overview, greenhouse gases emissions and economic analysis. *Journal for Cleaner Production, 111*, 262–278. 10.1016/j.jclepro.2015.08.083

Lin, C. S. K., Pfaltzgraff, L.A., Herrero-Davila, L., Mubofu, E. B., Abderrahim, S., and Clark, J. H. (2013). Food waste as a valuable resource for the production of chemicals, materials and fuels: current situation and global perspective. *Energy and Environmental Science, 6*, 426-464. 10.1039/c2ee23440h

López, G. J. A., Suárez, E. F., Vargas, G. M. C., López, M. J., Jurado, M. M., and Moreno, J. (2015). Dynamics of bacterial microbiota during lignocellulosic waste composting: studies upon its structure, functionality and biodiversity. *Bioresource Technology, 175*, 405-416. 10.1016/j.biortech.2014.10.123

Makan, A. (2015). Windrow co-composting of natural casings waste with sheep manure and dead leaves. *Waste Management, 42*, 17-22. 10.1016/j.wasman.2015.04.019

Mamma, D. and Christakopoulos, P. (2014). Biotransformation of citrus by-products into value added products. *Waste and Biomass Valorization, 5*, 529-549. 10.1007/s12649-013-9250-y

Mohee, R., Mauthoor, S., Bundhoo, Z. M. A., Somaroo, G., Soobhany, N., and Gunasee, S. (2015). Current status of solid waste management in small island developing states: A review. *Waste Management, 43,* 539-549. 10.1016/j.wasman.2015.06.012

Muscolo A., Papalia, T., Settineri G., and Mallamaci C., (2018). Are raw materials or composting conditions and time that most influence the maturity and/or quality of composts? Comparison of obtained composts on soil properties. *Journal for Cleaner Production 193*, 93-101. 10.1016/j.jclepro.2018.05.204

Nwanze, K. and Clark, O. G. (2019). Optimizing Heat Extraction from Compost. *Compost Science and Utilization*, 1-10. 10.1080/1065657X.2019.1686443

Panwar, D., Panesar, P. S., and Chopra, H. K. (2019). Recent Trends on the Valorization Strategies for the Management of Citrus By-products. *Food Reviews International*, 1-30. 10.1080/87559129.2019.1695834

Ortiz-Sánchez, M., Solarte-Toro, J. C., Orrego-Alzate, C. E., Acosta-Medina, C. D., and Cardona-Alzate, C. A. (2020). Integral use of orange peel waste through the biorefinery concept: an experimental, technical, energy, and economic assessment. *Biomass Conversion and Biorefinery*, 1-15. 10.1007/s13399-020-00627-y

Procaña (2019). *Subproductos de la caña*. http://www.procana.org/new/quienes-somos/subproductos-y-derivados-de-la-ca%C3%B1a.html#

Raveh, E., Goldenberg, L., Porat, R., Carmi, N., Gentile, A., and La Malfa, S. (2020). Conventional Breeding of Cultivated Citrus Varieties. In Gentile, A., La Malfa, S., Deng, Z. (Eds.) *The Citrus Genome* (pp. 33-48). Cham, Switzwerland: Springer. 10.1007/978-3-030-15308-3_4

Ravindran, B., and Sekaran, G. (2010). Bacterial composting of animal fleshing generated from tannery industries. *Waste Management, 30*(12), 2622-2630. 10.1016/j.wasman.2010.07.013

Sánchez, M. M. A., Serramiá, N., Civantos, C. G., Fernández, H. A., and Roig, A. (2010). Greenhouse gas emissions during composting of two-phase olive mill wastes with different agroindustrial by-products. *Chemosphere, 81*, 18-25. 10.1016/j.chemosphere.2010.07.022

Satari B. and Karimi K. (2018). Citrus processing wastes: Environmental impacts, recent advances, and future perspectives in total valorization. *Resources, Conservation and Recycling, 129*, 153-167. 10.1016/j.resconrec.2017.10.032

SEDARPA, (2019). *Estadísticas agrícolas*. http://www.veracruz.gob.mx/agropecuario/cierre-2017-estadisticasagricolas/

Shafawati, S. N. and Siddiquee, S. (2013). Composting of oil palm fibres and Trichoderma spp. as the biological control agent: a review. *International Biodeterioration and Biodegradation, 85*, 243-253. 10.1016/j.ibiod.2013.08.005

Singh, J. and Kalamdhad, A. S. (2014). Potential for composting of green phumdi biomass of Loktak lake. *Ecological Engineering, 67*, 119-126. 10.1016/j.ecoleng.2014.03.086

Taghizadeh-Alisaraei, A., Hosseini, S.H., Ghobadian, B., and Motevali, A. (2017). Biofuel production from citrus wastes: a feasibility study in Iran. *Renewable and Sustainable Energy Reviews, 69*, 1100-1112. 10.1016/j.rser.2016.09.102

Tang, Z., Yu, G., Liu, D., Xu, D., and Shen, Q. (2011). Different analysis techniques for fluorescence excitation-emission matrix spectroscopy to assess compost maturity. *Chemosphere 82*, 1202-1208. 10.1016/j.chemosphere.2010.11.032

Teixeira, D. L., de Matos, A. T., and Melo, E. D. C. (2015). Resistance to forced airflow through layers of composting organic material. *Waste Management, 36*, 57-62. 10.1016/j.wasman.2014.12.003

Tian, Y., Chen, L., Gao, L., Michael Jr., F. C., Keener, H. M., Klingman, M., and Dick, W. A. (2012). Composting of waste paint sludge containing melamine resin and the compost's effect on vegetable growth and soil water quality. *Journal of Hazardous Materials, 243*, 28-36. 10.1016/j.jhazmat.2012.09.013

Tuomela, M., Vikman, M., Hatakka, A., and Itävaara, M. (2000). Biodegradation of lignin in a compost environment: a review. Bioresource Technology 72, 169-183. 10.1016/S0960-8524(99)00104-2

Wei, Y., Li, J., Shi, D., Liu, G., Zhao, Y., and Shimaoka, T. (2017). Environmental challenges impeding the composting of biodegradable municipal solid waste: a critical review. *Resources, Conservation and Recycling, 122*, 51-65. 10.1016/j.resconrec.2017.01.024

Yang, R., Su, Y.-Z., Wang, T., and Yang, Q. (2016). Effect of chemical and organic fertilization on soil carbon and nitrogen accumulation in a newly cultivated farmland. *Journal Integrative Agriculture 15*(3), 658-666. 10.1016/S2095-3119(15)61107-8

Zema, D. A., Calabrò, P. S., Folino, A., Tamburino, V., Zappia, G., and Zimbone, S. M. (2018). Valorisation of citrus processing waste: A review. *Waste management*, *80*, 252-273. 10.1016/j.wasman.2018.09.024

Zhang, L. and Sun, X. (2016). Improving green waste composting by addition of sugarcane bagasse and exhausted grape marc. *Bioresource Technology,* 2018, 33-343. 10.1016/j.biortech.2016.06.097

Zhao, B., O'Connor, D., Zhang, J., Peng, T., Shen, Z., Tsang, D. C. W., and Hou, D. (2018). Effect of pyrolysis temperature, heating rate, and residence time on rapeseed stem derived biochar. *Journal of Cleaner Production, 174*, 977-987. 10.1016/j.jclepro.2017.11.013

Research Article / Industrial engineering http://dx.doi.org/10.15446/ing.investig.v40n3.81696

Analysis of the Relationship Between IT and Industry 4.0 Technologies with Internationalization and Business Performance

Análisis de la Relación de TI y de las Tecnologías de la Industria 4.0 con la Internacionalización y Rendimiento Empresarial

Omar Alexánder León García[1] and Eliana Rocio Baez Landeros[2]

ABSTRACT

The impact of information technologies (IT) on organizations has attracted the attention of professionals and academics in recent years. In this article, since Industry 4.0 technologies have not yet been implemented at the same level as general purpose information technologies, they are analyzed separately to understand their relationship with internationalization and business performance. Through a survey of 168 companies in the city of Bogotá, Colombia, different regression analyses were performed according to a research model proposed to estimate the effects of the determined variables. It was evident that the companies with a greater use of IT present a higher degree of internationalization. Likewise, the proposed model demonstrates the direct relationship between the use of IT and Industry 4.0 technologies and business performance. The software R studio was used for statistical analysis.

Keywords: IT, Industry 4.0, internationalization, business performance

RESUMEN

El impacto de las tecnologías de la información (TI) en las organizaciones ha atraído la atención de profesionales y académicos en los últimos años. En este artículo, dado que las tecnologías de la Industria 4.0 aún no han sido implementadas al mismo nivel que las tecnologías de la información de uso general, se analizan de manera separada para entender su relación con la internacionalización y el rendimiento empresarial. A través de una encuesta a 168 empresas en la ciudad de Bogotá, Colombia, se realizaron diferentes análisis de regresión de acuerdo con un modelo de investigación propuesto para estimar los efectos de las variables determinadas. Se hizo evidente que las empresas con mayor uso de TI presentan un mayor grado de internacionalización. Asimismo, el modelo propuesto demuestra la relación directa entre el uso de TI y las tecnologías de la Industria 4.0 con el rendimiento empresarial. Para el análisis estadístico se utilizó el software R studio.

Palabras clave: TI, Industria 4.0, internacionalización, rendimiento empresarial

Received: August 13th, 2019
Accepted: August 13th, 2020

Introduction

The contribution of IT in different organizational processes is well known. This is often analyzed from a theoretical perspective in terms of the investment that companies should make in this resource in order to increase their productivity, because the market is requesting it or the competition is already doing it. However, some companies are not clear about the true effect of these technologies within their organization. For this reason, quantifying the impact that IT has on business processes is essential to analyze its potential and challenges, and also to propose methodologies for its proper implementation. In this sense, the impact of IT on companies substantially improves business efficiency especially thanks to the use of the Internet, since it provides opportunities in different fields such as the development of new products and services. As a result, these technologies offer a wider market that contributes to distribution channels and customer service regardless of geographical distances. They facilitate contact between clients and suppliers, and provide improvements in the processing and storage of organizational information (González, Alfaro-Azofeifa, and Alfaro-Chamberlain, 2005).

Currently, new technologies developed under the concept of Industry 4.0 have been incorporated. The use of these resources is a fundamental factor for the optimization of company processes. These technologies have evolved in different aspects, such as 3D printing, reverse engineering, big data, and artificial intelligence, which generates a huge

[1] Production Engineer, Universidad EAN, Colombia. Ph.D. mechanical engineering and electrical energy, Mondragon Unibersitatea, Spain. Affiliation: Professor, Compensar Unipanamericana Fundación Universitaria (GIIS), Colombia. E-mail: omarleon@unipanamericana.edu.co
[2] Systems Engineer, Compensar Unipanamericana Fundación Universitaria, Colombia. Affiliation: Research intern, Compensar Unipanamericana Fundación Universitaria (GIIS), Colombia. E-mail: erbaez@unipanamericana.edu.co

How to cite: León, O. A. and Baez, E. R. (2020). Analysis of the Relationship Between IT and Industry 4.0 Technologies with Internationalization and Business Performance. *Ingeniería e Investigación*, *40*(3), 89-99. 10.15446/ing.investig.v40n3.81696

impact of transformation in the manufacturing industry and services. They also facilitate consumer analysis, offering new business alternatives and giving companies the ability to adapt to market changes. (Lasi, Fettke, Kemper, Feld, and Hoffmann, 2014).

In this context, a new business management model based on technology is proposed (Lu, Riddick, and Ivezic, 2016), making information processing more effective and efficient, meeting the needs of the value chain from product design to customer service and involving all the parts who are interested in the product life cycle. (Garbi and Loureiro, 2013). With this advancements, one of the main goals of the use of IT in terms of cost optimization and reduction of deadlines is expected to be met, thus reaching new markets and mitigating risks in organizations (Cerezo-Narváez, Otero-Mateo, Rodríguez-Pecci, and Pastor-Fernández, 2018). According to this, companies have turned these technologies into a resource of great strategic influence for the improvement of business performance, generating changes in the increase of their productivity and in their management model (Baena, Cano, Jarrin, and Pérez, 2014).

The use of IT has allowed companies to reach new international markets, offering benefits in efficiency, saving time and travel expenses, being an intermediary of assertive communication between two stakeholders, facilitating the construction of relationships, and providing negotiation alternatives (Baena et al., 2014). Internationalization is an opportunity for companies to be projected totally or partially into an international environment and different commercial flows (Leandro, 2009). The influence of IT on this activity can facilitate exports and favor the development of foreign trade operations that were only possible for multinational companies (Barber and Cobos, 2002).

To determine whether these aspects truly affect an organization, it is important to analyze their relationship with business performance. Studies on the relation between the use of IT and business performance have been especially aimed at understanding the so-called Solow paradox, which claims that the computer age can be seen everywhere but in the productivity statistics (Wilkesmann and Wilkesmann, 2018). Therefore, it has been considered important to determine the type of relation between IT, the technologies of Industry 4.0, and business performance.

It is important to understand that, due to the size of the company and the fact that some of the technologies of the industry 4.0 are still expensive, the penetration of these technologies has not been widespread. Although today's market offer has brought a wider range of solutions with options of all kinds, it is necessary need to measure the use of IT (general use technologies) and Industry 4.0 technologies separately in this study, while also avoiding the bias towards the low amount of information obtained from the analyzed companies.

That said, the objective of this study is to contribute to the literature and the business sector through the empirical demonstration of the relation between the use of IT and Industry 4.0 technologies in the degree of internationalization and business performance. This analysis aims to demonstrate that companies who use these technologies to a greater degree will also increase in the afore-mentioned aspects.

The document begins with the theoretical background on the relation between IT, Industry 4.0 technologies, internationalization, and business performance, on which the fundamentals of this study are based. The second part presents the development, methodology, and the model used for analysis. Then, the results regarding the relation between the identified variables are presented. The study ends with the conclusions and contributions that can be offered in this field of research.

Framework

The conceptual and theoretical elements that guide this research are presented below, starting with the relationship of information technologies with internationalization, continuing with its relationship with business performance, and ending with the relationship of Industry 4.0 technologies with business performance.

IT and internationalization

Internationalization is defined as the entry of a company into new business beyond geographical limits. Companies that implement this expansion strategy can use IT to coordinate their resources and capacities through different markets that facilitate the realization of economies of scope, thus allowing companies to operate in several markets simultaneously. (Mai and Ness, 1999; Ray, Xue, and Barney, 2013). Thus, the bet on IT that companies make plays in favor of the use of some specific firm assets across the borders of the country (Chari, Devaraj, and David, 2007).

Some research shows that the development of efficient virtual relations through computers and mobile devices is a significant factor in the creation of more effective work relationships, regardless of distance (Aren, José, and Baena Rojas, 2013). The use of IT leads companies to build customer-supplier relationships during the process in question. As a result, this allows a better contact between the negotiators, providing multiple communication options (visual, verbal, or written), focusing on the importance of negotiation, saving travel expenses, and allowing better access to communication (Aren et al., 2013). In this way, through the interaction and behavior of these electronic media, companies integrate IT within their business structure as a channel of negotiation in internationalization, thus yielding more efficient results as opposed to traditional face-to-face negotiation and obtaining a faster offer.

The reviewed studies show that the impact on internationalization is a positive function of the level of IT investment made by the company (Chari et al., 2007; Liu and Ravichandran, 2008; McElheran, 2014). This impact could be significantly positive (for companies with high IT investment), significantly negative (for companies with low IT investment), or neutral (for companies with medium level with IT investment) (Chari et al., 2007).

Empirical research mostly relates internationalization to IT, as its expense or investment significantly complements this strategy and is generally related to business results. (Chari, Devaraj, and David, 2008; Esteben, 2007; Hu, Zhang, and Teng, 2011; Liu and Ravichandran, 2008; Ravichandran, Liu, Han, and Hasan, 2009; Shin, 2006). In these studies, investment in IT is evidenced as the most frequent way to measure this variable for analysis. However, this investment has been treated by the literature as a black box, and its impact on measured performance is considered within a narrow context, while the level of adoption and use of IT can place said investment in perspective (Devaraj and Kohli, 2003). For this impact to be evident within the processes of the value chain, it is considered appropriate consider the use of these technologies as a unit of analysis, which few studies are evident and has been suggested in others. (Devaraj and Kohli, 2003; Sandulli, Fernández-Menéndez, Rodríguez-Duarte, and López-Sánchez, 2012).

Hypothesis (H1.1). Companies with activity abroad (internationalization) have a higher level of IT use than companies that do not have activity abroad. **(H1.2).** Companies with a greater use of IT present a higher degree of internationalization.

IT and Business Performance

Since the end of twentieth century, there is a strong impetus for companies to adopt the Internet as a way to improve their business behavior. This has spread throughout most sectors of the economies of developed countries (Alwahaishi and Snášel, 2013). In this sense, some of the improvements that companies have experienced with the incorporation of IT refer to the coordination among collaborators, the ability to catalogue and share the base knowledge of the company, the expansion of traditional boundaries of the business areas, and the efficiency due to the processing of information and the promotion of innovation (Dewett and Jones, 2001). This reveals that IT is a key element within companies to ensure economic growth (K. S. Ravichandran and Lorenz, 2007), and they are a way to remain competitive by improving productivity, profitability, and quality of operations (Devaraj and Kohli, 2003).

Currently, companies can see the continuous search to achieve compliance with their objectives and improve their business performance. For this, the use of IT offers a positive impact on these factors, not only in terms of productivity, but also at the level of processes, quality of services, customer satisfaction, and innovation in the market (Brynjolfsson and Hitt, 2000). In addition, the use of IT meets the objective of improving business performance in aspects such as business strategies, service design, policy creation in the sector, implementation of informative websites, and other search alternatives for different partners and customers in the market (Barbosa and Ayala, 2017). Consequently, the effects of IT can influence companies not only in sales growth sales, but also in process standardization and optimization.

According to Patalas-Maliszewska (2012), the implementation of these technologies depends on the intention of a company to achieve its strategic objectives (e.g. in the implementation of an ERP). In this sense, the aim is that through the implementation of information technologies companies can increase their productivity, profitability, efficiency of client relationships, human resources management effectiveness and improve the reliability of the financial reports.

On the other hand, a company's ignorance of these technologies prevents its growth, since the benefits of its use are not recognized. These benefits include the fact that IT leads to an increase in productivity by reducing costs; thus allowing companies to increase the quality and output of new products. (Brynjolfsson and Hitt, 2000). In this sense, the innovations achieved by IT have offered opportunities to improve processes and develop new business models and applications (Kim, 2013). IT also helps companies expand their competitive advantage potential by allowing them to carry out activities and products at a lower cost (Porter, 2015), extending the marketing potential of the company, facilitating the formation of business networks, and streamlining the transmission of information and knowledge regardless of distances (Vilaseca and Torrent, 2003).

Hypothesis (H2). The relation between IT and the performance of the company will depend partially on its IT usage, thus, a greater investment will encourage greater business performance.

Industry 4.0 Technologies and business performance

The term 'Industry 4.0' was coined in Germany with reference to the digital transformation of the industry, which came to be known as 'intelligent factory' or 'industrial internet', as well as to name the fourth industrial revolution (Morales, España, Zárate, González, and Frías, 2017). Thus, Industry 4.0 constitutes an easy alternative to adapt to changes in accordance with market demand, based on the interconnection of all elements, generating great satisfaction for customers and improving performance within companies.

The advancement of technology has accompanied the development of the industry from the early adoption of mechanical systems to support production processes, up to the current, highly automated assembly lines. It responded and adapted to the current, dynamic requirements and demands from the market. Under the concept of Industry 4.0, amazing growth in the advancement and adoption of information technology and social networks has increasingly influenced the perception of consumers regarding product innovation, quality, variety, and delivery speed (Lee, Kao, and Yang, 2014). The advancement of these technologies will allow the development of smart factories through innovation at each stage of the value chain (Matthyssens, 2019), as well as having a positive role in companies sustainable business performance and opening great opportunities for industries (Büchi, Cugno, and Castagnoli, 2020; Haseeb, Hussain, Ślusarczyk, and Jermsittiparsert, 2019).

According to industry and research experts, this industrial revolution was caused by the Internet, which allowed communication between humans and machines in cybernetic

systems through large networks (Frank, Dalenogare, and Ayala, 2019), giving rise to digital transformation within companies. This transformation has been directly supported in the development and use of technologies such as Big Data, with the ability to collect, store, and analyse large amounts of data; or the simulation of virtual industrial plant models with specialized engineering that allow management and control, as well as ensuring the quality of processes and products. Other technologies such as 3D printing, internet of things, cloud computing, wireless payment modes or recent advances in artificial intelligence, make the industry 4.0 have an increasingly greater penetration in organizations. The use of these technologies makes the flow of information between the production line, the level of business management and the management of the supply chain be more transparent and organized, reducing labour costs and providing a better working environment (Fatorachian and Kazemi, 2020; Lee et al., 2014).

Another important aspect of advancement in the technologies of Industry 4.0 is the integration of robotics, with the ability to be programmed to execute tasks at specific times and dates (Bahrin, Othman, Azli, and Talib, 2016). In this way, these advances have covered various fields such as biotechnology, genetics, and nanotechnology, giving a way to the development of artificial intelligence and additive manufacturing (3D printing), which implies a qualitative and significant leap for the companies in their value management chain (Mon, Del Giorgio, De María, Figuerola, and Querel, 2018).

Thereupon, companies have been structuring themselves in manufacturing by using computers connected to the Internet and in advanced analysis platforms, where they process data to produce final products. Likewise, the communication offered by these technologies is synchronized by means of devices that allow the connection of production plants, offices, and dependencies, where they can share information and communicate in real time, regardless of their location.

On the other hand, due to the size of the company and the high cost of some of these technologies, their adoption has not been easy. Such is the case of 3D printing, which in the beginning was a very expensive manufacturing and whose materials were very scarce. However, there are several types of materials compatible with this form of manufacture that are currently available. Nowadays, the market offer has brought a wider range of solutions with options of all kinds, which is why access to them has begun to be democratized. In this sense, SMEs implement the use of technologies that are within their reach, such as cloud computing, which has a great impact on the management of their information involving reduction of capital and operating costs regarding aspects such as database management, email management, business intelligence, customer relationship management, and business resource planning, among others. Through these tools, companies aim to demonstrate a competitive advantage in the manufacturing market, essentially in the smart industry for a better management and a positive impact on their business performance (Morales et al., 2017).

Hypothesis (H3). In accordance with the above, we can define the following as a study hypothesis: companies that have a greater use of Industry 4.0 technologies improve their business performance.

Thus, a general research model is presented which aims to relate the study variables and verify the hypotheses. These variables are the support for the design of the research instruments that were applied to the study sample. This model shows that the IT variables and technologies of Industry 4.0 affect the degree of internationalization and performance of companies. This can be seen in Figure 1.

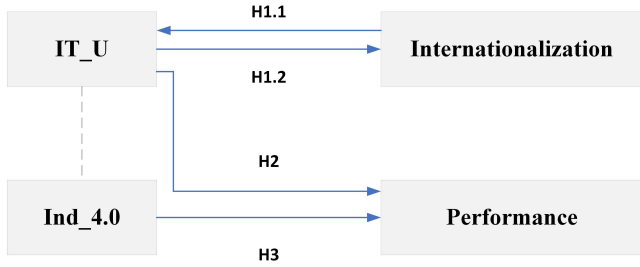


Figure 1. Proposed Research Model.
Source: Authors

Methods

This study was carried out through the implementation of a research instrument aimed at managers of SMEs established in the city of Bogotá D. C., Colombia. To collect the information, a structured questionnaire was used as primary source. This instrument was sent to managers of companies (to principal directors or to whom they delegate) established in the city of Bogotá, Colombia.

The sample was formed according to the obtained response rates and according to the characteristics defined in the population, discarding those companies that, by answering the questionnaire, did not comply with these characteristics. In total, 195 questionnaires were received, out of which 27 were cancelled because they were not properly completed. The final sample consisted of 168 companies that answered the questionnaire submitted in online and physical format, during the period from February to September, 2018 (95% confidence level, pq = 0,50: 10,5%). From the obtained sample, 26,2% of the companies classified as Other Services; 19,1% as Information and Communications; 17,4% as Manufacturing Industries; 9,5% as Professional, Scientific, and Technical Activities; 5,2% as Accommodation and Food Services; and 4,3% as Construction with.

For the empirical analysis, the data set of the sample was analyzed by means of statistical tests according to the variables used. For the first Hypothesis (11) a bivariate test was performed to verify the significant differences between

the uses of IT in two identified company categories. For this case, because the variable IT_U did not behave normally in the two categories, it was necessary to use a non-parametric test (Whitney U). Subsequently, the regression of ordinary least squares (OLS) estimated the relation between the use of IT and Industry 4.0 technologies with the degree of internationalization and company performance (Hypotheses 1.2, 2 and 3).

Measurement of variables

IT Use variable (*IT_U*)

The variable IT_U refers to the use that a company gives to the Information Technologies that it has implemented. Managers were inquired to determine how much a set of selected technological tools were being used according to the reviewed studies, including both IT and communications elements. The list of selected tools was suggested by the International Telecommunication Union (ITU) and the OECD, and identified in other studies (Albarracín, Erazo, and Palacios, 2014; Guzmán, Serna, de Lema, Enríquez, and Adame, 2010). This list includes the use of computers, internet access, web presence, e-mail, VoIP, social networks, e-recruitment, e-learning, online services to customers, e-commerce, electronic banking, ERP, SCM, CRM, technological surveillance, and e-government.

This variable consists of the arithmetic mean obtained from a five-point Likert scale, in which the directors indicated the level of use of each of the presented technologies. To validate this measure, the reliability of the scale was verified by Cronbach's alpha (0,932) and KMO (0,889), and communalities higher than 0,4.

Use of Industry 4.0 Technologies (*Ind_4.0*)

To measure this variable, some of the main technological tools of Industry 4.0 were identified (3D printing, cloud/client architecture, big data, intelligent machines, augmented reality, wireless payments, IoT). To this effect, the directors were asked to indicate what the level of use for each of the selected tools is. The list of tools was suggested by trends recognized by ITU studies. This variable consists of the arithmetic mean obtained from a five-point Likert scale. To validate this measure, the reliability of the scale was verified by Cronbach's alpha (0,842) and KMO (0,843), and communalities greater than 0,6.

Business performance (*Performance*)

Business performance refers to the level of success of a company (Chelliah, Sulaiman, and Yusoff, 2010). While performance in most cases is recognized in accordance with quantifiable measures such as income, sales, market share, etc., other measures are customer satisfaction, skill development, improvement of the workflow, etc. (Khan, Khalique, and Nor, 2014). This makes performance a multidimensional concept which depends on both financial and non-financial indicators, but the number of indicators must be consistent with the strategic management style of the organizations (Muralidharan, 1997).

In this sense, the performance of the company WAS measured through the scale developed by Schuh, Potente, Wesch-Potente, Weber, and Prote (2014). This scale has been tested empirically and has produced solid reliability and validity. In addition, this measurement of the performance of the company involves data reported by the manager, who can examine the financial and non-financial dimensions of performance within the organization (sales growth, income growth, growth in the number of employees, net profit margin, product and service innovation, innovation process, product/service quality, variety of products/services, adoption of new technologies, and customer satisfaction). A Likert scale of five points was used to evaluate the performance in the most recent years of the companies. The use of this measure is appropriate in this study because it is convenient when studying small and medium size companies (Dess and Robinson, 1984) since they allow a broader definition of the performance of the company (Cronbach's alpha 0,890, KMO 0,791, and communalities higher than 0,6).

Internationalization

The internationalization variable was measured in two ways. There was a dichotomous categorical measure that identified two categories of companies: those that have international activity (1) and those that do not perform international activities (0). It was expected that companies with international activity have a greater degree of IT use. A second measure of this variable was a composite index suggested by Sullivan (1994) through the use of two indicators: the intensity of operations abroad (external sales ratio over firm total sales) and its geographic scope (number of countries in which the companies operate), these being the two most commonly used indicators to measure the scope of international operations (Tallman and Li, 1996). These two measures allow the proposed analyses to be carried out.

Results and Discussion

IT and Internationalization

Under Hypothesis 1.1, variables IT_U and internationalization were contrasted. To analyze the relationship between the variables of this hypothesis, a bivariate test was used in order to verify the significant differences between the use of IT in the two company categories (companies that have international activity and companies that do not present international activity). By means of the test for independent samples, the equality of means of the dependent variable in each of the defined categories was analyzed. If the assumption of equality of means is right it is assumed that the variables are independent, since the mean would be the same regardless of the company group. If the null hypothesis can be rejected, it is assumed that there is a relation between the variables.

Before this analysis, the distribution of the populations should be revised, which means that the samples must come from

populations distributed in a normal way. According to the size of the sample in this study, a Kolmogorov-Smirnov statistic was obtained below the level of significance established for the No category (0,017) and higher for the Yes category (0,200), with which it can be stated that the variable does not behave normally in both categories and makes it necessary to use a non-parametric test. This is shown in Table 1. One of the most used tests in this case is the Whitney U. It is a nonparametric test with which differences are identified between two populations based on the analysis of two independent samples.

Table 1. Descriptive analysis

	Type	N	Mean	Std. Deviation	Std. Error
IT_U	No	105	3,3679	0,6708	0,0651
	Yes	63	3,9120	0,7958	0,1021

Source: Authors

The average IT_U of the companies with international activity was higher than those without, in the sense proposed in Hypothesis 1.1. To check if this difference is explanatory, the value of asymptotic (bilateral) significance of the Z statistic (0,000) was analyzed. The Whitney U analysis led us to conclude that the null hypothesis was rejected and the mean IT_U is similar in both groups. With these values, it can be indicated that the companies that present a higher level of IT use are more likely to carry out activity abroad (Hypothesis 1.1). We decided to complement this analysis by determining if the level of IT use has a direct relation with the degree of internationalization. In this sense, Hypothesis 1.2, which defines this association, was raised. For this analysis, only companies with international activity (63) were taken to measure their degree of internationality, as stated above, through a composite index proposed by Sullivan (1994) and the use of two indicators: the intensity of operations abroad and their geographic scope. Specifically, principal components factorial analysis was used to reduce the two variables. This multivariate method suggests that these relations can be explained from a series of variables called factors, where the number of factors is much smaller than that of the actual variables (Piera and Carrasco, 2010).

For factor analysis to provide a useful composite index, it is necessary to correlate the two variables and load them into a single factor. For this case, the two variables had a correlation of 0,473 ($p < 0,00$) and were loaded in a component that explains 73,6% of the cumulative variance. Each variable presented a load of 0,858 in the extracted component. These results from the factor analysis indicate that composite measure is a useful index of internationalization. Having obtained the values of the variables, we proceeded to perform the correlation analysis of the two variables of interest, IT_U and degree of internationalization. The Correlation Coefficient (Pearson) had a positive value (0,336) with a statistically significant coefficient ($p = 0,008$), so it can be concluded that both variables are associated in the population from which the analysed sample comes, and that association shows a direct correlation. With these correlation values, it was possible to complement the statistical study through linear regression analysis to evaluate this relation. To this effect, the following model was considered, using a linear regression by OLS (ordinary least squares). It is described using Equation (1):

$$\text{Internationalization level} = b_0 + b_1 IT_{U_i} + b_2 Size_i + b_3 Age_i + \varepsilon_i \quad (1)$$

The regression results (Adjusted $R^2 = 0,096$) indicate that 9,6 % of the variability of the degree of internationalization is associated with the use of IT. In turn, the ANOVA table showed that the variables are linearly related to a level of significance of 0,008 (Durbin-Watson 1,476). Likewise, a non-standard coefficient for UTIC of 0,108 was identified, with VIF values close to 1 (1,000). It is shown in Table 2. With these results, it can be concluded that the two variables are either associated or linearly related in the population from which the sample comes, and that the relation of the variables is direct. These results show that companies with a greater degree of use of IT have a greater degree of internationalization, thus confirming Hypothesis 1.2.

Table 2. Model summary

Model	R	R^2	Adjusted R^2	Std. Error	FChange	Sig. F Change
1	0,333[a]	0,111	0,096	0,24944	7,590	0,008

a. Predictors: (Constant), IT_U

b. Dependent Variable: Internationalization

Source: Authors

The results of these analyses indicate that information technology is a key factor for companies to expand their markets beyond the borders of their home country. Since greater internationalization implies an increase in the number of countries in which the company operates, information processing promoted by IT is especially relevant to take advantage of a greater geographical reach. It is necessary to highlight in the results that, despite a positive and highly significant influence of the use of IT on the degree of internationalization (which is in line with the findings of (Chari et al., 2008; Honarmand, 2011), it should be moderate when extrapolating the results of this analysis, since although they indicate a real relationship between the predictors and the response variable, the adjusted R^2 is low. Similarly, it is important to note that, contrary to what was expected, no significant influence of company age and size was found.

IT and Business Performance

The study assumes that the greater the use of IT, the higher the performance of a company. For this analysis, a correlation analysis was performed between these two variables: IT_U and Performance. It can be seen in Figure 2.

It has been determined that the correlation coefficient (Pearson) has a positive value (0,422) with a statistically significant coefficient ($p = 0,000$), so it can be concluded that both variables are directly associated. With these correlation values, it is possible to complement the statistical study

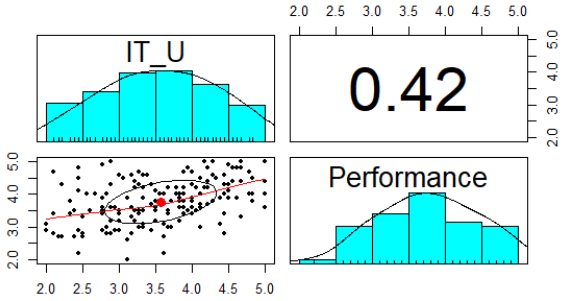


Figure 2. Statistical Analysis of IT_U and Performance.
Source: Authors

by means of a linear regression analysis to evaluate the relation. The Kolmogorov-Smirnov test showed that the data meet the requirements of normality. The following is the analysis model, using a linear regression by OLS. The variables size and age of the company have been included in the regression model for control purposes. This is described using Equation (2)

$$\text{Performance} = b_0 + b_1 IT_{U_i} + b_2 Size_i + b_3 Age_i + \varepsilon_i. \quad (2)$$

The determination coefficient R^2 (0,256) indicates that 25,6% of the variability of business performance is explained by the model. The ANOVA statistic (analysis of variance) has a low value of 0,000, which confirms that the variables are linearly related. This is shown in Table 3. A non-standardized coefficient is also identified for IT_U 0,331, with values close to 1 VIF (1,024) in the variables used, which indicates that multicollinearity did not affect the results. These results show that companies with a higher degree of IT use have a higher level of business performance. Therefore, Hypothesis 2 is confirmed.

Table 3. Model summary

Model	R	R^2	Adjusted R^2	Std. Error	FChange	Sig. F Change
1	0,506[a]	0,256	0,242	0,58071	18,818	0,000

a. Predictors: (Constant), IT_U
b. Dependent Variable: Performance

Source: Authors

With this analysis, it has been possible to statistically demonstrate this relation, concluding that companies that make greater use of IT will see it reflected on a better performance of the businesses they have entered. These results are in agreement with Aydiner, Tatoglu, Bayraktar, and Zaim (2019) regarding the use of technologies and information systems to improve business performance. It is important to clarify that, in order to obtain the expected results, it is necessary to align these technologies with the business strategy so that it has better influence on business performance, which also agrees with Luftman, Lyytinen, and Zvi (2017). As expected, the size variable shows a significant value (0,010), corroborating the importance and influence of this characteristic on the use of technologies and business performance, which, in turn, is in accordance with what was exposed by Spinelli, Dyerson, and Harindranath (2013). In contrast, the age variable does not represent a significant value.

Industry 4.0 Technologies and Business Performance

Hypothesis 3 assumes that, as companies have greater use of Industry 4.0 technologies, business performance will also be higher. For this analysis, a correlation analysis was made between these two variables: Ind_4 and Performance. It can be seen in Figure 3.

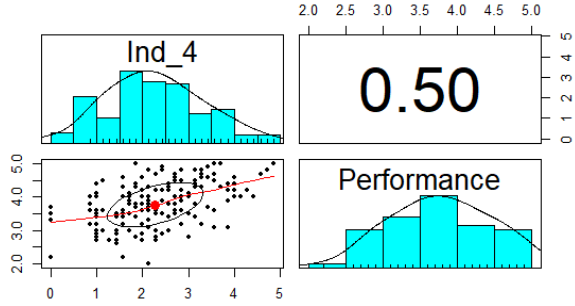


Figure 3. Statistical Analysis of the Industry 4.0 Technology and Performance.
Source: Authors

It has been determined that the correlation coefficient (Pearson) has a positive value (0,500) with a statistically significant coefficient (p = 0,000), so we can conclude that both variables are directly associated. The Kolmogorov-Smirnov test showed that the data meet the requirements of normality. The following is the analysis model, using a linear regression by OLS (ordinary least squares). The variables size and age of the company have been included in the regression model for control purposes. This is described using Equation (3):

$$\text{Performance} = b_0 + b_1 Ind_{4.0_i} + b_2 Size_i + b_3 Age_i + \varepsilon_i. \quad (3)$$

The adjusted R^2 coefficient of determination (0,302) indicates that the model explains 30,2% of the business performance variability. The ANOVA statistic (analysis of variance) has a low value (0,000), which confirms that the variables are linearly related. It also identifies a non-standardized coefficient for IT_U 0,294 (VIF 1,062). This is shown in Table 4. These results show that companies with a higher degree of use of Industry 4.0 technologies have a higher level of business performance, thus confirming Hypothesis 3. These results would be in line with the evidence in the literature claiming that the incorporation of these new technologies to optimize processes of companies, as well as their use within organizations, will have a positive impact on business performance (Dalenogare, Benitez, Ayala, and Frank, 2018; Haseeb et al., 2019).

Regarding the control variables, only company size was found to be significant, which is in line with what was stated by Horváth and Szabó (2019), in that large companies are more prepared to implement Industry 4.0 technologies which provides opportunities for performance measurement and evaluation.

Table 4. Model summary

Model	R	R^2	Adjusted R^2	Std. Error	F Change	Sig. F Change
1	0,561[a]	0,315	0,302	0,55732	25,116	0,000

a. Predictors: (Constant), ind_4.0
b. Dependent Variable: Performance

Source: Authors

To complement this analysis, according to the descriptive results obtained in the sample, the average Industry 4.0 Technology use in the analyzed companies is 1,90 out of 5,0. This indicates that such technologies continue to have low implementation levels, which agrees with other studies (Stentoft, Jensen, Philipsen, and Haug, 2019). By analyzing the use of each of these technologies, and their possible influence on business performance, it is observed that cloud computing (3,1) is the one with the highest valuation. This indicates a commitment by companies to this technology. Likewise, companies have been increasing the use of the Internet of Things (2,09). Contrary to this, the use of autonomous robots (1,25), and additive manufacturing (1,33) present the lowest valuation and show a still incipient development within the companies and industries of the region. One of the main reasons why such technologies have lower use may be high purchase prices and low availability of supplies necessary for their operation.

Conclusions

The results derived from this research contribute to the theories based on the relation between the use of IT and the strategies developed by companies to implement their internationalization activities to improve business performance. The study shows that IT technological resources and Industry 4.0 technologies provide tools for organizations to boost foreign markets and positively stimulate business performance.

The theoretical framework presented in this document contributes to IT managers and those responsible for internationalization strategies in decision-making within the company to leverage these resources, promote greater internationalization of the organization, and improve business performance.

Companies make the decision to internationalize for various reasons (to grow, increase sales, gain prestige or promote an innovative product, among others), but there are still many that consider internationalization a privilege that only large corporations can access. However, SMEs can leverage the use of IT and reap their great benefits to develop strategies that allow them to reach other geographical areas and improve their performance.

It is then evident that the development of new information and communication technologies has allowed companies to get closer to their customers, responding more quickly to their needs and adapting to the dynamic environment of the current economy. These technologies (internet access, e-recruitment, e-learning, online services to customers, e-commerce, electronic banking, ERP, SCM, CRM, IoT, augmented reality, artificial intelligence, 3D printing, etc.) are driving new business models and major changes in the internal structure of organizations that aim for internationalization and performance improvement.

After the application of the analysis model in the variables of this research, we verified that companies with a greater use of IT have a higher rate of internationalization, unlike companies that do not use it within their business model. Thus, the impact of these technologies influences the knowledge of new markets abroad and the opportunity to be identified in an international context.

According to current trends in the use of emerging technologies and the results obtained in research, we found that an increase in business performance is characterized by the use of Industry 4.0 technologies (with cloud computing and Internet of Things being increasingly used), which transitions companies towards an intelligent industry management, while also involving changes in manufacturing and process management for business development and growth.

Future research can focus on the specific analysis of the relation between each of the identified technologies and the internationalization indexes and the financial and non-financial elements of business performance.

Acknowledgements

This paper is the result of an in-progress research project, developed in the GIIS research group and funded by the Compensar Unipanamericana Fundación Universitaria and the company Draco Servicios in the city of Bogotá, Colombia. The addresses are Avenida Calle 32 #17-30 and Carrera 7H bis #159-23, respectively.

References

Albarracín, E. J. G., Erazo, S. C. R., and Palacios, F. C. (2014). Influencia de las tecnologías de la información y comunicación en el rendimiento de las micro, pequeñas y medianas empresas colombianas. *Estudios Gerenciales, 30*(133), 355-364. 10.1016/j.estger.2014.06.006

Alwahaishi, S. and Snášel, V. (2013). Consumers' acceptance and use of information and communications technology: A UTAUT and flow based theoretical model. *Journal of Technology Management and Innovation, 8*(2), 61-73. 10.4067/S0718-27242013000200005

Aren, C., José, A., and Baena Rojas, J. J. (2013). Retos en la implementación de las TIC para el proceso de negociación internacional. *Cuadernos de Administración* (*Universidad del Valle*), *29*(50), 153-163. 10.25100/cdea.v29i50.51

Aydiner, A. S., Tatoglu, E., Bayraktar, E., and Zaim, S. (2019). Information system capabilities and firm performance: Opening the black box through decision-making performance and business-process performance. *International Journal of Information Management, 47*, 168-182. 10.1016/j.ijinfomgt.2018.12.015

Cano, J. A. and Baena, J. J. (2015). Tendencias en el uso de las tecnologías de información y comunicación para la negociación internacional. *Estudios Gerenciales*, *31*(136), 335-346. 10.1016/j.estger.2015.03.003

Bahrin, M. A. K., Othman, M. F., Azli, N. N., and Talib, M. F. (2016). Industry 4.0: A review on industrial automation and robotic. *Jurnal Teknologi, 78*(6-13), 137-143. 10.11113/jt.v78.9285

Barber, J. P. and Cobos, A. (2002). La aceleración del proceso de internacionalización de la empresa: el caso de las international new ventures españolas. *Informacion Comercial Española-Monthly Edition*, 9-22.

Barbosa, D. M. E. and Ayala, A. H. (2017). El uso de las TICs en las PYMES exportadoras. *Dimensión empresarial, 15*(1), 184-205.

Brynjolfsson, E. and Hitt, L. M. (2000). Beyond computation: Information technology, organizational transformation and business performance. *Journal of Economic Perspectives, 14*(4), 23-48. 10.1257/jep.14.4.23

Büchi, G., Cugno, M., and Castagnoli, R. (2020). Smart factory performance and Industry 4.0. *Technological Forecasting and Social Change*, 150, 119790. 10.1016/j.techfore.2019.119790

Cerezo-Narváez, A., Otero-Mateo, M., Rodríguez-Pecci, F., and Pastor-Fernández, A. (2018). Transformación digital de requisitos en la industria 4.0: caso de plataformas navales. *DYNA-Ingeniería e Industria*, *93*(4), 448-456. 10.6036/8636

Chari, M. D. R., Devaraj, S., and David, P. (2007). International diversification and firm performance: Role of information technology investments. *Journal of World Business, 42*(2), 184-197. 10.1016/j.jwb.2007.02.004

Chari, M. D. R., Devaraj, S., and David, P. (2008). Research note - The impact of information technology investments and diversification strategies on firm performance. *Management Science, 54*(1), 224-234. 10.1287/mnsc.1070.0743

Chelliah, S., Sulaiman, M., and Yusoff, Y. M. (2010). Internationalization and performance: Small and medium enterprises (SMEs) in Malaysia. *International Journal of Business and Management, 5*(6), 27. 10.5539/ijbm.v5n6p27

Dalenogare, L. S., Benitez, G. B., Ayala, N. F., and Frank, A. G. (2018). The expected contribution of Industry 4.0 technologies for industrial performance. *International Journal of Production Economics, 204*, 383-394. 10.1016/j.ijpe.2018.08.019

Dess, G. G. and Robinson, R. B. (1984). Measuring organizational performance in the absence of objective measures: the case of the privately–held firm and conglomerate business unit. *Strategic Management Journal, 5*(3), 265-273. 10.1002/smj.4250050306

Devaraj, S. and Kohli, R. (2003). Performance impacts of information technology: is actual usage the missing link? *Management Science, 49*(3), 273-289. 10.1287/mnsc.49.3.273.12736

Dewett, T. and Jones, G. R. (2001). The role of information technology in the organization: a review, model, and assessment. *Journal of Management, 27*(3), 313-346. 10.1177/014920630102700306

Esteben, M. (2007). Can information technology enable profitable diversification? An empirical examination. *Journal of Engineering and Technology Management - JET-M, 24*(3), 167-185. 10.1016/j.jengtecman.2007.05.006

Fatorachian, H. and Kazemi, H. (2020). Impact of Industry 4.0 on supply chain performance. *Production Planning and Control*, 1-19. 10.1080/09537287.2020.1712487

Frank, A. G., Dalenogare, L. S., and Ayala, N. F. (2019). Industry 4.0 technologies: Implementation patterns in manufacturing companies. *International Journal of Production Economics, 210*, 15-26. 10.1016/j.ijpe.2019.01.004

Garbi, G. P. and Loureiro, G. (2013). Business-product-service Portfolio Management. In Bil, C., Mo, J., and Stjepandić, J. (Eds) *20th ISPE International Conference on Concurrent Engineering* (pp. 137-146), Lansdale, PA: IOS Press. 10.3233/978-1-61499-302-5-137

González, R. M., Alfaro-Azofeifa, C., and Alfaro-Chamberlain, J. (2005). *TICs en las PYMES de Centroamérica: impacto de la adopción de las tecnologías de la información y la comunicación en el desempeño de las empresas*. Cartago, Costa Rica: Idrc.

Guzmán, G. M., Serna, M. d. C. M., de Lema, D. G. P., Enríquez, L. A., and Adame, M. G. (2010). La influencia de las TICs en el rendimiento de la PyME de Aguascalientes. *Investigación y Ciencia, 18*(47), 57-65.

Haseeb, M., Hussain, H. I., Ślusarczyk, B., and Jermsittiparsert, K. (2019). Industry 4.0: A solution towards technology challenges of sustainable business performance. *Social Sciences, 8*(5), 154. 10.3390/socsci8050154

Honarmand A. (2011). *IT Adoption and the Internationalization of SMEs* (Master's thesis, Department of Business Administration, Technology and Social Sciences (ETS), Luleå University of Technology, Sweden).

Horváth, D. and Szabó, R. Z. (2019). Driving forces and barriers of Industry 4.0: Do multinational and small and medium-sized companies have equal opportunities? *Technological Forecasting and Social Change, 146*, 119-132. 10.1016/j.techfore.2019.05.021

Hu, T., Zhang, X., and Teng, W. Y. (2011). Understanding Information Technology Configuration in Business Diversification: A Political View. *Information Technology Journal, 10*(3), 470-477. 10.3923/itj.2011.470.477

Khan, M. W. J., Khalique, M., and Nor, R. M. (2014). Exploring the Measurements of Organizational Performance. *Organizational Research Methods, 16*(1). 10.2139/ssrn.2073275

Kim, Y. (2013). *The Korean wave: Korean media go global*. London: Routledge. 10.4324/9781315859064

Lasi, H., Fettke, P., Kemper, H.-G., Feld, T., and Hoffmann, M. (2014). Industry 4.0. *Business and Information Systems Engineering, 6*(4), 239-242. 10.1007/s12599-014-0334-4

Leandro, A. A. (2009). El proceso de internacionalización de empresas. *TEC empresarial, 3*(3), 18-25. https://dialnet.unirioja.es/servlet/articulo?codigo=3202468

Lee, J., Kao, H.-A., and Yang, S. (2014). Service innovation and smart analytics for industry 4.0 and big data environment. *Procedia CIRP, 16*, 3-8. 10.1016/j.procir.2014.02.001

Liu, Y. and Ravichandran, T. (2008). A comprehensive investigation on the relationship between information technology investments and firm diversification. *Information Technology and Management, 9*(3), 169-180. 10.1007/s10799-008-0042-1

Lu Y., Riddick F., and Ivezic N. (2016) The Paradigm Shift in Smart Manufacturing System Architecture. In: Näas I. et al. (eds) *Advances in Production Management Systems. Initiatives for a Sustainable World. APMS 2016. IFIP Advances in Information and Communication Technology, Volume 488* (pp 767-776). Springer, Cham. 10.1007/978-3-319-51133-7_90

Luftman, J., Lyytinen, K., and Zvi, T. B. (2017). Enhancing the measurement of information technology (IT) business alignment and its influence on company performance. *Journal of Information Technology, 32*(1), 26-46. 10.1057/jit.2015.23

Mai, L.-W. and Ness, M. R. (1999). Canonical correlation analysis of customer satisfaction and future purchase of mail-order speciality food. *British Food Journal, 101*(11), 857-870. 10.1108/00070709910301373

Matthyssens, P. (2019). Reconceptualizing value innovation for Industry 4.0 and the Industrial Internet of Things. *Journal of Business and Industrial Marketing, 34*(6), 1203-1209. 10.1108/JBIM-11-2018-0348

McElheran, K. (2014). Delegation in multi-establishment firms: Evidence from I.T. purchasing. *Journal of Economics and Management Strategy, 23*(2), 225-258. 10.1111/jems.12054

Mon, A., Del Giorgio, H., De María, E., Figuerola, C., and Querel, M. (2018). *Evaluación del desarrollo tecnológico para la definición de Industrias 4.0*. Paper presented at the XX Workshop de Investigadores en Ciencias de la Computación (WICC 2018, Universidad Nacional del Nordeste). http://sedici.unlp.edu.ar/handle/10915/67063

Morales, P. G., España, J. A. A., Zárate, J. E. G., González, C. C. O., and Frías, T. E. R. (2017). La nube al servicio de las pymes en dirección a la industria 4.0. *Pistas Educativas, 39*(126). http://www.itcelaya.edu.mx/ojs/index.php/pistas/article/view/1034/845

Muralidharan, R. (1997). Strategic control for fast-moving markets: updating the strategy and monitoring performance. *Long Range Planning, 30*(1), 64-73. 10.1016/S0024-6301(96)00097-0

Patalas-Maliszewska, J. (2012). Assessing the Impact of Erp Implementation in the small Enterprises. *Foundations of Management, 4*(2), 51-62. 10.2478/fman-2013-0010

Piera, P. J. F. and Carrasco, C. A. (2010). El análisis factorial como técnica de investigación en psicología. *Papeles del psicólogo, 31*(1), 18-33. http://www.papelesdelpsicologo.es/pdf/1793.pdf

Porter, M. E. (2015). *Ventaja competitiva: creación y sostenimiento de un desempeño superior*. México, D.F.: Grupo Editorial Patria.

Ravichandran, K. S. and Lorenz, U. (2007). Engulfment of apoptotic cells: signals for a good meal. *Nature Reviews Immunology, 7*(12), 964. 10.1038/nri2214

Ravichandran, T., Liu, Y., Han, S., and Hasan, I. (2009). Diversification and firm performance: Exploring the moderating effects of information technology spending. *Journal of Management Information Systems, 25*(4), 205-240. 10.2753/mis0742-1222250407

Ray, G., Xue, L., and Barney, J. B. (2013). Impact of information technology capital on firm scope and performance: The role of asset characteristics. *Academy of Management Journal, 56*(4), 1125-1147. 10.5465/amj.2010.0874

Sandulli, F. D., Fernández-Menéndez, J., Rodríguez-Duarte, A., and López-Sánchez, J. I. (2012). The productivity payoff of information technology in multimarket SMEs. *Small Business Economics, 39*(1), 99-117. 10.1007/s11187-010-9297-0

Schuh, G., Potente, T., Wesch-Potente, C., Weber, A. R., and Prote, J.-P. (2014). Collaboration Mechanisms to increase Productivity in the Context of Industrie 4.0. *Procedia CIRP, 19*, 51-56. 10.1016/j.procir.2014.05.016

Shin, N. (2006). The impact of information technology on the financial performance of diversified firms. *Decision Support Systems, 41*(4), 698-707. 10.1016/j.dss.2004.10.003

Spinelli, R., Dyerson, R., and Harindranath, G. (2013). IT readiness in small firms. *Journal of Small Business and Enterprise Development, 20*(4), 807-823. 10.1108/JSBED-01-2012-0012

Stentoft, J., Jensen, K. W., Philipsen, K., and Haug, A. (2019). *Drivers and Barriers for Industry 4.0 Readiness and Practice: A SME Perspective with Empirical Evidence*. Paper presented at the Proceedings of the 52nd Hawaii International Conference on System Sciences. 10.24251/HICSS.2019.619

Sullivan, D. (1994). Measuring the degree of internationalization of a firm. *Journal of International Business Studies*, *25*, 325-342. 10.1057/palgrave.jibs.8490203

Tallman, S. and Li, J. (1996). Effects of international diversity and product diversity on the performance of multinational firms. *Academy of Management Journal, 39*(1), 179-196. 10.2307/256635

Vilaseca, J. and Torrent, J. (2003). *Las TIC y las Transformaciones de la Empresa Catalana*. Barcelona: Universidad Oberta de Catalunya.

Wilkesmann, M. and Wilkesmann, U. (2018). Industry 4.0–organizing routines or innovations? *VINE Journal of Information and Knowledge Management Systems, 48*(2), 238-254. 10.1108/VJIKMS-04-2017-0019

Suscripción a la Revista *I&I*

Instrucciones de Suscripción:

1. Consignar **$54 000 COP**, valor correspondiente a **1 año** y **tres ejemplares** de la revista en la cuenta **No. 01272007- 4** (No. Concepto **20181015**) a nombre del Fondo Especial de la Facultad de Ingeniería — Suscripción Revista, en cualquier oficina del **Banco Popular** .
2. Enviar una copia del recibo de consignación junto con los siguientes datos:

   Suscripción a nombre de:

   Email:

   Dirección de envío (País, ciudad,...):

   Teléfono celular:

   Fecha: dd/mm/yyy

"Tecnología e innovación con tradición y excelencia"

ISSN: 0120-5609 (print)
ISSN: 2248-8723 (online)
E-mail: revii_bog@unal.edu.co
Página web: http://revistas.unal.edu.co/index.php/ingeinv

UNIVERSIDAD NACIONAL DE COLOMBIA

Suscripción a la Revista *I&I*

Instrucciones de Suscripción:

1. Consignar **$54 000 COP**, valor correspondiente a **1 año** y **tres ejemplares** de la revista en la cuenta **No. 01272007- 4** (No. Concepto **20181015**) a nombre del Fondo Especial de la Facultad de Ingeniería — Suscripción Revista, en cualquier oficina del **Banco Popular** .
2. Enviar una copia del recibo de consignación junto con los siguientes datos:

   Suscripción a nombre de:

   Email:

   Dirección de envío (País, ciudad,...):

   Teléfono celular:

   Fecha: dd/mm/yyy

INGENIERÍA E INVESTIGACIÓN

"Tecnología e innovación con tradición y excelencia"

ISSN: 0120-5609 (print)
ISSN: 2248-8723 (online)
E-mail: revii_bog@unal.edu.co
Página web: http://revistas.unal.edu.co/index.php/ingeinv

UNIVERSIDAD NACIONAL DE COLOMBIA

Instructions for Authors

Editorial Committee reserves the copyright to printing any material and its total or partial reproduction, as well as the right to accept submitted material or reject it. It also reserves the right to make any editorial modification which it thinks fit. In such event, the author of the submitted material in question will receive the evaluators' recommendations for changes to be made in writing. If an author accepts them, the revised (or rewritten) article must be submitted with the suggested changes having been made by the date fixed by the journal to guarantee its publication in the programmed issue.

The process to be followed for publishing an article in the journal

The article must be uploaded into the journal's OJS website, see the guidelines for article submission in the Authors guide section in our website http://www.revistas.unal.edu.co/index.php/ingeinv/article/view/59291/56815. Any manuscript must be sent using journal's template (6 pages length max.) and must be accompanied by the license agreement, addressed to the journal's editor, Prof. Andrés Pavas, stating that all authors involved in the work in question agree to it being submitted for consideration in the *Ingeniería e Investigación journal*.

Article and License templates are available on:
http://www.revistas.unal.edu.co/index.php/ingeinv/index

Once an article has been received by the journal, the corresponding author will be notified by e-mail and the peer-review process will be begun. Following this evaluation, authors will then be informed whether their article has been accepted or not. If accepted, authors must deal with the respective corrections recommended by the evaluators and the Editorial Committee's final decision. If it is to be published.

Content

All articles being considered by the committee for possible publication in the *Ingeniería e Investigación journal* must consist of the following parts:

- Title, abstract and keywords must be written in Spanish and English. The title must clearly explain the contents of the article in question, written in normal title form and be preferably brief. The abstract should contain around 200 words in Spanish and English, as well as including the methods and materials used, results obtained and conclusions drawn.
- An Introduction must be given. It must describe article's general purpouse, including its main objective, referring to any previous work and the scope of the current article.
- Conclusions must be drawn. This section must provide the implication of the results found and their relationship to the proposed objective.
- Bibliographical references must be given (an explanation and example of how to set them out is given later on).
- Acknowledgements (Optional). These should be brief and mention any essential support received for carrying out the work being reported.
- Appendix (Optional).

Scientific and technological research articles must also include:

- Experimental development. This must be written giving sufficient details for the subject to be fully understood by readers, including descriptions of any procedures involved.
- Results. These must give a clear explanation and interpretation of the findings. If it is necessary, a brief, focused discussion about how given results can be interpreted.

It is required that the bibliographical references for all articles are included at the end of the article, given in alphabetical order of first authors' surnames and mentioned in the text and, since May 2014, it is asked that the authors use the American Psychological Association (APA) style for citation and references:

- Articles published in journals:

Author, A. A., Author, B. B., & Author, C. C. (year). Article title. Journal Title, volume number(issue number), page numbers.

Del Sasso, L. A., Bey, L. G. & Renzel, D. (1958). Low-scale flight ballistic measurements for guided missiles. Journal of the Aeronautical Sciences, 15(10), 605-608

Author, A. A., & Author, B. B. (year). Article title. Journal Title, volume number(issue number), page numbers. Retrieved from http://www.xxxxxxxxxxxxxxxx

Gaona, P. A. (2014). Information visualization: a proposal to improve search and access digital resources in repositories. Ingeniería e Investigación, 34(1), 83-89. Retrieved from http://www.revistas.unal.edu.co/index.php/ingeinv/article/ view/39449

- Books:

Author, A. A. & Author, B. B. (year). Title of work. Location: Publisher.

Turner, M. J., Martin, H. C. & Leible, R. C. (1964). Further development and applications of the stiffness method, Matrix Methods of Structural Analysis. New York: the Macmillan Co.

- Conference papers and symposium contributions:

Uribe, J. (1973, September). The effects of fire on the structure of Avianca building, Paper presented at National Seminar concerning Tall Buildings, Bogotá, Colombian School of Engineering.

- Theses or undergraduate projects:

Patton, F. D. (1906). Multiple modes of shear failure in rock-related materials (PhD thesis, University of Illinois).

Further information can be obtained by:

Contacting the Editorial Team (Email: revii_bog@unal.edu.co) or Prof. Andrés Pavas (Editor-in-Chief.
Email: fapavasm@unal.edu.co)

The *Ingeniería e Investigación* journal's office is located at: Ciudad Universitaria, Facultad de Ingeniería, Edificio CADE. Telefax: (57-1) 3165000 Ext. 13674. Bogotá - Colombia.