Ingeniería e Investigación
Journal

Abbreviated Journal Title: **Ing. Investig.**

Frequency
Continuous periodicity (three issues per year)

Cover Layout
Gabriela Rojas Castro

Proofreader
José Daniel Gutiérrez-Mendoza

Layout Artist
David Mauricio Valero

For additional information, please contact:
revii_bog@unal.edu.co
Bogotá - Colombia
2024

Table of Contents

Mechanical Engineering, Mechatronics, and Materials Science

Systems and Computer Engineering

Education in Engineering

Research Article/Agricultural Engineering http://doi.org/10.15446/ing.investig.104844

Stress Evaluation Using Finite Elements in a Manual Agricultural Tool

Evaluación de esfuerzos mediante elementos finitos en una herramienta agrícola de acción manual

Laura Delgado-Bejarano[1], Hugo González-Sánchez[2], and Germán García-Monsalve[3]

ABSTRACT

This study addresses the imperative need for efficient hand-held agricultural tools, particularly in challenging contexts like hillside agriculture, by focusing on the redesign and evaluation of a manual tillage tool. The objective is to comprehensively assess the stress and fatigue life of a redesigned tool, considering different manufacturing materials such as steels (AISI/SAE 4140, 4130, 1060), A356 aluminum, and nodular cast irons. Employing finite element method simulations and the Von Mises equation, this research confirms an optimal performance within elastic limits for all materials, mitigating the risks of plastic deformation or breakage during normal operation, with Von Mises stresses ranging from 8.39 to 16.30 MPa. All the tools yielded optimal results, meeting the critical requirements for soil penetration resistance, reporting no fatigue failures, and exhibiting useful life values over 1.75 x 1013 years. In terms of ergonomics, A356 aluminum stands out, as it is less heavy and implies a lower effort by the operator, promoting efficient tillage without compromising comfort. This research provides nuanced insights for the design of agricultural tools, emphasizing the harmonious balance between efficiency, longevity, and operator comfort in sustainable practices.

Keywords: handheld tools, elastic limits, fatigue, sustainable agriculture, ergonomics

RESUMEN

Este estudio aborda la imperiosa necesidad de herramientas agrícolas manuales que sean eficientes, especialmente en ámbitos desafiantes como la agricultura en laderas, centrándose en el rediseño y evaluación de una herramienta de labranza manual. El objetivo es evaluar exhaustivamente la tensión y la vida útil ante la fatiga de una herramienta rediseñada considerando diferentes materiales de fabricación como aceros (AISI/SAE 4140, 4130, 1060), aluminio A356 y hierros fundidos nodulares. Empleando simulaciones del método de elementos finitos y la ecuación de Von Mises, esta investigación confirma un rendimiento óptimo dentro de los límites elásticos para todos los materiales, mitigando los riesgos de deformación plástica o rotura durante la operación normal, con tensiones de Von Mises que van de 8.39 a 16.30 MPa. Todas las herramientas presentaron resultados óptimos, cumpliendo con los requisitos críticos de resistencia a la penetración del suelo, sin reportar fallas por fatiga y demostrando valores de vida útil superiores a 1.75 x 1013 años. En términos de ergonomía, el aluminio A356 se destaca por ser menos pesado e implicar un menor esfuerzo por parte del operador, promoviendo una labranza eficiente sin comprometer la comodidad. Esta investigación proporciona percepciones matizadas para el diseño de herramientas agrícolas, enfatizando el equilibrio armonioso entre eficiencia, longevidad y comodidad del operador en prácticas sostenibles.

Palabras clave: herramientas manuales, límites elásticos, fatiga, agricultura sostenible, ergonomía

Received: September 18th, 2022
Accepted: January 16th, 2024

Introduction

Currently, the development of efficient and functional handheld agricultural tools is of great significance for advancing the engineering of the agricultural sector. Research works that propose use-based modifications to current designs use are a priority for manufacturing innovative agricultural equipment and tools aimed at the modernization of relatively small rural areas (Toni *et al*., 2017). In this regard, hillside agriculture constitutes a particular case, where high slopes hinder tillage via motorized machines. In this type of land, it is necessary to use highly efficient, manual agricultural tools that generate a low impact on the soil and feature ergonomic designs adjusted to the anthropometric characteristics of the farmer (Sims *et al*., 1998).

[1] Agricultural engineer., Universidad Nacional de Colombia, Colombia. Affiliation: Researcher, Universidad Nacional de Colombia, Colombia. Email: ladelgadobe@unal.edu.co

[2] Agricultural engineer., Universidad Nacional de Colombia, Colombia. MSc in Materials Engineering, Universidad Nacional de Colombia, Colombia, PhD in Materials Engineering, Universidad Nacional de Colombia, Colombia, Affiliation: professor, Universidad Nacional de Colombia, Colombia. Email: hagonzal@unal.edu.co

[3] Mechanical engineer., Universidad Nacional de Colombia, Colombia. MSc in Materials Engineering, Universidad Nacional de Colombia, Colombia, PhD in Materials Engineering, Universidad Nacional de Colombia, Colombia, Affiliation: professor, Universidad Nacional de Colombia, Colombia. Email: glgarcia@unal.edu.co

Hoes are among the most commonly employed manual agricultural tools. A hoe is an instrument with a generally curved metal blade that is used to graze, break soils with high hardness, and cut thin roots, among others (González, 2018; Sanahuja, 1971). In addition, it has been reported that mattocks and hoes have been widely used for agricultural tillage since the Neolithic period and by different cultures (Mazoyer and Roudart, 2006). According to Byzantine manuscripts, the hoe was the most widely used tool in the Hesiod era, and it appeared years later in the British Isles during the late Mesolithic. This tool is likely responsible for the emergence of agriculture as we know it. Sanahuja (1971) refers to findings in Iberian deposits and Roman villas in Catalonia, among which some agricultural tools such as mattocks and hoes stand out. When comparing their design against those currently marketed, this author noted little variation in their geometry and manufacture (forged metals). In addition to agricultural use, there are documents that mention the use of hoes as weapons during battles between the Irish and the Spanish (O'Donell and De Estrada, 2014).

Today, there is a variety of hoes, which differ in size, geometry, materials, and use type. Moreover, handheld hoes are regarded as tools of great importance for agriculture and are likely the most widely used in tillage and soil clearing operations worldwide, which is why many farmers depend on them (Sims *et al*., 1998). This may be the main reason why they have evolved to the point of being commercially differentiated by the aforementioned characteristics (Ashburner and Kienzle, 2013). In this context, abrasive wear is an important factor that affects functionality, and, according to Agudelo (2013), it is a significant issue in soil removal and agriculture, increasing the costs associated with energy consumption, repair, part replacement, and the downtime generated by these activities. Agudelo (2013) estimates that the costs generated by friction and wear on agricultural tools, both mechanized and manual, can represent between 0.3 and 3.0% of a country's GDP, and that this situation is more critical in developing countries. Similarly, Yazici (2011) mentions that performing preventive maintenance can save around 337 million dollars a year.

The foregoing justifies undertaking innovation processes in design, materials, use techniques, and ergonomics, among others, in order to achieve greater efficiency in agricultural tillage with hand tools.

This work presents an evaluation of the stress generated by impact load on a manual tillage tool, considering different manufacturing materials and using the finite element method (FEM) in computational simulations. In particular, the employed hoe, modeled for furrowing work, features an improved geometric design, and it was manufactured with materials resistant to wear and impact, *i.e.*, AISI/SAE 4140, 4130, and 1060 steel; A356 aluminum castings; and nodular cast irons with different surface treatments (annealed and normalized). It should be noted that some of the materials mentioned above have already been used for the manufacture of agricultural tools and have been referenced by various authors in different patents. As per a patent on the construction of vertical plows (Balvanz and Eldora, 2019), AISI/SAE 4140 steel is used to make moldboards in the plows and is implemented in the construction of backhoes with high resistance to wear (Gegel, 1998). Yu and Bhole (1990) report the high resistance of AISI/SAE 4130 steel, hence its great usefulness in the manufacture of plow tools. Teko *et al*. (2018) have reported this material in regard to the construction of blades for sugar cane harvesters. As for AISI/SAE 1060 steel, De Almeida Luz (2019) reports their use in furrowers. A356 aluminum has been reported in the manufacture of machine elements such as pistons, engine blocks, and car wheels (Ahmed *et al*., 2016; Charco, 2017; Fernanda *et al*., 2018). Finally, Bednář *et al*., (2013) have used nodular cast irons to manufacture tools for subsoilers, given that this material has a high resistance to abrasive and erosive wear and is regarded as a ductile iron (Enríquez Berciano and Tremps Guerra, 2012).

For stress evaluation via the FEM, a maximum power of 300 W (Botta, 2003) was considered as a critical value, which is reached upon the impact between a fertile soil and a tool, considering the dynamic movement of the latter as resembling a particle that hits the former for a short period of time. The simulations were carried out using the ANSYS software, taking the Von Mises equation in three dimensions as a reference, to determine the elastic stress concentration sites of each of the materials used for manufacturing the tools. *Stress* is denoted as the uniaxial tensile stress that generates a combination of maximum distortion energy, considered as the actual combination of applied stresses, expressed in a two-dimensional or three-dimensional manner (Cruz-Avilés *et al*., 2018; Norton, 2011).

Most manual farming tools available in the market are designed considering the force they can apply to the soil, their producibility, and their efficiency, but the comfort of the operator is rarely taken into account. It has been reported that most of the agricultural workforce suffers from, pain in the neck, shoulders, arms, and trunk (Benos *et al*., 2020; Yusuf and Yusuf, 2006). Thus, designing a tool that allows for the dissipation of forces, enables proper cultivation, and reduces the effort required by the operator, would contribute to greater efficiency in farming tasks, without compromising the workers' quality of life.

Therefore, the objective of this research was to evaluate the stresses and lifecycle of a manual agricultural tool through the FEM, analyzing tools of different materials when subjected to impact loads.

Materials and methods

Manual tool modeling in computer software

Using the SolidWorks software, the redesign of a conventional hoe (quadrangular) was modeled. Through image processing, a final 3D model with the same dimensions and geometry as the actual tool used in the soil preparation process was

obtained. To create the design, photographs of the tool were taken from the side, top, and front views. Measurements of length, width, and height were taken, and, based on them, the photographs were scaled in the software to generate a model (Figures 1 and 2).

Figure 1. Manual tool taken as a reference for modeling in SolidWorks
Source: Authors

Figure 2. SolidWorks model of the manual tool
Source: Authors

Mechanical properties and materials used for the simulation

To perform the FEM stress simulations, the following groups of materials were considered: nodular cast irons (annealed and normalized), low-alloy steels with medium-carbon content (AISI/SAE 4140 and 4130), carbon steel (AISI/SAE 1060), and cast aluminum (A356).

The FEM analysis considered an impact between the tip of the manual tool and the soil, simulating the energy that a farmer would exert during tillage. Through calculations using Newton's third law (action and reaction), the simulation results were found to be equivalent, *i.e.*, the force exerted by the tool against the soil was of the same magnitude as that exerted by the soil on the tool, but in the opposite direction. Regarding the magnitude of the impact force, the critical value reported by Botta (2003) was considered. This author concluded that, during a normal working day, a person exerts a force of approximately 60 N, at a speed of 1.1 m/s (equivalent to 66 W), albeit stating that, for some instants, an operator can impart a power close to 300 W on the agricultural tool. This value was taken as a reference in our simulations, which were performed using the ANSYS 2019 R2 V19.4 software. The mechanical module was employed, which required entering the characteristics and properties of each material. Molten aluminum (A356) was modeled based on information on physical properties obtained from Mott (2009) and Matweb (2016). The data for the nodular cast iron were obtained from Agudelo (2013), Angus (1976), and Herring (2018). In the case of steels, information from Matweb (2016) and Budynas and Nisbett (2015) was employed. The mass of the materials was calculated based on the density values obtained in the aforementioned sources and the volume provided by SolidWorks, *i.e.*, 9.2977×10^{-5} m^3. Based on the constant velocity, a static structural analysis was performed.

Table 1 presents the mechanical properties and the mass calculations considered for each hand tool material. This information was used in Equation (1) to model the dynamics of the impact during tillage, considering the tool as a particle.

$$\int \underline{F} dt = m \triangle \vec{v} \quad (1)$$

Table 1. Mechanical properties of the materials used for the impact force calculations and the FEM stress simulation

Material	Mass (kg)	Elastic modulus (GPa)	Poisson's ratio	Density (kg/m³)	Elastic limit (MPa)	Tensile strength (MPa)
AISI/SAE 4140	0.73	205.00	0.29	7850.00	675.00	1020.00
AISI/SAE 4130	0.73	205.00	0.29	7850.00	460.00	731.00
AISI/SAE 1060	0.73	205.00	0.29	7850.00	427.00	779.00
Annealed nodular cast iron	0.66	168.33	0.29	7120.00	275,79	413.69
Normalized nodular cast iron	0.66	168.33	0.29	7120.00	482.63	689.48
A356 aluminum permanent mold cast	0.25	72.40	0.33	2670.00	179.00	255.00

Source: Agudelo (2013), Angus (1976), Budynas and Nisbett (2015), Herring (2018), Matweb (2016), Mott (2009)

Where:

- F: impact force due to weight (N)
- dt: differential time for impact (s)
- $m\Delta\vec{v}$ mass variation due to the speed of the tool during the impact (kg-m/s)

By solving the integral in Equation (1), the formula for the linear motion of an impacting particle was obtained, as shown in Equation (2).

$$F\left(\Delta t\right) = mv_f - mv_i \tag{2}$$

Where:

- m : mass of the tool (kg)
- v_i : initial speed of the tool (m/s)
- v_f : final velocity just before impact (m/s)

To obtain the impact force, the initial velocity (v_i) was set as 0, considering that the tool starts from a resting position. Following Botta (2003), a final speed of 1,1 m/s was assumed, as well as an impact time 0.001 s. For similar situations, the analysis of Marcon (2018) was taken into account, where the maximum impact time used was 10 ms. However, for the impact to be more critical, a minimum value was taken, *i.e.*, approximately one thousandth of a second.

In accordance with the above, the impact force for each hoe was calculated with the following values, as represented in Equation (3).

$$F = \frac{mv_f}{\left(\Delta t\right)} = \frac{\left(mass\,of\,hoe\right)\left(1.1\frac{m}{s}\right)}{0.001s} \tag{3}$$

During the impact, it was considered that the operator applied a force of 272.73 N. This value was added to that obtained in Equation (3) and tabulated as the total force for each material (Table 2).

Table 2. Impact force of a hoe on the soil for each tool material

Material	Force (N)	Fy (N)	- Fz (N)
AISI/SAE 4140	1075.58	1036.47	287.44
AISI/SAE 4130	1075.58	1036.47	287.44
AISI/SAE 1060	1075.58	1036.47	287.44
Annealed nodular cast iron	1000.92	964.52	267.49
Normalized nodular cast iron	1000.92	964.52	267.49
Aluminum A356	545.80	525.95	145.86

Source: Authors

Simulation and finite element model

In the FEM simulations, a mesh size of 2 mm was selected for optimal and precise results. 207 799 nodes and 132 830 tetrahedral elements were obtained, as this type of mesh allows for a better adjustment of the tool's geometry, adapting to changes in shape and to smaller areas. In general, the mean quality was 0.7617, and reductions in this regard were caused by sudden changes in geometry. However, considering that the tool has a complex shape, the reported values are good, as they are closer to 1, which denotes the highest quality (Seeni *et al.*, 2020). Additionally, the aspect ratio was 2.3041 – a good-quality mesh has an aspect ratio between 1 and 5. The Jacobian ratio obtained was 0.9984, with the ideal value being 1 (SolidWorks, 2022). In addition, greater precision can be obtained with a larger number of nodes (Dutt, 2015; Molino *et al.*, 2003). The static module of the ANSYS finite element software was used for modeling and calculating the maximum Von Mises stress after assigning a type of material to the studied tools. Then, the boundary conditions were selected, *i.e.*, the type of stress, corresponding to the force exerted by the farmer when using the furrowing hoe; and the impact force of each material, concentrated on the tip of the tool, using approximately one third of the total area (2.9 x 10^{-3} m^2). This procedure is illustrated in Figure 3. Here, the area in red receives the impact. The direction follows the rib (tangential to the tool and perpendicular to the soil), which is considered the most critical factor. The tool stub is centered on the z-axis, implying that the forces are distributed in the y- and z-axes (Table 2). Regarding the distribution, a tool angle of 15.5° was taken into account. The magnitude of the force applied by each material is presented in Table 2. Each material was simulated with these parameters.

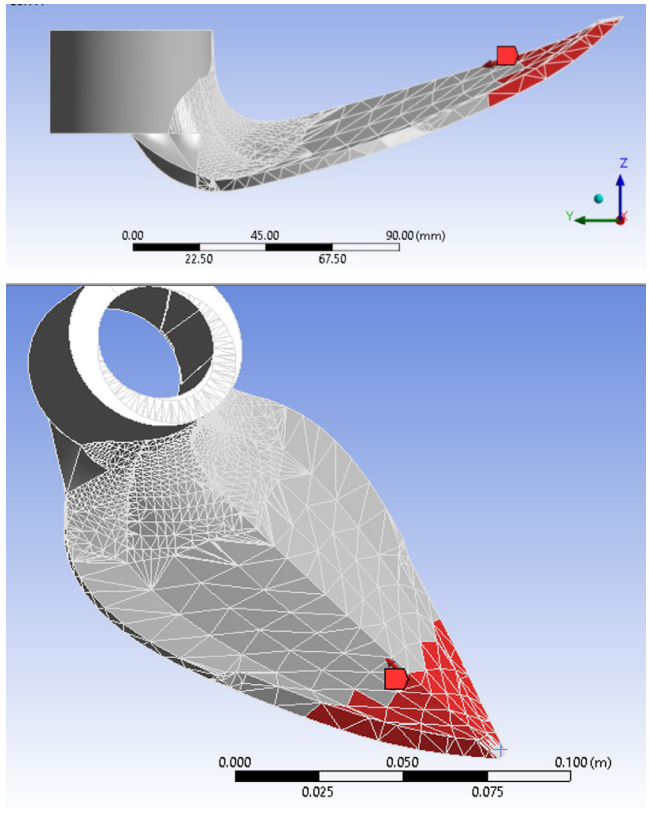


Figure 3. Zone of greatest influence of the impact force on the soil
Source: Authors

It is important to note that, internally, the software calculates the Von Mises stresses that can produce failures due to deformation or breakage. For a deformation of the material to occur, it is necessary to equal or exceed the elastic limit stress in tension for each material. In the event of a break, the maximum stress at break must be equaled or exceeded.

The Von Mises stress can be expressed as in Equation (4) for a three-dimensional system using the XYZ components (Budynas and Nisbett, 2015).

$$\sigma' = \frac{1}{\sqrt{2}}\left[\left(\sigma_x - \sigma_y\right)^2 + \left(\sigma_y - \sigma_z\right)^2 + \left(\sigma_z - \sigma_x\right)^2 + 6\left(\tau_{xy}^2 + \tau_{yz}^2 + \tau_{zx}^2\right)\right]^{\frac{1}{2}} \quad (4)$$

Where:

- σ' : Von Mises stress
- σ_x : normal stress in the x-direction
- σ_y : normal stress in the y-direction
- σ_z : normal stress in the z-direction
- τ_{xy} : shear stress in the xy plane
- τ_{yz} : shear stress in the yz plane
- τ_{zx} : shear stress in the zx plane

In order to compare the values obtained from the simulation against those of the manual calculations, the simplified Von Mises equation was used:

$$\sigma' s = \sqrt{\frac{\left(\sigma_1 - \sigma_2\right)^2 + \left(\sigma_2 - \sigma_3\right)^2 + \left(\sigma_3 - \sigma_1\right)^2}{2}} \quad (5)$$

Where:

- $\sigma' s$: simplified Von Mises stress
- $\sigma_1 = \sigma_x$: normal stress in the x-direction
- $\sigma_2 = \sigma_y$: normal stress in the y-direction
- $\sigma_3 = \sigma_z$: normal stress in the z-direction

The percent error is presented in Equation (6).

$$\%E = \frac{FEM\ Value - Manual\ Value}{Manual\ Value} * 100 \quad (6)$$

Lifecycle simulation

Using the ANSYS software, a simulation of the tool's lifecycle was carried out, based on the fatigue parameters for each of the materials, in addition to using the Von Mises stresses generated in the previous simulations and the geometry of the tool. We predicted the number of cycles that the tool could withstand under the conditions provided by the stress analysis, considering a reversible stress and Goodman relation. For an eight-hour workday and a work efficiency of 70%, the number of cycles in a conventional ploughing day was determined, finally obtaining the number of days that the tool can work before failing due to fatigue.

Results and discussion

The Von Mises theory, also known as the *maximum distortion energy theory*, offers a versatile framework for understanding the behavior of diverse materials across various service conditions. These conditions encompass factors such as stress application rates, impact severity, and working temperatures, as well as specific design characteristics like material properties, defects, and geometry. This theory posits that yielding and failure occur when the total deformation energy per unit volume reaches or surpasses the energy corresponding to the material's yield strength (Budynas and Nisbett, 2015; Vanegas Useche, 2018). According to Pérez-González *et al*. (2011) the Von Mises criterion is a good option for materials with a ductile tendency, but it is not suitable for materials with a brittle tendency (in such materials, they recommend using another one, such as the Christensen criterion). The materials used in this research exhibit a ductile tendency, as it is common to find rock fragments in agricultural soils, where a sudden impact can more frequently lead to cracks or catastrophic failures in the material. In the case of the materials under study, the Von Mises equivalent stress values, calculated based on normal stresses, consistently remained below the yield strength of each material. This observation indicates that the actual stresses experienced during impact did not surpass the material's elastic limit, highlighting that the material's performance primarily resides within its elastic zone.

The computational results reveal that the geometric design of the tool's tip facilitates stress distribution and concentration in localized areas, effectively dissipating the deformation energy generated during operation. While the computational graphs demonstrate stress concentration in small regions relative to the tool's size, none of the materials exceeded their yield limits or elastic strength. This reaffirms that the materials predominantly operate within their elastic zones.

Table 3. Von Mises elastic stresses for each tool material

Material	Von Mises Stress (MPa)	Elastic limit (MPa)
AISI/SAE 4140	16.30	675.00
AISI/SAE 4130	16.30	460.00
AISI/SAE 1060	16.30	427.00
Annealed nodular cast iron	15.17	275.79
Normalized nodular cast iron	15.17	482.63
Aluminum A356 permanent mold cast	8.39	179.01

Source: Authors

For the steels, stress magnitudes between 69.67 x 10^{-4} and 16.30 MPa were obtained (Figure 4). In general, the concentration of Von Mises stresses in the conic area where the cape is installed was evident for each type of material,

mainly due to the change in the geometry of the tool and the presence of the moment generated by the cape during agricultural work. Particularly, the highest stress magnitudes were obtained in the area of the nerve, as it has a greater contact with the soil and due to the change in section from its pointed geometry, implying areas with high stress concentrations. This is similar to the results of Cazin *et al.*, (2020), who found that the maximum efforts occurred in joints and geometry changes.

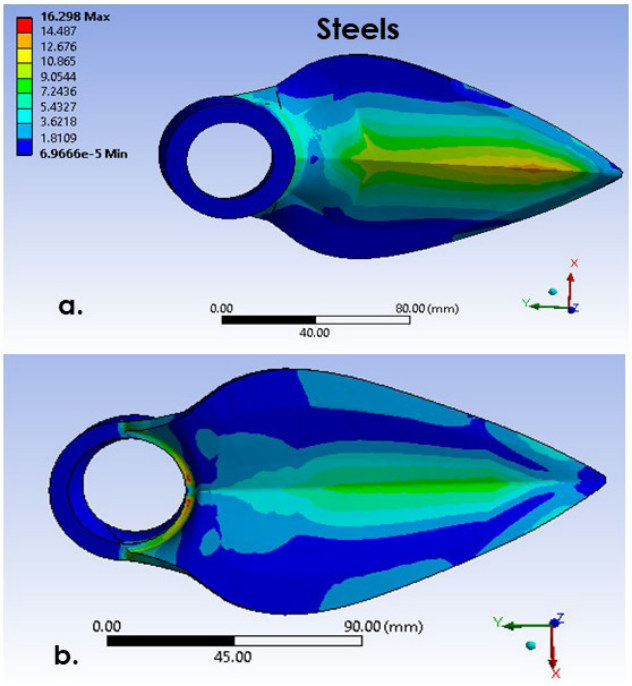


Figure 4. Von Mises stress calculated for steel tools**.** Views: a) front, b) back.
Source: Authors

The nodular steel tools simulated (Figure 5) reported values between 64.85 x 10^{-4} and 15.17 MPa, with a behavior similar to that of steels regarding the stress concentration areas. These values, although slightly lower (by 7%) than those of steels, do not represent great differences in the Von Mises stress calculated in this study. However, given the nature of the materials, they could eventually exhibit a different wear behavior, which was not within the scope of this study.

In the case of the A356 aluminum tool, the calculated Von Mises stress ranged from 40.46 x 10^{-4} to 8.39 MPa (Figure 6), which is significantly lower than the values obtained for steels and iron castings. The main reason for this difference is that the aluminum hoe is significantly lighter than the others, as well as the impact force exerted on the tip of the tool. The maximum values are observed in the nerve and the area of connection with the end of the furrowing hoe, maintaining a behavior similar to that of the other materials, as a result of the geometry of the tool and the stress concentration areas. In addition, the tillage conditions and the impact model are similar for all the studied materials.

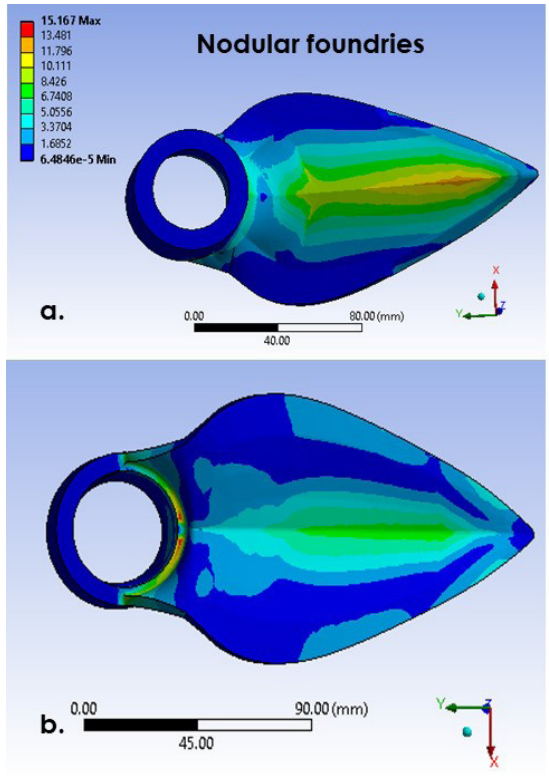


Figure 5. Von Mises stress calculated for the nodular foundry tools. Views: a) front, b) Back.**Source:** Authors

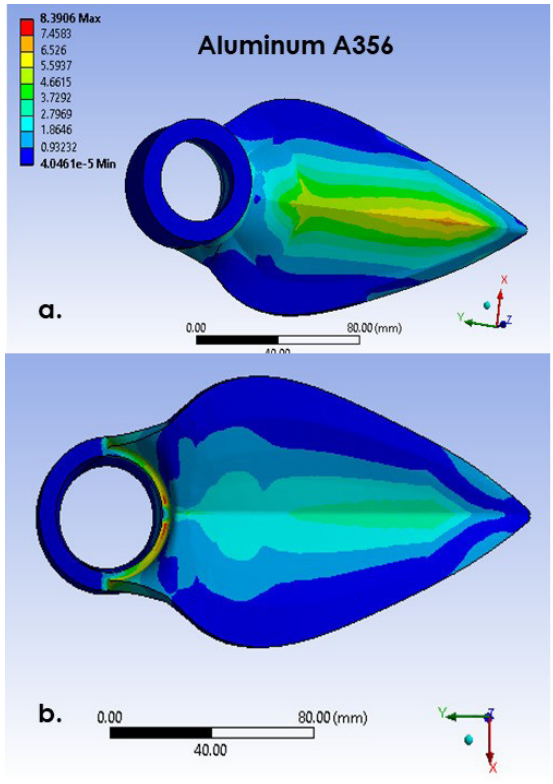


Figure 6. Stress generated in the aluminum tool. Views: a) front, b) back.
Source: Authors

As per Von Mises theory, the elastic limit of the material and the maximum stress generated must be compared. In Table 3, it is evident that none of the furrowing hoe materials exceeds the creep limit, *i.e.*, no tool material will fail under the established working conditions, which are similar to those proposed by Leslie and Aguila (2014) for the FEM analysis of a multiple chisel plough, where the tool did not any exhibit failures caused by exceeding the elastic limit. This leads to the hypothesis that the tool materials' useful life could be long and lack plastic deformations. Consequently, the tools' lifecycle may be influenced by tribological factors such as wear.

Table 4 presents the results obtained in the simplified von Mises stress analysis, in addition to the percent error.

Table 4. Simplified Von Mises elastic stresses for each tool material

Material	Simplified Von Mises stress σ' (MPa)		%E
	Manual	FEM	
AISI/SAE 4140	4.93	5.27	6.90
AISI/SAE 4130	4.93	5.27	6.90
AISI/SAE 1060	4.93	5.27	6.90
Annealed nodular cast iron	4.58	4.90	6.99
Normalized nodular cast iron	4.58	4.90	6.99
Aluminum A356 permanent mold cast	2.50	2.69	7.60

Source: Authors

Note that the percentage errors are below 8%. This indicates a satisfactory outcome of the simulation, and this error can be attributed to the geometry and distribution of the areas during the simplified calculations.

According to the previous analysis and given that the tool is unlikely to fail by exceeding its elastic limit, a fatigue analysis was carried out to determine the number of lifecycles to which the different tool materials could be subjected before exhibiting damage to their microstructure. Figures 7, 8, and 9 present the lifecycles for the steel, iron, and aluminum foundries, respectively. These analyses were performed while assuming constant loads, without any variability during different agricultural tasks. However, it should be indicated that lifecycles can be affected because of the interaction between the tool and a heterogeneous and anisotropic material, such as agricultural soils. Furthermore, it is known that fatigue is not a critical factor of failure in this type of tool, as discussed below.

As shown in Figure 7, the number of lifecycles for the tool under load conditions is very high. In the case of AISI/SAE 4130 steel, there is a stress concentration zone similar to that obtained in the Von Mises simulation, with values of up to 7.65 x 10^{29} lifecycles before exhibiting failures, particularly in the area adjacent to the nerve and on the cape, representing the notching areas subjected to higher concentrations of stress. Additionally, it is important to note that, for the AISI/SAE 1060 and 4140 steels, a clear visual reference is not given since they have higher tensile strength and lifespan.

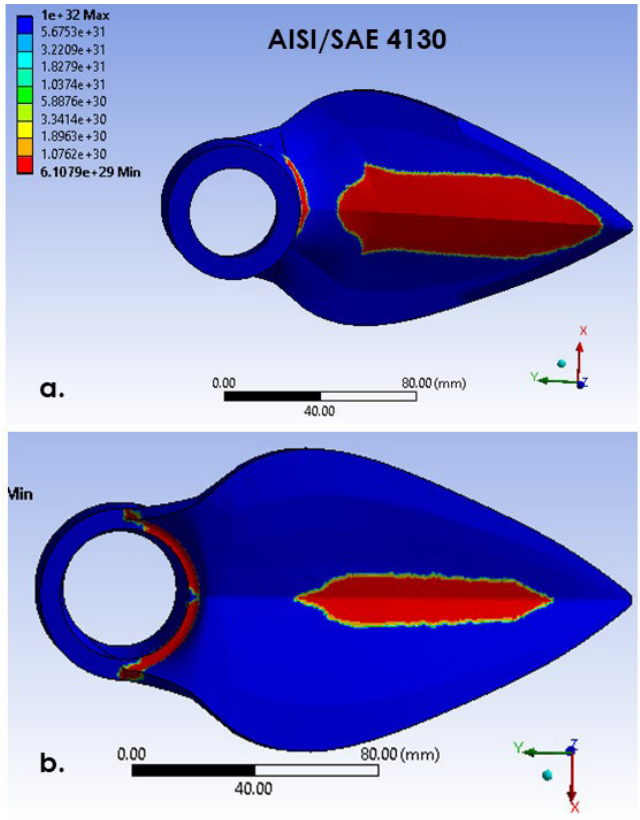


Figure 7. Simulated lifecycles of the 4130 steel tool. Views: a) front, b) back.
Source: Authors

The opposite occurs with iron castings, although the normalized nodular casting has a higher tensile strength than the annealed one. However, the fatigue properties of the latter are better, implying that the tool material subjected to the standardization process exhibits a greater susceptibility to failure due to fatigue. Those differences lead to the hypothesis that the microstructure of iron foundries influences fatigue cycles due to the heat treatment of the tool. However, the number of cycles necessary for fatigue to occur is very high; given the analyzed conditions, it takes about 3.06 x 10^{29} cycles for the tool to fail.

Figure 8 shows a higher stress concentration zone in the standardized nodular casting tool, implying a difference in the number of lifecycles when compared to that shown in Figure 7 for AISI/SAE 4130 steel, whose fatigue resistance is greater. The latter also has a better distributed tool geometry.

In the fatigue simulation, the A356 aluminum tool has the lowest fatigue lifecycles, since it is made of the material with the lowest tensile strength and has the lowest ductility coefficient. However, Figure 9 shows that, except for the stress concentration area, the lifecycles are similar

to those of the other tools (1 x 10^{32} cycles). It is worth noting that A356 aluminum is the material that supports the lowest load.

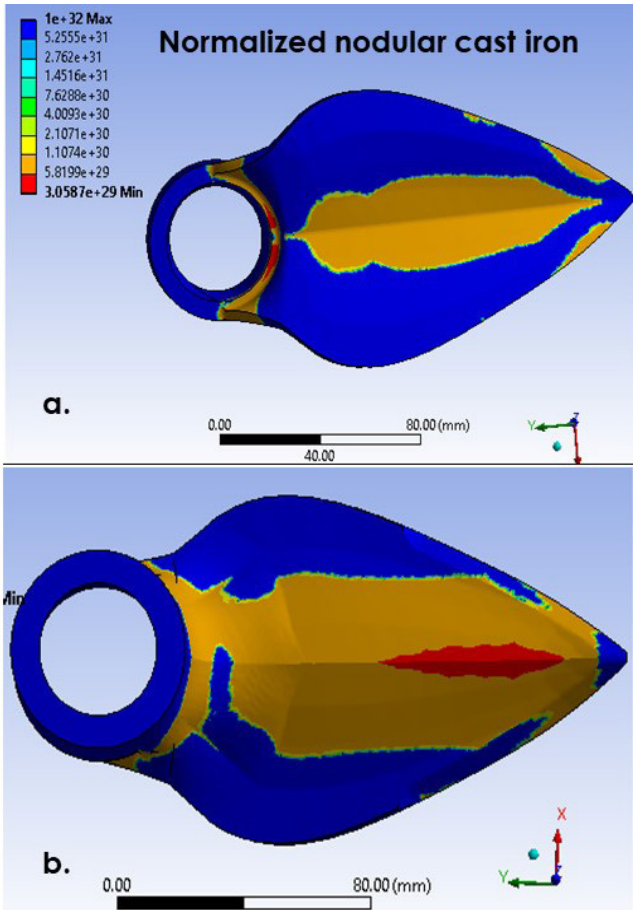


Figure 9. Simulated life cycles for the A356 aluminum tool. Views: Views: a) front, b) back.
Source: Authors

The application of loads below the creep limit on agricultural tools generates changes in the internal microstructure of each material. When this process occurs repeatedly under different conditions, it can damage the tool over time, as it is subjected to fluctuating and variable stresses. This type of failure is called *high cycling failure*, and it occurs when the stresses are very far from the creep limit. In the case of the studied metals, the number of lifecycles needed for failure is very high. This could only be demonstrated after the tools have been operated for much of their useful life and/or by performing specific microstructure and even nanostructure tests (Espejo and Hernández, 2017).

Table 5 presents the parameters calculated regarding the number of days that the A356 aluminum tool could withstand under traditional working conditions. This material was selected because, according to the fatigue results, it is the one with the most critical fatigue lifecycles.

Note that, under normal working conditions, the tool would have a very long lifespan, in the order of thousands of years. Thus, the critical factor for a damage to the tool is determined through other mechanisms, such as abrasive wear given the soil-tool interaction, but not through fatigue.

Table 5. Fatigue parameters

Parameter	Value	Units
Impacts/day	18 144	Cycles/day
A356 aluminum lifecycles	7.60 x 10^{19}	Cycles
Lifespan of A356	8.38 x 10^{14}	Weeks
	2.09 x 10^{14}	Months
	1.75 x 10^{13}	Years

Source: Authors

From a perspective of ergonomics and the principle of lowest energy expenditure by the operator, the furrowing hoe with the best performance is the one designed with aluminum casting; its lower weight implies the lowest impact force, hence the lower energy requirements. The stress required for penetrating the soil should be considered, since the tool must withstand the soil's resistance without failing. Jaramillo (2014) reports that a soil resistance value greater than 35.5 kg/cm^2 (3.48 MPa) constitutes the limit for root development in a great proportion of crops. Still, considering the load necessary to break the soil in very critical situations, none of the tools exceeds its elastic limit, indicating that the agricultural soil does not offer sufficient resistance to cause the tool to fail due to impact load. Therefore, based on the stress and fatigue analysis, all the studied materials have a potentially long useful life, in addition to the fact that all of them exceed the critical agricultural soil stress, allowing for an adequate decompaction. If worker ergonomics are considered, A356 is the recommended furrower material. However, to select the optimal material, it is necessary to perform an abrasive wear analysis, given that, although there is no fatigue failure, the loss of useful life can be caused mainly by the abrasion generated by agricultural soils, especially those with a higher concentration of sand particles, which in turn causes the loss of shape and mass of the tool (Singh *et al*., 2020). Abrasive wear is considered to be one of the most serious issues in agricultural tools, reporting losses of up to 40 million dollars in countries such as Australia, which are associated with the replacement of tools due to wear by continuous use (Malvajerdi, 2023). Thus, we suggest conducting future research focused on the abrasive wear of each of the materials studied herein, in order to enhance the selection process.

Conclusions

The Von Mises stress calculated for the six manual tool materials, which were simulated using the ANSYS software, did not exceed the elastic limit. Therefore, under normal operating conditions, plastic deformation is unlikely, as well as the breakage of the materials, due to impact against the soil. Finite element analysis determined that the material

with the highest Von Mises stress was AISI/SAE 4140 steel (16.30 MPa), and the lowest value was reported by A356 aluminum (8.39 MPa).

Through the ANSYS fatigue analysis, it was predicted that the tools would not exhibit fatigue failures and would feature a great number of operating cycles, with the most critical value reported for A356 aluminum, corresponding to a useful life of 1.75×10^{13} cycles. Meanwhile, the lowest critical value regarding fatigue was observed in AISI/SAE 1060, steel with 1×10^{32} cycles. Hence, all the materials are suitable for tool design, offering a long lifespan and meeting the stress requirements for completing their tasks.

As for ergonomics, A356 aluminum exhibits a good behavior, given that it is less heavy and, consequently, the efforts generated are lower, enabling an efficient tillage while ensuring the operator's comfort.

Acknowledgements

We would like to thank Universidad Nacional de Colombia, Medellín campus, for funding the project titled *Centro de excelencia (CE) para la innovación en mecanización y opciones energéticas para la agricultura familiar* [Excellence center for innovation regarding mechanization and energy options for family agriculture], with Hermes code 52937.

CRedit author statement

Laura Delgado Bejarano: conceptualization, formal analysis, investigation, writing-original draft, writing (review and editing), visualization.
Hugo González Sánchez: conceptualization, project administration, resources, writing (review and editing).
Germán García Monsalve: conceptualization, project administration, resources, writing (review and editing).

References

Agudelo, P. (2013). *Estudio de desgaste de recubrimientos duros en una herramienta de arado tipo cincel* Master's thesis, Universidad Nacional de Colombia. https://repositorio.unal.edu.co/handle/unal/20967

Ahmed, A., Wahab, M. S., Raus, A. A., Kamarudin, K., Bakhsh, Q., and Ali, D. (2016). Mechanical properties, material and design of the automobile piston: An ample review. *Indian Journal of Science and Technology, 9*(36). https://doi.org/10.17485/ijst/2016/v9i36/102155

Angus, H. T. (1976). Mechanical, physical and electrical properties of cast iron. In Cast Iron: Physical and Engineering Properties (2nd ed., pp. 34–160). Elsevier. https://doi.org/10.1016/B978-0-408-70933-0.50007-4

Ashburner, J. E., and Kienzle, J. (2013). *Agricultural hand tools in emergencies: Guidelines for technical and field officers*. FAO Plant Production and Protection Paper.

Balvanz, J., and Balvanz, R. (*2019*). *Vertical tillage tool. U.S. Patent No. 10,314,221 B2*. https://patents.google.com/patent/US10314221B2/en

Bednář, R., Votava, J., Červinka, J., and Fajman, M. (2013). Suitability of technical materials for machinery subsoilers for soil tillage. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis, 61*(1), 9–16. https://doi.org/10.11118/actaun201361010009

Benos, L., Tsaopoulos, D., and Bochtis, D. (2020). A review on ergonomics in agriculture. part II: Mechanized operations. *Applied Sciences* (*Switzerland*), *10*(10), 3484. https://doi.org/10.3390/app10103484

Botta, G. (2003). *Guía de clases maquinaria agrícola, Licenciatura en Negocios Agropecuarios*. Universidad de La Pampa, Facultad de Agronomía.

Budynas, R. G., and Nisbett, J. K. (2015). *Diseño de ingeniería mecánica de Shigley* (10th ed.). McGraw Hill.

Charco, J. L. (2017). *Fabricación de un prototipo de block de motor de combustión interna de dos tiempos por manufactura aditiva indirecta con una aleación de aluminio al silicio* [Undergraduate thesis, Escuela Politécnica Nacional]. https://bibdigital.epn.edu.ec/handle/15000/18849

Cruz-Avilés, A., Ortiz-Domínguez, M., Suarez-López, Y., RiveraLanderos, E. A., Lechuga Canto, C. B., Muños-Sánchez, Y., Santana-Robles, F., and García-Pérez, C. (2018). Teoría de falla bajo carga estática. *Ingenio y Conciencia Boletín Científico de La Escuela Superior Ciudad Sahagún, 5*(10), 3319. https://doi.org/10.29057/ess.v5i10.3319

De Almeida Luz, W. (2019). *Desgaste em ferramentas de máquinas agrícolas: revisão bibliográfica* [Thesis, Universidade Tecnológica Federal Do Paraná]. http://repositorio.roca.utfpr.edu.br/jspui/bitstream/1/14778/1/LD_COEME_2019_2_05.pdf

Dutt, A. (2015). Effect of mesh size on finite element analysis of beam. *International Journal of Mechanical Engineering, 2*(12), 8-10. https://doi.org/10.14445/23488360/IJME-V2I12P102

Enríquez Berciano, J. L., and Tremps Guerra, E. (2012). *Fabricación de fundición nodular*. Fundidores. https://oa.upm.es/14450/1/Fundición_nodular.pdf

Espejo, É., and Hernandez, H. (2017). *Análisis de fallas de estructuras y elementos mecánicos* (1st ed.). Universidad Nacional De Colombia.

Fernanda, T., Gatamorta, F., and Oliveira, M. (2018). A *Influência das Condições Ambientais na Porosidade de rodas de ligas leves*. https://doi.org/10.26678/ABCM.CONEM2018.CON18-1018

Gegel, G. A. (1998). *Earthworking machine ground engaging tools having cast-in-place abrasion and impact resistant metal matrix composite components* (*Patent No. 5,743,033*). Google Patents.

González Cobas, J. (2018). Lexicografía y sistematicidad. Acerca de los nombres que designan instrumentos agrícolas. *Revista de Lexicografía, 22*, 135-150. https://doi.org/10.17979/rlex.2016.22.0.3324

Herring, D. (2018). *Heat treatment of cast irons*. Industrial Heating. https://www.industrialheating.com/articles/94644-heat-treatment-of-cast-irons

Jaramillo, D. F. (2014). *El suelo: origen, propiedades, espacialidad* (2nd ed.). Universidad Nacional de Colombia, Sede Medellín.

Leslie, J. F. M., and Aguila, M. V. G. (2014). Análisis del comportamiento resistivo del implemento de trabajo del multiarado MAU-250. *Acta Agronomica, 64*(1), 22-29. https://doi.org/10.15446/acag.v64n1.42849

Malvajerdi, A. S. (2023). Wear and coating of tillage tools: A review. *Heliyon, 9*(6), e16669. https://doi.org/10.1016/j.heliyon.2023.e16669

Marcon, P. (2018). *Agent of deterioration: Physical forces*. Government of Canada. https://www.canada.ca/en/conservation-institute/services/agents-deterioration/physical-forces.html

Matweb (2016). *Online Materials Information resource*. http://www.matweb.com/

Mazoyer, M., and Roudart, L. (2006). *A history of world agriculture*. Earthscan. http://base.dnsgb.com.ua/files/book/Agriculture/History-of-Agriculture/A-History-of-World-Agriculture.pdf

Molino, N., Bridson, R., and Fedkiw, R. (2003). *Tetrahedral mesh generation for deformable bodies*. https://graphics.stanford.edu/papers/meshing-sig03/meshing.pdf

Mott, R. L. (2009). *Resistencia de materiales* (vol. 5). Pearson Educación.

Norton, R. (2011). *Diseño de máquinas: un enfoque integrado* (4th ed.). Pearson Educación.

O'Donell, H., and De Estrada, D. (2014). Presencia irlandesa en la milicia española. *Revista Internacional de Historia Militar, 92*, 555844. https://dialnet.unirioja.es/servlet/libro?codigo=555844

Pérez-González, A., Iserte-Vilar, J. L., and González-Lluch, C. (2011). Interpreting finite element results for brittle materials in endodontic restorations. *BioMedical Engineering OnLine, 10*(1), 44. https://doi.org/10.1186/1475-925X-10-44

Sanahuja, E. (1971). Instrumental de hierro agrícola e industrial de la época ibero-romana en cataluña. *Pyrenae, 7*, 61-110. https://core.ac.uk/download/pdf/39114287.pdf

Seeni, A., Rajendran, P., Hussin, M., and Ismail, F. (2020). Errors and uncertainties in simulation of steady, viscous flow past a circular cylinder at re = 20 using a systematic approach. *INCAS Bulletin, 12*(3), 203-217. https://doi.org/10.13111/2066-8201.2020.12.3.17

Sims, B. G., O'Neill, D. H., Walle, R. J., Ellis-Jones, J., Rivera Rosero, A., and Jirón Estrada, J. (1998). Ergonomic evaluation of hand-hoes for hillside weeding and soil preparation in Honduras. *Ceiba, 39*(2), 183-189.

Singh, J., Chatha, S. S., and Sidhu, B. S. (2020). Abrasive wear behavior of newly developed weld overlaid tillage tools in laboratory and in actual field conditions. *Journal of Manufacturing Processes, 55*, 143-152. https://doi.org/10.1016/j.jmapro.2020.03.040

SolidWorks (2022). *Mesh quality checks*. https://shorturl.at/qzEKO

Teko, D. R., Rouge, B., and Cormier, J. (2018). *Base cutter blade* (*Patent No. US 2018/0139901 A1*). https://patentimages.storage.googleapis.com/fc/a4/9d/21d94e1962fdee/US20180139901A1.pdf

Toni, S., Corrêa, C. C., de Souza, D. C., Muzeka, G., and Rossa, Ü. B. (2017). Enxada iracema, cavador tatuí e subsolador vangador de solo como novas ferramentas agrícolas como promotoras de tecnologias as propriedades rurais. *Anais Da Mostra de Ensino, Pesquisa, Extensão e Cidadania (MEPEC), 2*, 83-87. https://publicacoes.ifc.edu.br/index.php/MEPEC/article/view/229

Vanegas Useche, L. V. (2018). *Diseño de elementos de máquinas*. Universidad Tecnológica de Pereira. https://doi.org/10.22517/9789587223019

Yazici, A. (2011). Wear behavior of carbonitride-treated ploughshares produced from 30MnB5 steel for soil tillage applications. *Metal Science and Heat Treatment, 53*, 46-51.

Yu, H.-J., and Bhole, S. D. (1990). Development of a prototype abrasive wear tester for tillage tool materials. *Tribology International, 23*(5), 309-316. https://doi.org/10.1016/0301-679X(90)90004-9

Yusuf, H., and Yusuf, D. (2006). *Ergonomics of selected hand tools for soil tillage in Nigeria* [Conference paper]. International Conference on Engineering Research & Development: Impact on Industry. https://www.researchgate.net/publication/365762299%0AErgonomics

Research Article/Chemical, Food, and Environmental Engineering http://doi.org/10.15446/ing.investig.106346

Effect of Convective Drying and Far-Infrared Radiation on the Physical Properties and Microstructure of *Yacón* Chips (*Smallanthus sonchifolius*)

Efecto del secado convectivo y radiación infrarroja lejana sobre las propiedades físicas y microestructura de hojuelas de yacón (*Smallanthus sonchifolius*)

Edwin O. Baldeón[1], Álvaro Bracamonte-Herrera[2], Andrés Soto-Torres[3], Walter F. Salas-Valerio[4], and Julio Vidaurre-Ruiz[5]

ABSTRACT

Convective drying is widely used in the food industry due to its simplicity and versatility, as it allows for better temperature control and heat distribution, which is essential for maintaining product quality. However, this method can be slower compared to infrared drying. The aim of this research was to evaluate the effect of convective and infrared drying on the physical properties and microstructure of *yacón* chips. An infrared dryer and a convection dryer were used to this effect, setting temperatures of 60, 70, and 80 ˚C for both methods. The color and texture properties, as well as the microstructure, changed with the increase in temperature. The greatest color variation in the *yacón* samples was reported by infrared drying at a temperature of 80 ˚ C. Regarding texture, there were no differences between the two methods. The microstructure of the *yacón* samples dried by hot air exhibited more significant cell damage, especially at 60 ˚C, in comparison with infrared drying, which showed a more microporous and compact structure. The results indicate that the method used impacted shrinkage: infrared drying produced a higher level of shrinkage when compared to convective drying. It is important to note that this is a significant physical change that can adversely affect the quality of dehydrated food. Infrared drying produced greater rehydration in comparison with convective drying. Similarly, improved rehydration was observed at a temperature of 70 °C. In conclusion, the infrared drying method, coupled with appropriate drying conditions, constitutes a good alternative for drying *yacón* chips.

Keywords: heat transfer, internal structure, micropores, drying, hot air, infrared

RESUMEN

El secado convectivo es ampliamente utilizado en la industria alimentaria debido a su simplicidad y versatilidad, ya que permite un mejor control de la temperatura y distribución del calor, lo cual es esencial para mantener la calidad del producto. Sin embargo, este método puede ser más lento en comparación con el secado por infrarrojos. El objetivo de esta investigación fue evaluar el efecto del secado convectivo y por infrarrojos en las propiedades físicas y la microestructura de las rodajas de yacón. Para ello, se utilizaron un secador por infrarrojos y un secador de convección, estableciendo temperaturas de 60, 70 y 80 ˚C para ambos métodos. Las propiedades de color y textura, así como la microestructura, cambiaron con el aumento de la temperatura. La mayor variación de color en las muestras de yacón correspondió al secado por infrarrojos a una temperatura de 80 ˚C. En cuanto a la textura, no hubo diferencias entre los dos métodos. La microestructura de las muestras de yacón secadas con aire caliente mostró un daño celular más significativo, especialmente a 60 ˚C, en comparación con el secado por infrarrojos, que mostró una estructura más microporosa y compacta. Los resultados indican que el método utilizado impactó en la contracción: el secado por infrarrojos produjo un nivel de contracción más alto en comparación con el secado convectivo. Es importante señalar que este es un cambio físico significativo que puede afectar negativamente la calidad de los alimentos deshidratados. El secado por infrarrojos produjo una mayor rehidratación en comparación con el secado convectivo. De manera similar, se observó una mejor rehidratación a una temperatura de 70 °C. En conclusión, el método de secado por infrarrojos, en combinación con condiciones de secado adecuadas, constituye una buena alternativa para el secado de rodajas de yacón.

Palabras clave: transferencia de calor, estructura interna, microporos, secado, aire caliente, infrarrojo

Received: April 4th, 2023
Accepted: January 11th, 2024

[1] Food engineer, Universidad Nacional Agraria de la Selva, Peru. PhD. In Food Engineering, Universitat Politècnica de València, España. Affiliation: Senior professor, Universidad Nacional Agraria La Molina, Peru. E-mail: vidaurrejm@lamolina.edu.pe

[2] Food engineer, Universidad Nacional Agraria La Molina, Peru. Affiliation: Young research, Universidad Nacional Agraria La Molina, Peru. E-mail: 20141340@lamolina.edu.pe

[3] Food engineer, Universidad Nacional Agraria La Molina, Peru. Affiliation: Young research, Universidad Nacional Agraria La Molina, Peru. E-mail: 20141372@lamolina.edu.pe

[4] Food engineer, Universidad Nacional Agraria La Molina, Peru. MSc. In Agriculture Engineering, Michigan State University, United States of America. Affiliation: Senior professor, Universidad Nacional Agraria La Molina, Peru. E-mail: wfsalas@lamolina.edu.pe

[5] Food engineer, Universidad Nacional Pedro Ruiz Gallo, Peru. PhD. In Food Science, Universidad Nacional Agraria La Molina, Peru. Affiliation: Associate professor, Universidad Nacional Agraria La Molina, Peru. E-mail: vidaurrejm@lamolina.edu.pe

Introduction

Yacón is a native crop of the Andes region, grown at various altitudes in Peru, Bolivia, Ecuador, and Argentina. Due to its medicinal properties, it has been part of the Andean diet for centuries and has spread to other countries like New Zealand, Europe, the USA, and Japan. In Europe, the Czech Republic has pioneered its cultivation, while it is gaining importance in Brazil, particularly in the State of São Paulo. The name *yacón* comes from the Quechua term *yaku*, meaning 'water', but this crop has different names in various regions. It belongs to the genus Smallanthus, *i.e.*, *Smallanthus sonchifolius*, in the family Asteraceae. In recent years, *yacón* has gained popularity in other areas outside its native region, given its nutritional profile and potential health benefits. It has become an important crop in some South American areas and has been exported to other countries for consumption and use in the food and supplement industry (Žiarovská *et al*., 2019; Choque Delgado *et al*., 2013).

Yacón is an Andean natural resource, similar in appearance to potatoes, with a sweet flavor and crispy pulp (Bernstein and Noreña, 2014). It has been proven that the consumption of yacón has beneficial effects on health, since it has antioxidant properties, reduces blood sugar levels, and is a potential probiotic (Valentová and Ulrichová, 2003). This is due to its high content of fructooligosaccharides and inulin, which promote the development of the colonic microbiota (Choque Delgado *et al*., 2013). Its consumption takes place after a period of exposure to the sun aimed at increasing its sweetness (Bernstein and Noreña, 2014). However, *yacón's* high water content (83-90%) and the presence of enzymes such as polyphenol oxidase and peroxidase make it a perishable food (Shi *et al*., 2013). Its shelf life in fresh form is approximately seven days under ambient conditions, exhibiting constant depolymerization of the fructooligosaccharide compounds, which are of interest to consumers (Perussello *et al*., 2014).

Different technological alternatives are being studied to preserve *yacón*, transforming the product into beverages, instant powders, sweets, purées, filters, and snacks, among others (Franco *et al*., 2016; Choque Delgado *et al*., 2013), always with the aim of not reducing its nutritional or bioactive compounds (Reis *et al*., 2012; Campos *et al*., 2016).

Convective drying is widely used in the food industry due to its simplicity and versatility. It allows for better temperature control and even heat distribution around the food, which is essential for maintaining product quality. However, this method can be slow in comparison with infrared drying, and high temperatures and extended drying times can adversely affect quality in terms of the taste, color, and texture of the food (Marques *et al*., 2023). Convective drying is the simplest method to decrease the moisture content of a product, as heat and mass transfer occur in its surroundings (Reis *et al*., 2012).

Recent research has shown that pretreatments such as osmotic dehydration or the application of organic acids improve the bioactive characteristics of *yacón*, even when dried at high temperatures (Khajehei *et al*., 2018; Campos *et al*., 2016; Perussello *et al*., 2014). However, convective drying (CD) affects sensory properties such as the color and texture of foods, which is why some studies have sought alternatives to make drying processes less drastic and more economical. An example of this is vacuum drying, which has shown positive effects on *yacón* slices exhibiting low browning, a golden yellow color (Reis *et al*., 2012).

Another alternative not yet studied in *yacón* is drying by infrared radiation, a type of non-contact heat transfer that harnesses the propagation of electromagnetic waves and does not require a medium. Infrared energy radiates to the heated surface and penetrates directly into the inner layer of the material. This energy is absorbed by molecules in different layers of the material, causing the vibrational energy level of the molecules to rise and fluctuate, generating heat and increasing the overall temperature. This is one of the most important advantages of infrared radiation, as it avoids energy losses and considerably maintains the original quality of the product. However, the initial acquisition and installation costs of infrared equipment are typically high, and precise temperature control can be a challenging task. Moreover, the limited penetration of infrared energy can affect the uniformity of dehydration in products with irregular geometries (D. Huang *et al*., 2021; Zeng *et al*., 2019).

Infrared drying provides faster dehydration by transferring energy directly to the food, helping to preserve product quality (*e.g.*, color and flavor). Additionally, it is more energy-efficient due to its shorter drying times.

Recent research has revealed the benefits of drying fruits using far-infrared radiation, as is the case of apples (El-Mesery and Mwithiga, 2015), kiwis (Zeng *et al*., 2019), and mangoes (Yao *et al*., 2020), among others. However, the effect of infrared drying on *yacón* chips at different temperatures has not yet been reported. Thus, the objective of this research was to evaluate the effects of convective and far-infrared radiation drying at three different temperature levels on the physical properties and microstructure of *yacón*.

This article consists of four sections. The first section covers the raw material conditioning process and provides a detailed description of the drying methods employed, as well as the analysis procedures conducted on the samples. The second section focuses on presenting the results obtained, encompassing the physical and microstructural properties of *yacón* chips. The third section of the article is devoted to discussing these results, which are analyzed and compared against those in the specialized literature. Finally, the fourth section provides the conclusions of this study and proposals for future research concerning the drying process of *yacón* chips. The results constitute a reference for studying the effects of far-infrared radiation drying on the physical quality

components of *yacón*, aiming for a functional product with acceptable characteristics.

Materials and methods

Raw material conditioning

Yacón roots (*Smallanthus sonchifolius*) were purchased from the Santa Rosa market, located in the district of La Molina, Lima, Peru. The moisture content of the fresh *yacón* samples was analyzed via the 930.04 method (AOAC, 2005). The results showed a moisture content of 89.48 ± 0.72%, which is similar to the values reported in previous studies (Corrêa *et al*., 2021). A selection process was carried out, eliminating the samples that exhibited physical damage, such as bruises or signs of microbial attack, evidenced in color changes on the surface. Then, the selected roots were washed with by immersion in potable water to eliminate any residues from the harvesting process. Disinfection was carried out with sodium hypochlorite (150 ppm). The water was removed from the outer part of the roots, and they were cut into 5 mm thick slices using a vegetable slicer (CL-52, Robot Coupe, United States). Chemical bleaching was performed by immersing the *yacón* slices in a 2% citric acid solution for 5 min at room temperature. This was done to inactivate the enzymes present in the *yacón* roots that are related to enzymatic browning (peroxidase and polyphenoloxidase).

Drying process

The *yacón* slices were placed on stainless steel mesh trays. They were dried in a convective cabin dryer (TAAC-PC, Edibon, Spain) at an air velocity of 0.91 m/s and in a far-infrared dryer (IRCDi8, IR Confort, Spain) at temperatures of 60, 70, and 80 °C. The temperature range for dehydrating *yacón* was selected while considering commercial working conditions; this range efficiently expedites the moisture removal process in *yacón*, saving both time and energy. The infrared dryer contained eight trays whose area and power were 0.24 m^2 and 221 W, respectively. Both drying operations ended when the product reached a humidity range of 8-10%. The dehydrated *yacón* samples were packed in polypropylene bags at room temperature for later characterization. An infrared moisture meter (MX-50, AND, Japan) was used to determine the moisture content of the dried *yacón* samples.

Yacón drying curves

The drying curve was constructed by plotting the moisture content *vs.* time, and the drying rate curve was plotted *vs.* the moisture content. The moisture content (X) was determined using Equation (1), which calculates the amount of water in kilograms per kilogram of dry matter.

$$X = \frac{W - W_{dm}}{W_{dm}} \quad (1)$$

where W represents the weight at a specific time, and W_{dm} denotes the weight of the dry matter.

The drying rate was calculated using Equation (2), which measures the amount of water removed in kilograms per kilogram of dry matter per hour.

$$\varnothing = -\frac{dX}{dt} = -\frac{\Delta X}{\Delta t} \quad (2)$$

where ϕ represents the drying rate ($kg_w/kg_{dm}h$), X is the free moisture (kg_w/kg_{dm}), and t represents the drying time (h).

Color determination

The color of the samples was expressed in accordance with the CIELAB parameters: L* (0=black; 100=white), a* (−a*=greenish, +a*=red), and b* (−b*=bluish, +b*=yellow). The colorimeter (CR-400, Konica Minolta, Japan) was placed vertically on the surface of the sample (laid on a white surface). Measurements were taken in three replicates. The total color difference (ΔE), indicating the color change between fresh and dry samples, was calculated using the following Equation (Yao *et al*., 2020):

$$E^* = \sqrt{\left(L_0^* - L_1^*\right)^2 + \left(a_0^* - a_1^*\right)^2 + \left(b_0^* - b_1^*\right)^2} \quad (3)$$

Texture analysis

The methodology proposed by Egea *et al*. (2012) was used, with some modifications. A texturometer (Model 3345, Instron, USA) was used, equipped with a 0.25-inch diameter plunger. The speed of the penetration test was 1.0 mm/s. The breaking point (expressed in Newtons) of the dehydrated *yacón* was determined.

Volumetric shrinkage (Sh)

To determine the initial volume (V_0) of the *yacón* samples, their diameter and thickness were measured using a digital Vernier caliper with a precision of 0.01 mm. The area of each sample was digitally analyzed using the ImageJ software, and its thickness was measured at four different points using the digital Vernier. Four replicates of the measurements were performed. Sh was calculated with respect to the volume of the sample at each time (V) and its initial volume (V_0) according to Equation (4) (Senadeera *et al*., 2020; Corrêa *et al*., 2021).

$$Sh = \left(1 - \frac{V}{V_0}\right) * 100 \quad (4)$$

where V and V_0 are the volume of the sample at each time and the initial volume (m^3), respectively.

Rehydration

To measure the rehydration ratio, 5 g of dried *yacón* were soaked in 150 ml of distilled water in a 250 ml beaker at 25 °C. After 1 h, the samples were placed on a piece of paper towel to remove any excess water from the surface. The rehydrated mass was then determined three times to ensure accuracy. Three replicates of the measurements were performed. The rehydration ratio was calculated using Equation (2) (Mugodo and Workneh, 2021; Corrêa *et al*., 2021).

$$\text{Rehydration ratio} = \frac{W_2}{W_1} \quad (5)$$

where the weights W_2 and W_1 (g) correspond to the drained and dried *yacón* samples, respectively.

Microstructure analysis

Photographs were taken using a scanning electron microscope (SEM). Fresh and dehydrated *yacón* chips were plated with gold under vacuum conditions using an automatic metallizer (Q150R, Quorum Technologies, England). The samples were subsequently analyzed using a SEM-Q250 from Thermo Scientific Analytical. The kV value was set as 15.00, and the gain was 300 µm.

Statistical analysis

All experiments were performed in triplicate, and the treatments were analyzed within a completely randomized design (CDR). The results were averaged, and an analysis of variance (ANOVA) was performed for each determination. Then, a comparison of means was made, using the Tukey test at a significance level of 0.05. All statistical analyses were performed using the Statgraphics Centurion XVII statistical package.

Results

Figure 1 shows the moisture contents (experimental) *vs.* the drying time for the two methods. The results indicate that the technique used has an effect on the drying rates of *yacón*. Using the far-infrared drying (ID) method increases the drying rate and reduces the time required to achieve a certain moisture content.

Figure 1 displays typical drying curves, characterized by two falling-rate periods and no apparent constant-rate period. However, it might be possible to have a very short constant-rate period at lower moisture values. The air temperature (60, 70, and 80 °C) impacted the drying kinetics of *yacón* (Figure 1a); increasing the temperature of the drying medium increased the drying potential and the moisture removal rates.

Table 1 shows the different color coordinates obtained for each drying treatment. In the case of CD, the luminosity of the *yacón* samples decreased as the drying temperature increased. This was also reported by ID. However, it should be noted that the greatest decrease in luminosity was obtained with ID at 80 °C. This phenomenon could be attributed to the duration of the process, since it was longer than 5 h, compared to 3 h for CD. This means that, at a temperature of 80 °C, the latter is faster than the former. The reduced values of the color coordinate a* caused yellowish tones in the *yacón* samples. In CD, the increase in a* was progressive as the drying temperature increased. In the case of ID, the increase in the value of a* was more abrupt, producing browner shades. As for b*, no significant changes were observed, indicating that the green tones remained constant.

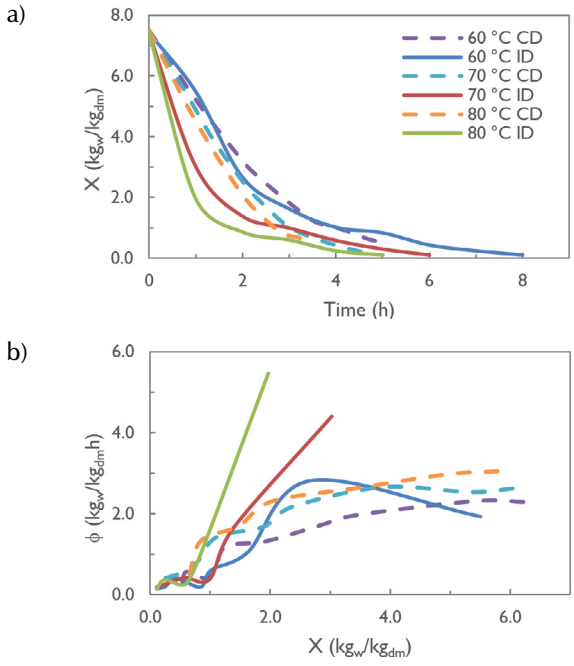


Figure 1. Drying curves of *yacón* for the two studied methods at different temperatures: a) moisture *vs.* time, b) drying rate curve. CD: convective drying, ID: far-infrared drying.

Table 1. Physical properties (color coordinates and texture) of dehydrated *yacón*

	T (°C)	Physical properties				
		Color				Texture
		L*	a*	b*	ΔE	Force (N)
CD	60	62.31 ± 0.38^{a}	1.05 ± 0.13^{a}	23.74 ± 2.07^{a}	27.10 ± 0.56^{a}	18.54 ± 2.36^{a}
	70	58.89 ± 0.95^{ab}	2.97 ± 0.21^{b}	27.88 ± 3.57^{a}	30.42 ± 0.62^{ab}	14.57 ± 1.72^{ab}
	80	56.99 ± 1.25^{b}	3.61 ± 0.91^{bc}	25.90 ± 3.53^{a}	32.38 ± 0.70^{b}	12.78 ± 1.62^{b}
ID	60	61.74 ± 3.10^{a}	0.71 ± 0.02^{a}	22.82 ± 0.60^{a}	22.78 ± 3.17^{ab}	18.67 ± 2.01^{a}
	70	56.71 ± 2.18^{b}	4.18 ± 0.29^{c}	25.86 ± 0.58^{a}	32.54 ± 2.55^{b}	16.80 ± 2.02^{ab}
	80	42.94 ± 0.58^{c}	5.50 ± 0.42^{d}	24.28 ± 0.40^{a}	46.44 ± 0.83^{c}	13.85 ± 2.04^{b}

Different letters (a, b, c) within the same column show significant differences between values (p<0.05). Data are reported as the mean of three replicates (n=3) ± standard deviation (SD).

The results showed that the greatest color variations with respect to the fresh *yacón* sample occurred during ID at 80 ˚C. On the other hand, the smallest variations were evidenced during ID at 60 ˚C.

Table 1 also shows the force values required to break a *yacón* slice. It was observed that, as the drying temperature increases, the maximum breaking point decreases. This is due to the fact that higher temperatures entail a greater removal of water from the matrix, making the product more fragile.

Table 2 summarizes the shrinkage obtained for each drying treatment. The results indicate that the drying method affects shrinkage, but the temperature has no effect. ID produced a higher level of shrinkage when compared to CD. Shrinkage is a significant physical change that can adversely affect the quality of dehydrated food.

Table 2. Physical properties (shrinkage and rehydration) of dehydrated *yacón*

	T (˚C)	Physical properties		
		Moisture (%)	Shrinkage (%)	Rehydration
CD	60	9.85 ± 0.74	72.68 ± 0.77^{a}	3.45 ± 0.06^{a}
	70	9.89 ± 1.02	74.97 ± 0.50^{a}	3.93 ± 0.04^{b}
	80	8.15 ± 0.35	75.32 ± 2.78^{a}	3.48 ± 0.07^{a}
ID	60	9.56 ± 0.74	82.14 ± 4.50^{b}	3.87 ± 0.09^{b}
	70	8.08 ± 1.89	79.64 ± 4.73^{b}	4.30 ± 0.13^{c}
	80	9.20 ± 1.13	81.60 ± 1.35^{b}	4.30 ± 0.16^{c}

Different letters (a, b, c) within the same column show significant differences between values ($p<0.05$).

Table 2 also shows the rehydration ratios. The data indicate a significant effect of both drying method and temperature. ID produced greater rehydration in comparison to CD. Similarly, improved rehydration was observed at a temperature of 70 ˚C.

Figure 2 shows the surface microstructure of the fresh and dried *yacón* samples under different drying conditions. The fresh sample (Figure 2a) exhibits round, elongated, compact, and well-structured parenchyma cells. CD produced severe surface shrinkage, manifesting as a brittle and poor-quality surface (Figures 2b and 2c). In contrast, ID showed more micropores and less shrinkage and deformation (Figures 2d and 2e).

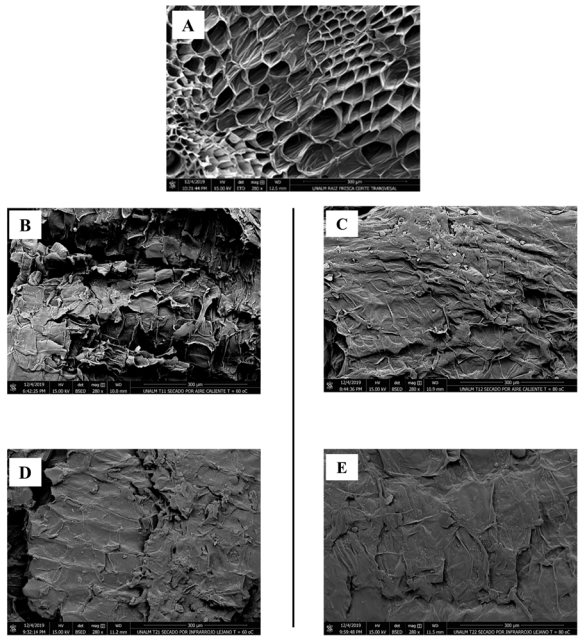


Figure 2. SEM of the surface of *yacón* root slices. A: control; B: convective drying at 60 ˚C; C: convective drying at 80 ˚C; D: infrared drying at 60 ˚C; E: infrared drying at 80 ˚C.

Discussion

ID was found to require shorter times than CD, as the dryer absorbs infrared radiation, causing heat to be released from the interior of the sample. As a result, water is carried from the inside of the sample to its surface, facilitating quick drying. The results were consistent with those of the specialized literature. For instance, during the drying process of *Ganoderma lucidum*, a study found that using an infrared dryer allowed reaching the desired moisture content faster and provided superior performance in comparison with CD (Naseri *et al*., 2023). The *yacón* chips exhibited higher initial drying rates, potentially due to the evaporation of surface moisture. As the moisture content decreased, the drying rates also decreased, suggesting that moisture diffusion played a significant role. The high drying rates observed could be linked to internal heat generation. The absence of a constant drying rate period may be due to the fact that the thin slices did not provide a consistent supply of moisture during drying (Sadeghi*et al*., 2020). Other studies have also reported increased drying rates with higher radiation intensity (Sadeghi *et al*., 2020; Naseri *et al*., 2023). Furthermore, the curve could indicate a constant drying rate stage if corrected for the actual exchange surface area of the sample (Marques *et al*., 2023).

In this study, the total color variation (ΔE) is an indicator of treatment severity with regard to coloration of the initial sample (Bernstein and Noreña, 2014; Ning *et al*., 2015). This can be attributed to the fact that infrared radiation generates rapid and intense heat inside the material, causing serious damage to cell tissues and increasing the likelihood

of contact between the substrate and the enzymes, thereby darkening the *yacón*. However, Nowak and Lewicki (2005) state that drying with infrared technology can prevent the excessive browning of food, giving it a better appearance than that provided by traditional methods. The magnitudes of color variation were on the order of those reported by Bernstein and Noreña (2014), who blanched *yacón* slices for 4 min with steam at 100˚C before subjecting them to a first drying period at 50˚C for 5 h and a second drying period at 75˚C for 5 h.

Several studies have observed that drying with infrared technology decreases hardness in comparison with CD. This is the result of the swelling of the matrix due to the heating of the starches and the solubilization of the pectin, which can produce infrared radiation (Nathakaranakule *et al*., 2010; Qi *et al*., 2014). However, this study noted no difference in this parameter between both drying technologies. As explained above, the infrared radiation exposure time is crucial to reap the benefits of this technology. Therefore, a correct selection of power and working conditions will be crucial to obtaining better results.

Studies on the drying of persimmon and *yacón* using hot air at varying temperatures showed that a value of 60˚C results in minimal shrinkage (Senadeera *et al*., 2020; Marques *et al*., 2023). Furthermore, Yang *et al*. (2020) investigated the drying of mushrooms using different drying methods, including hot air and infrared technologies. They found that there is greater shrinkage in ID. This was also achieved in *yacón* slices. According to Mugodo and Workneh (2021), the techniques employed in drying mango slices significantly impact rehydration. This impact could include irreversible damage, causing cell rupture and shrinkage. This integrity loss and shrinkage subsequently decrease the hydrophilic properties of the cell. As a result, it cannot absorb enough water to fully hydrate the product (Corrêa *et al*., 2021).

The dried *yacón* microstructure was similar to that of dried hemp, kiwi, and beet products (Junqueira *et al*., 2018; Jiang *et al*., 2018; Pham *et al*., 2018). The observed shrinkage is due to the heat transfer caused by CD from the outside to the inside, causing the surface water to evaporate rapidly and forming hard membranes and surface cells with irregular contractions (Zhu et. al, 2022). In contrast, ID reported less shrinkage and more micropores, as one advantage of this technology is rapid heating, which causes the internal cell cavity to heat and expand, producing more micropores in the structure (X. Huang *et al*., 2021). This benefits the surface since the cells are subjected to less structural stress, thus forming more open structures with good morphology in the dehydrated *yacón*. Similarly, Puente-Díaz *et al*. (2020) found changes in the microstructure of dried Physalis fruit puree using the infrared-assisted refractory window method at different temperatures, obtaining a rough surface with evident water loss. In their study, Marques *et al*. (2023) analyzed the microstructure of *yacón* to explain the shrinkage phenomenon during CD. They identified two distinct cell types: xylem vessels and parenchyma cells. These researchers found that xylem vessels have thicker cell walls than parenchyma cells, making them less susceptible to collapsing during drying. Their discovery sheds light on the cause behind the cell wall collapse observed in *yacón* slices dried at 60˚C, and it demonstrates one of the advantages attributed to ID to improve the quality of *yacón* chips, in the form of enhancement in the physical structure and integrity of the product.

Conclusions

This research reported, for the first time, the effect of infrared drying on the physical and structural properties of *yacón* chips. It was found that a high temperature (80 ˚C) causes significant changes in the color and texture parameters of the product in both in infrared and convective drying. However, the most substantial changes were perceived at the structural level, as the infrared dried *yacón* chips showed a better morphology. Therefore, applying this technology cannot only increase the number of micropores in the internal structure. It can also improve the heat and mass transfer process, as well as the appearance of *yacón*. Although this work provides the first insight into the benefits of this technology, other aspects should also be explored, such as combined treatments using different technologies, pre-treatments to efficiently avoid browning, and the effect of different power levels during infrared drying.

Acknowledgements

The authors are grateful to the Vice-President for Research of Universidad Nacional Agraria La Molina for the financial support provided under the UNALM-MINEDU agreement (Resolution No. 0410-2019-R-UNALM).

Conflicts of interest

The authors of this paper declare that they have no conflicts of interest.

CRediT author statement

All authors: conceptualization, methodology, software, formal analysis, investigation, data curation, writing (original draft, review, and editing).

References

Bernstein, A., and Noreña, C. P. Z. (2014). Study of thermodynamic, structural, and quality properties of yacon (*Smallanthus sonchifolius*) during drying. *Food and Bioprocess Technology, 7*(1), 148-160. https://doi.org/10.1007/s11947-012-1027-y

Campos, D., Aguilar-Galvez, A., and Pedreschi, R. (2016). Stability of fructooligosaccharides, sugars and colour of yacon

(*Smallanthus sonchifolius*) roots during blanching and drying. *International Journal of Food Science and Technology, 51*(5), 1177-1185. https://doi.org/10.1111/ijfs.13074

Choque Delgado, G. T., da Silva Cunha Tamashiro, W. M., Maróstica Junior, M. R., and Pastore, G. M. (2013). Yacon (*Smallanthus sonchifolius*): A functional food. *Plant Foods for Human Nutrition, 68*(3), 222-228. https://doi.org/10.1007/s11130-013-0362-0

Corrêa, J. L. G., Lopes, F. J., de Mello Junior, R. E., de Jesus Junqueira, J. R., de Angelis Pereira, M. C., and Salvio, L. G. A. (2021). Dried yacon with high fructooligosaccharide content. *Journal of Food Process Engineering, 44*(12), 13884. https://doi.org/10.1111/jfpe.13884

de Jesus Junqueira, J. R., Corrêa, J. L. G., de Mendonça, K. S., de Mello Júnior, R. E., and de Souza, A. U. (2018). Pulsed vacuum osmotic dehydration of beetroot, carrot and eggplant slices: Effect of vacuum pressure on the quality parameters. *Food and Bioprocess Technology, 11*(10), 1863-1875. https://doi.org/10.1007/s11947-018-2147-9

Egea, M. B., Borsato, D., Silva, R. S. dos S. F. da, and Yamashita, F. (2012). Osmo-dehydrated functional product containing fructo-oligosaccharides: Physical, chemical and sensorial characteristics. *Brazilian Archives of Biology and Technology, 55*(6), 927-936. https://doi.org/10.1590/S1516-89132012000600017

El-Mesery, H. S., and Mwithiga, G. (2015). Performance of a convective, infrared and combined infrared-convective heated conveyor-belt dryer. *Journal of Food Science and Technology, 52*(5), 2721-2730. https://doi.org/10.1007/s13197-014-1347-1

Franco, T. S., Perussello, C. A., Ellendersen, L. N., and Masson, M. L. (2016). Effects of foam mat drying on physicochemical and microstructural properties of yacon juice powder. *LWT - Food Science and Technology, 66,* 503-513. https://doi.org/10.1016/j.lwt.2015.11.009

Huang, D., Yang, P., Tang, X., Luo, L., and Sunden, B. (2021). Application of infrared radiation in the drying of food products. *Trends in Food Science and Technology, 110,* 765-777. https://doi.org/10.1016/j.tifs.2021.02.039

Huang, X., Li, W., Wang, Y., and Wan, F. (2021). Drying characteristics and quality of Stevia rebaudiana leaves by far-infrared radiation. *LWT, 140,* 110638. https://doi.org/10.1016/j.lwt.2020.110638

Jiang, Y., Lawrence, M., Ansell, M. P., and Hussain, A. (2018). Cell wall microstructure, pore size distribution and absolute density of hemp shiv. *Royal Society Open Science, 5*(4), 171945. https://doi.org/10.1098/rsos.171945

Khajehei, F., Hartung, J., and Graeff-Hönninger, S. (2018). Total phenolic content and antioxidant activity of yacon (*Smallanthus sonchifolius* poepp. and endl.) chips: Effect of cultivar, pre-treatment and drying. *Agriculture, 8*(12), 183. https://doi.org/10.3390/agriculture8120183

Marques, B., Perré, P., Casalinho, J., Tadini, C. C., Plana-Fattori, A., and Almeida, G. (2023). Evidence of iso-volume deformation during convective drying of yacón: An extended van Meel model adapted to large volume reduction. *Journal of Food Engineering, 341*, 111311. https://doi.org/10.1016/j.jfoodeng.2022.111311

Mugodo, K., and Workneh, T. S. (2021). The kinetics of thin-layer drying and modelling for mango slices and the influence of differing hot-air drying methods on quality. *Heliyon, 7*(6), e07182. https://doi.org/10.1016/j.heliyon.2021.e07182

Naseri, M., Movagharnejad, K., and Nanvakenari, S. (2023). Comparison of the dried properties of Ganoderma lucidum produced by the convective dryer and infrared dryer. *Scientific Reports, 13*(1), 12636. https://doi.org/10.1038/s41598-023-39883-z

Nathakaranakule, A., Jaiboon, P., and Soponronnarit, S. (2010). Far-infrared radiation assisted drying of longan fruit. *Journal of Food Engineering, 100*(4), 662-668. https://doi.org/10.1016/j.jfoodeng.2010.05.016

Ning, X., Lee, J., and Han, C. (2015). Drying characteristics and quality of red ginseng using far-infrared rays. *Journal of Ginseng Research, 39*(4), 371-375. https://doi.org/10.1016/j.jgr.2015.04.001

Nowak, D., and Lewicki, P. P. (2005). Quality of infrared dried apple slices. *Drying Technology, 23*(4), 831-846. https://doi.org/10.1080/DRT-200054206

Perussello, C. A., Kumar, C., De Castilhos, F., and Karim, M. A. (2014). Heat and mass transfer modeling of the osmo-convective drying of yacon roots (*Smallanthus sonchifolius*). *Applied Thermal Engineering, 63*(1), 23-32. https://doi.org/10.1016/j.applthermaleng.2013.10.020

Pham, N. D., Martens, W., Karim, M. A., and Joardder, M. U. H. (2018). Nutritional quality of heat-sensitive food materials in intermittent microwave convective drying. *Food & Nutrition Research, 62,* 1-11. https://doi.org/10.29219/fnr.v62.1292

Puente-Díaz, L., Spolmann, O., Nocetti, D., Zura-Bravo, L., and Lemus-Mondaca, R. (2020). Effects of infrared-assisted refractance window™ drying on the drying kinetics, microstructure, and colour of Physalis fruit purée. *Foods, 9*(3), 343. https://doi.org/10.3390/foods9030343

Qi, L.-L., Zhang, M., Mujumdar, A. S., Meng, X.-Y., and Chen, H.-Z. (2014). Comparison of drying characteristics and quality of shiitake mushrooms (*Lentinus edodes*) using different drying methods. *Drying Technology, 32*(15), 1751-1761. https://doi.org/10.1080/07373937.2014.929588

Reis, Felipe R., Lenzi, M. K., and Masson, M. L. (2012). Effect of vacuum drying conditions on the quality of yacon (*Smallanthus sonchifolius*) slices: Process optimization toward colour quality. *Journal of Food Processing and Preservation, 36*(1), 67-73. https://doi.org/10.1111/j.1745-4549.2011.00555.x

Reis, F. R., Lenzi, M. K., de Muñiz, G. I. B., Nisgoski, S., and Masson, M. L. (2012). Vacuum drying kinetics of yacon (*Smallanthus sonchifolius*) and the effect of process conditions on fractal dimension and rehydration capacity. *Drying Technology, 30*(1), 13-19. https://doi.org/10.1080/07373937.2011.611307

Sadeghi, E., Haghighi Asl, A., and Movagharnejad, K. (2020). Optimization and quality evaluation of infrared-dried kiwifruit slices. *Food Science & Nutrition, 8*(2), 720-734. https://doi.org/10.1002/fsn3.1253

Senadeera, W., Adiletta, G., Önal, B., Di Matteo, M., and Russo, P. (2020). Influence of different hot air drying temperatures on drying kinetics, shrinkage, and colour of persimmon slices. *Foods, 9*(1), 101. https://doi.org/10.3390/foods9010101

Shi, Q., Zheng, Y., and Zhao, Y. (2013). Mathematical modeling on thin-layer heat pump drying of yacon (*Smallanthus sonchifolius*) slices. *Energy Conversion and Management, 71,* 208-216. https://doi.org/10.1016/j.enconman.2013.03.032

Valentová, K., and Ulrichová, J. (2003). *Smallanthus sonchifolius* and *Lepidium meyenii* - Prospective Andean crops for the prevention of chronic diseases. *Biomedical Papers, 147*(2), 119-130. https://doi.org/10.5507/bp.2003.017

Yang, R.-L., Li, Q., and Hu, Q.-P. (2020). Physicochemical properties, microstructures, nutritional components, and free amino acids of *Pleurotus eryngii* as affected by different drying methods. *Scientific Reports, 10*(1), 121. https://doi.org/10.1038/s41598-019-56901-1

Yao, L., Fan, L., and Duan, Z. (2020). Effect of different pretreatments followed by hot-air and far-infrared drying on the bioactive compounds, physicochemical property and microstructure of mango slices. *Food Chemistry, 305,* 125477. https://doi.org/10.1016/j.foodchem.2019.125477

Zeng, Y., Liu, Y., Zhang, J., Xi, H., and Duan, X. (2019). Effects of far-infrared radiation temperature on drying characteristics, water status, microstructure and quality of kiwifruit slices. *Journal of Food Measurement and Characterization, 13*(4), 3086-3096. https://doi.org/10.1007/s11694-019-00231-3

Žiarovská, J., Padilla-González, G. F., Viehmannová, I., and Fernández, E. (2019). Genetic and chemical diversity among yacon [*Smallanthus sonchifolius* (Poepp. et Endl.) H. Robinson] accessions based on iPBS markers and metabolomic fingerprinting. *Plant Physiology and Biochemistry, 14*, 183-192. https://doi.org/10.1016/j.plaphy.2019.05.020

Zhu, J., Liu, Y., Zhu, C., and Wei, M. (2022). Effects of different drying methods on the physical properties and sensory characteristics of apple chip snacks. *LWT, 154,* 112829. https://doi.org/10.1016/j.lwt.2021.112829

Research Article/Chemical, Food, and Environmental Engineering http://doi.org/10.15446/ing.investig.108212

The Formalism of Chemical Thermodynamics Applied to an Oscillatory Multistep Chemical System

El formalismo de la termodinámica química aplicado a un sistema químico oscilatorio con múltiples etapas

Jean P. Montoya[1], Alexander Contreras-Payares[2], and Daniel Barragán[3]

ABSTRACT

The thermodynamic optimization of a process focuses on consumption, production, and efficient use of energy. The unsteady-state nature of batch reactor processing requires describing the set of processes' dynamic behavior for energy optimization. This work aims to apply the formalism of chemical thermodynamics to a multistep chemical system in a batch reactor, aiming for a dynamic description of its evolution to the equilibrium state. As the system of study, we selected a mathematical model called *the Oregonator*, derived from the mechanism of the oscillating Belousov-Zhabotinsky reaction. In the methodology, we used the reaction quotient to evaluate the Gibbs function, the thermodynamic affinity, and the entropy generation as a function of the reaction extent. The results show that the overall reaction fulfills the thermodynamic fundamentals of chemical equilibrium, despite having a non-stoichiometric coefficient. However, the multistep coupled reaction system does not allow verifying compliance with the thermodynamic foundations of chemical equilibrium. We conclude that it is necessary to improve thermodynamic formalism to describe multistep chemical processes as a function of a global reaction extent variable. In this scenario, the entropy production rate emerges as a promising quantity.

Keywords: chemical process simulation, chemical equilibrium, the Oregonator, Belousov-Zhabotinsky reaction, oscillating reactions

RESUMEN

La optimización termodinámica de un proceso se enfoca en el consumo, la producción y el uso eficiente de la energía. La naturaleza de estado no estacionario del procesamiento en reactores tipo *batch* requiere describir el comportamiento dinámico del conjunto de procesos en términos de optimización de energía. Este trabajo tiene como objetivo aplicar el formalismo de la termodinámica química a un sistema químico de múltiples etapas en un reactor tipo *batch*, buscando obtener una descripción dinámica de su evolución hacia el estado de equilibrio. Como sistema de estudio, seleccionamos el modelo matemático conocido como *el Oregonator*, que se deriva del mecanismo de la reacción oscilante de Belousov-Zhabotinsky. En la metodología, utilizamos el cociente de reacción para evaluar la función de Gibbs, la afinidad termodinámica y la generación de entropía en función del grado de avance de la reacción. Los resultados muestran que la reacción global cumple con los fundamentos termodinámicos del equilibrio químico, a pesar de tener un coeficiente no estequiométrico. Sin embargo, el sistema de reacción con múltiples pasos acoplados no permite verificar el cumplimiento de los fundamentos termodinámicos del equilibrio químico. Concluimos que es necesario mejorar el formalismo termodinámico para describir los procesos químicos de múltiples pasos en función de una variable global de avance de la reacción. En este escenario, la velocidad de generación de entropía surge como una cantidad prometedora.

Palabras clave: simulación de procesos químicos, equilibrio químico, el Oregonador, reacción de Belousov-Zhabotinsky, reacciones oscilantes

Received: April 08th, 2023
Accepted: August 18th, 2023

Introduction

The study and characterization of chemical equilibrium are of great importance for describing processes that depend on the phase stability of fluids, such as separation in reactors and reactive and azeotropic distillation in columns (Prausnitz *et al.*, 1998; Saunders and Miodownik, 1998; Tosun, 2021). On the other hand, in reacting systems, the chemical equilibrium defines the upper bound of transformation of a chemical reaction under the initial conditions and the imposed constraints. A closed system that evolves at constant T and p must satisfy the constraints of directionality and equilibrium associated with thermodynamic potentials (Rosenberg and Klotz, 2008; Honig, 2020).

Let us consider a reaction system composed of n simultaneous chemical reactions coupled to each other through intermediary chemical substances. The set of chemical equations that satisfies the law of mass conservation and describes the reactant system composed

[1] Chemical engineer, Universidad Nacional de Colombia, Colombia. MSci in Chemistry and Ms in the Teaching of Sciences, Universidad Nacional de Colombia, Colombia. Email: jpmontoy@unal.edu.co

[2] Chemist, Universidad de Córdoba, Colombia. Email: alcontreraspa@unal.edu.co

[3] Corresponding author. Chemist, Universidad Nacional de Colombia, Colombia. DrSci in Chemistry, Universidad Nacional de Colombia, Colombia. Associate professor, Universidad Nacional de Colombia, Colombia. Email: dalbarraganr@unal.edu.co

of m substances can be written compactly, as shown in Equation (1):

$$\sum_{i=1}^{n} \vartheta_{ir} R_i = 0, (r:1,2,\ldots,m) \tag{1}$$

where ϑ_{ir} is the stoichiometric coefficient of the substance R_i in the *r*-th chemical reaction. The R_i substances distribute between reactants with a negative ϑ_{ir} coefficient and products with a positive ϑ_{ir} coefficient.

The change in the amount of moles N_i of the substance R_i in the *r*-th chemical reaction is determined by the reaction extent ϑ_r, as follows:

$$dN_{ir} = \vartheta_{ir} d\xi_r \tag{2}$$

The total change in the amount of the substance R_i in the reaction system is calculated via Equation (3):

$$dN_i = \sum_{r=1}^{m} \vartheta_{ir} d\xi_r \tag{3}$$

The reaction extent ξ_r, as a state variable of the reaction system, allows evaluating the change in the Gibbs function of a chemical system described by Equation (1) by following Equation (4):

$$dG = \left(\frac{\partial G}{\partial T}\right)_{p,\xi} dT - \left(\frac{\partial G}{\partial p}\right)_{T,\xi} dp + \sum_{i=1}^{n}\left(\sum_{r=1}^{m} \mu_i \vartheta_{ir} d\xi_r\right) \tag{4}$$

At constant T and p, and due to the initial conditions of the set of chemical reactions, the change in the reaction extent has an arbitrary sign. Therefore, the equilibrium condition $dG_{T,p} = 0$ must be satisfied by the closed system of chemical reactions; based on Equation (4), this is described for each *r* reaction in Equation (5):

$$\sum_{i=1}^{n} \mu_i \vartheta_{ir} = 0 \tag{5}$$

The equilibrium condition in Equation (5) can also express a generalized force, *i.e.*, the thermodynamic affinity A, according to Equation (6):

$$A_r = -\sum_{i=1}^{n} \vartheta_{ir} \mu_i = \sum_{a\{reactants\}} \vartheta_a \mu_a - \sum_{b\{products\}} \vartheta_b \mu_b = 0 \tag{6}$$

A contracted form of Equation (6) can be written in terms of a reaction quotient and an equilibrium constant (Dutt, 1985). From Equation (4) and Equation (6), and due to the irreversible nature of chemical reactions, for a reacting mixture in a closed system at constant T and p the Gibbs function reaches a minimum at the chemical equilibrium defined by a thermodynamic constant. The evolution towards and at the equilibrium state of the system described by Equation (1) satisfies the conditions given in Equation (7).

$$\begin{cases} \left(\dfrac{\partial A_r}{\partial \xi_r}\right)_{T,p} < 0 \\ A_r = 0 \end{cases} \tag{7}$$

Thus, as the reaction progresses toward the chemical equilibrium, the thermodynamic affinity decreases until it becomes equal to zero (Solaz-Portolés, 2011; Kondepudi and Prigogine, 2014).

The above-presented formalism allows for a direct account of the irreversible isothermal transformation of matter during a chemical reaction through the generation of entropy (*thermodynamic dissipation*) described in Equation (8) (Prigogine, 1961; Kondepudi and Prigogine, 2014; Barragán *et al*., 2021).

$$\sigma = \frac{d_i S}{dt} = \sum_{r=1}^{m} \frac{A_r}{T} \frac{d\xi_r}{dt} \geq 0 \tag{8}$$

As per Equation (8), the thermodynamic dissipation is zero at the chemical equilibrium, since the affinity and the net reaction rate are zero.

Although the above is widely accepted and duly documented in many texts (Eu and Al-Ghoul, 2018), great doubts arise when applying it to concrete systems (Beretta and Gyftopoulos, 2015), such as those raised by the discovery of oscillating chemical reactions in the 1950s. By then, agreeing on chemical oscillations and the second law of thermodynamics was unclear (Kiprijanov, 2016). Two decades later, once experimentation validated the first observations of chemical oscillations and provided an understanding on how the second law of thermodynamics rationalizes these systems, a fruitful field of research began, known as *nonlinear chemical dynamics* (Epstein and Pojman, 1998). There is extensive development in the formalism and the applications of the chemical equilibrium of phases. However, there are still doubts about the chemical equilibrium in reacting systems, particularly its teaching and the conceptual errors that can affect it (Quílez-Pardo and Solaz-Portolés, 1995; Banerjee, 1995; Rogers *et al*., 2000; Patiño-Sierra and Barragán, 2022; Martínez-Grau *et al*., 2014).

In this work, we test the application of thermodynamics formalism to multistep reacting systems using the mathematical model known as the *Oregonator*, derived from the chemical mechanism of the Belousov-Zhabotinsky (BZ) reaction.

Methodology

This section describes the Oregonator model and the kinetic parameters used in this study. It presents the equations that govern the dynamics of the model's overall reaction and the system of equations for each of the reaction steps. We apply the formalism of thermodynamics to derive the expressions for the reaction coefficient and the Gibbs function. All equations are solved numerically by using an integration subroutine for systems of stiff differential equations, which is available in the MATLAB software.

Multistep chemical system: the Oregonator

The Oregonator model describes the main characteristics of the oscillatory dynamics of the BZ reaction, both in an open and closed system (Field *et al*., 1972; Field and Noyes, 1974). The Oregonator comprises five reaction steps, four for a set of inorganic reactions of the reaction BZ and one for organic reactions: Equations O1 to O5 in Table 1. The reaction Og is the overall reaction of the model, obtained as follows: Og=2O1+O2+O3+O4+O5. The overall reaction Og satisfies the following: A and B reactants have integer stoichiometric coefficients, intermediates X and Z are absent, and the stoichiometric coefficient f is uniquely related to Y. The variables in the Oregonator model correspond to the following chemical species in the BZ reaction: A to sodium bromate, B to malonic acid, Z to cerium (IV) ion catalysts, X to bromous acid, Y to bromide ion, and P to hypobromous acid. Equation O3 is the chemical instability of the model, which is the autocatalytic production of bromous acid. Equation O5 represents a set of organic reactions, with malonic acid, the organic substrate, kinetically controlling the concentration of molecular bromine and bromide ions, which are the species driving the positive feedback loop in the reaction. The stoichiometric coefficient f in equation O5 accounts for the complexity of the chemical reaction and acts as a bifurcation parameter in the Oregonator model (Field and Noyes, 1974). Although there are many versions of the Oregonator model, whose majority does not include the B variable, in this work, we employ the original version derived from the BZ reaction mechanism to use the organic substrate concentration as a dynamical control parameter (Table) 1 (Tyson, 1981, 1982).

Mathematical model: the overall reaction Og

When applying Equation (2) to the overall reaction Og, the change in each variable as a function of the reaction extent is as follows: $A = A_0 - 2\xi$, $B = B_0 - \xi$, $Y = Y_0 - (3-\boldsymbol{f})\xi$, $P = P_0 + 5\xi$.

The reaction rate, expressed as a function of the reaction extent by and applying the mass action law to Og, is shown in Equation (9).

$$\frac{d\xi}{dt} = k_g \left(A_0 - 2\xi\right)^2 \left(B_0 - \xi\right)\left[Y_0 - (3-\boldsymbol{f}\xi)\right]^{3-f} - k_{-g}\left(P_0 + 5\xi\right)^5 \tag{9}$$

Table 1. Set of equations and kinetic constants of the Oregonator model (Tyson, 1981, 1982)[21-22]

Equation	Reaction step	Forward kinetic constant	Backward kinetic constant
O1	$A+Y \leftrightharpoons X+P$	$k_1 = 2,0$ $mol^{-1}Ls^{-1}$	$k_{-1} = 3,3$ $mol^{-1}Ls^{-1}$
O2	$X+Y \leftrightharpoons 2P$	$k_2 = 2\times10^9$ $mol^{-1}Ls^{-1}$	$k_{-2} = 5\times10^{-10}$ $mol^{-1}Ls^{-1}$
O3	$A+X \leftrightharpoons 2X+Z$	$k_3 = 1\times10^4$ $mol^{-1}Ls^{-1}$	$k_{-3} = 33$ $mol^{-2}L^2s^{-1}$
O4	$2X \leftrightharpoons A+P$	$k_4 = 4\times10^7$ $mol^{-1}Ls^{-1}$	$k_{-4} = 1\times10^{-10}$ $mol^{-1}Ls^{-1}$
O5	$B+Z \leftrightharpoons \boldsymbol{f}Y$	$k_5 = 5,0\, mol^{-1}Ls^{-1}$	$k_{-5} = 1\times10^{-11}$ $mol^{1-f}L^{f-1}s^{-1}$
Og	$2A+B+(3-\boldsymbol{f})Y \leftrightharpoons 5P$	$k_g = k_1^2k_2k_3k_4k_5$	$k_{-g} = k_{-1}^2k_{-2}k_{-3}k_{-4}k_{-5}$

Source: Authors

Mathematical model: reaction steps O1 to O5

When applying Equation (3) to the system of equations O1-O5, the change in each of the variables, expressed as a function of the progress of each reaction r, is as follows: $A = A_0 - \xi_1 - \xi_3 + \xi_4$, $B = B_0 - \xi_5$, $X = X_0 + \xi_1 - \xi_2 + \xi_3 - 2\xi_4$, $Y = Y_0 - \xi_1 - \xi_2 + \boldsymbol{f}\xi_5$, $Z = Z_0 + \xi_3 - \xi_5$, $P = P_0 + \xi_1 + 2\xi_2 + \xi_4$. Thus, by applying the law of mass action, Equations (10)-(14) show the reaction rate of the steps in the Oregonator model:

$$d\xi_1/dt = k_1\left(A_0 - \xi_1 - \xi_3 + \xi_4\right)\left(Y_0 - \xi_1 - \xi_2 + \boldsymbol{f}\xi_5\right) - k_{-1}\left(X_0 + \xi_1 - \xi_2 + \xi_3 - 2\xi_4\right)\left(P_0 + \xi_1 + 2\xi_2 + \xi_4\right) \tag{10}$$

$$d\xi_2/dt = k_2\left(X_0 + \xi_1 - \xi_2 + \xi_3 - 2\xi_4\right)\left(Y_0 - \xi_1 - \xi_2 + \boldsymbol{f}\xi_5\right) - k_{-2}\left(P_0 + \xi_1 + 2\xi_2 + \xi_4\right)^2 \tag{11}$$

$$d\xi_3/dt = k_3\left(A_0 - \xi_1 - \xi_3 + \xi_4\right)\left(X_0 + \xi_1 - \xi_2 + \xi_3 - 2\xi_4\right) - k_{-3}\left(Z_0 + \xi_3 - \xi_5\right)\left(X_0 + \xi_1 - \xi_2 + \xi_3 - 2\xi_4\right)^2 \tag{12}$$

$$d\xi_4/dt = k_4\left(X_0 + \xi_1 - \xi_2 + \xi_3 - 2\xi_4\right)^2 - k_{-4}\left(A_0 - \xi_1 - \xi_3 + \xi_4\right)\left(P_0 + \xi_1 + 2\xi_2 + \xi_4\right) \tag{13}$$

$$d\xi_5/dt = k_5\left(B_0-\xi_5\right)\left(Z_0+\xi_3-\xi_5\right)- k_{-5}\left(Y_0-\xi_1-\xi_2+\boldsymbol{f}\xi_5\right)^f \quad (14)$$

Thermodynamics: the overall reaction Og

We assign arbitrary values to the standard chemical potential for the chemical species variables of the overall reaction Og, taking the following values as reference: $\mu_A^0=-5\times10^4\ J/mol$, $\mu_B^0=-10^5\ J/mol$, $\mu_P^0=-10^5\ J/mol$. From Equation (4), at constant T and p , the change in the Gibbs function during reaction r is shown in Equation (15).

$$\left(\frac{dG}{d\xi}\right)_{T,p}=\Delta_r G=\sum_{i=1}^{n}\vartheta_i\mu_i \quad (15)$$

Equation (15) is written as Equation (16) after using the equation for the chemical potential $\mu_i=\mu_i^0+RTlnC_i$, where C_i indicates the molar concentration of substance i .

$$\Delta_r G=\Delta_r G^0+RTlnQ \quad (16)$$

Here, Q is the reaction quotient of the reaction, and, at the chemical equilibrium, the reaction quotient is equal to the equilibrium constant $Q=K$, thus obtaining Equation (17).

$$\Delta_r G^0=-RTlnK \quad (17)$$

Equation (17) is fundamental in providing consistency to the thermodynamic calculations of chemical equations. For the overall reaction Og in Equation (15), Equation (18) is obtained.

$$\Delta_r G^0=5\mu_P^0-2\mu_A^0-\mu_B^0-\left(3-\boldsymbol{f}\right)\mu_Y^0 \quad (18)$$

where the standard chemical potential of variable Y is defined from Equation (17) as follows:

$$\mu_Y^0=\frac{5\mu_P^0-2\mu_A^0-\mu_B^0+RTlnK}{\left(3-\boldsymbol{f}\right)} \quad (19)$$

Finally, for the overall reaction Og, we have the following expressions for the reaction quotient and the Gibbs function, as a function of the reaction extent, *i.e.*, Equations (20) and (21):

$$Q=\frac{\left(P_0+5\xi\right)^5}{\left(A_0-2\xi\right)^2\left(B_0-\xi\right)\left(Y_0-\left(3-\boldsymbol{f}\right)\xi\right)^{(3-f)}} \quad (20)$$

$$G=2\mu_A\left(A_0-2\xi\right)+\mu_B\left(B_0-\xi\right)+\left(3-\boldsymbol{f}\right)\mu_Y\left(Y_0-\left(3-\boldsymbol{f}\right)\xi\right)+5\mu_P\left(P_0+5\xi\right) \quad (21)$$

Thermodynamics: reaction steps O1 to O5

For the Oregonator reaction steps, we assign arbitrary values to the standard chemical potentials of the variables A and P :

$$\mu_A^0=-5\times10^4\frac{J}{mol},\ \mu_P^0=-10^5\ J/mol \quad (22)$$

The values of the standard chemical potentials for the variables B , X , Y , and Z are assigned according to Equations O1-O5 and (17) to guarantee thermodynamic consistency in simulations and calculations:

$$\mu_X^0=\frac{\mu_A^0+\mu_P^0+RTln\left(k_4/k_{-4}\right)}{2} \quad (23)$$

$$\mu_Y^0=\mu_X^0+\mu_P^0-\mu_A^0+RTln\left(k_1/k_{-1}\right) \quad (24)$$

$$\mu_Z^0=\mu_A^0-\mu_X^0-RTln\left(k_3/k_{-3}\right) \quad (25)$$

$$\mu_B^0=f\,\mu_Y^0-\mu_Z^0+RTln\left(k_5/k_{-5}\right) \quad (26)$$

The overall reaction's thermodynamic equations apply similarly to Equations O1-O5.

Results and discussion

Thermodynamics: reaction steps O1 to O5

Figure 1 shows the dynamics of the variables A , B , Y , and P of the overall reaction Og over time. Curves emerge from the integration of Equation (9) to the initial conditions indicated in the legend, subsequently replacing the result in $=A_0-2\xi$, $B=B_0-\xi$, $Y=Y_0-\left(3-f\right)\xi$, $P=P_0+5\xi$. To better appreciate the result of the integration, we use an equilibrium constant smaller than the Oregonator, with a value of 10^7 . This does not affect the evolution of the Og reaction towards equilibrium. As expected, the species that act as reactants, A , B , Y , decrease until they reach equilibrium, while the product P increases.

From the results of Equation (9), and by using Equation (19), we compute the thermodynamic quantities describing the evolution of the overall reaction Og towards the equilibrium state.

Figure 2 shows, also based on Equation (6), the thermodynamic affinity as a function of the reaction extent for different initial conditions of variable B (*i.e.*, the organic substrate). Note that the thermodynamic affinity decreases continuously and smoothly until it reaches zero at equilibrium, which satisfies the conditions in Equation (7). It is also important to note that the initial values for B were selected to be the limiting quantity according to the

stoichiometry of the overall reaction Og. This is why the reaction proceeds until the reaction extent is very close to the value of the limit quantity B.

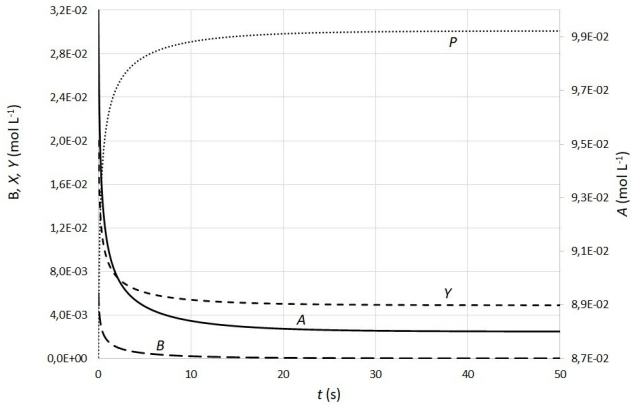


Figure 1. Dynamics of the overall reaction Og of the Oregonator over time. Curves in the figure are identified with the corresponding variable. The initial values for the variables are $B_0 = 3\times10^{-3}$, $A_0 = 10^{-1}$, $Y_0 = 10^{-2}$ and $P_0 = 10^{-4}$. The values for parameters f and k_s/k_{-s} are 0.48 and 10^7, respectively.

Source: Authors

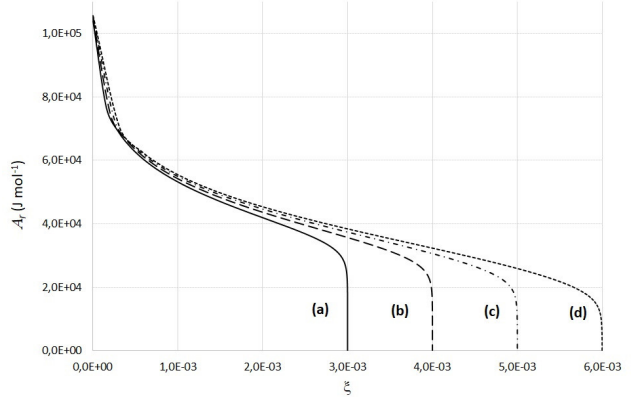


Figure 2. Thermodynamic affinity as a function of the reaction extent of the overall reaction Og. Curves in the figure correspond to a different initial value for the variable B, as follows: (a) 3×10^{-3}, (b) 4×10^{-3}, (c) 5×10^{-3}, (d) 6×10^{-3}. Other initial values are $A_0 = 10^{-1}$, $Y_0 = 10^{-2}$, and $P_0 = 10^{-4}$. The values for parameters f and k_s/k_{-s} are 0.48 and 10^7, respectively.

Source: Authors

From the reaction extent obtained using Equation (9), we calculate the reaction quotient of the overall reaction Og, defined in Equation (20), and the Gibbs function of the system, described in Equation (21). Figure 3 shows the Gibbs function for different initial conditions of variable B. Note that the Gibbs function decreases as a function of the reaction quotient, which changes until it becomes equal to the value of the equilibrium constant. Thus, the Gibbs function decreases continuously and smoothly until the reaction reaches chemical equilibrium, which is univocally defined when the reaction quotient equals the equilibrium constant.

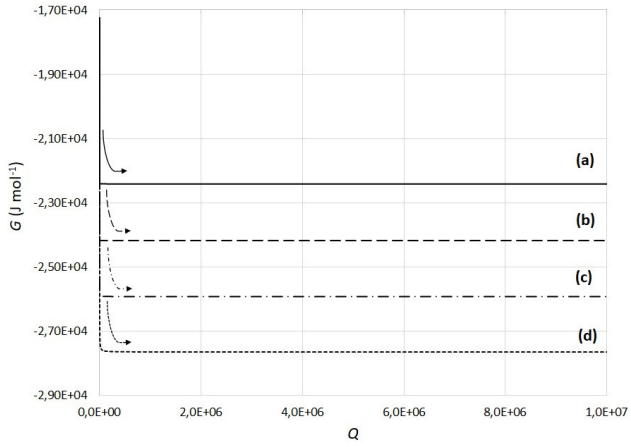


Figure 3. Gibbs function as a function of the reaction quotient of the overall reaction Og. Curves in the figure correspond to a different initial value for the variable B, as follows: (a) 3×10^{-3}, (b) 4×10^{-3}, (c) 5×10^{-3}, (d) 6×10^{-3}. Other initial conditions are $A_0 = 10^{-1}$, $Y_0 = 10^{-2}$, and $P_0 = 10^{-4}$. The values for parameters f and k_s/k_{-s} are 0.48 and 10^7, respectively.

Source: Authors

Finally, from the reaction extent and the reaction quotient, we calculate the change in the chemical reaction Gibbs function, $\Delta_r G$, defined in Equation (16). Figure 4 shows that $\Delta_r G$ increases continuously and smoothly until it becomes equal to zero when the reaction quotient is equal to the equilibrium constant. Thus, from Figures 2, 3, and 4, we verify that the global reaction Og satisfies the requirements of classical thermodynamics in chemical equilibrium, *i.e.*, the thermodynamic affinity as a function of advance is zero, the Gibbs function of the system has a minimum as a function of the reaction quotient, the Gibbs function of the reaction is zero as a function of the reaction quotient, and the equilibrium constant uniquely defines the chemical equilibrium.

The formalism and methodology presented in this section are commonly applied in studying chemical processes. In modeling the homogeneous gas-phase oxidation of sulfur dioxide and titanium (IV) chloride, thermodynamic affinity and the Gibbs function describe the evolution towards chemical equilibrium, approaching zero as a function of the overall reaction extent (Koukkari *et al.*, 2018; Koukkari and Pajarre, 2021), as shown in Figures 2 and 4. A similar approach applies to characterizing the equilibrium of thermal biomass conversion and methanation processes (Kangas, 2015; Kangas *et al.*, 2017).

There are also many applications of thermodynamic formalism devoted to studying the optimization of overall chemical processes in batch reactors, like optimizing the reaction extent and entropy production in sulfuric acid decomposition (Wang *et al.*, 2016). There is also some criticism of the formalism of thermodynamics also. A modified Gibbs function seems necessary to correct inaccuracies in the prediction of equilibrium models, positing that heat exchange between the system and the surroundings is not reversible. As a result, the modified Gibbs function corrects

the prediction of the chemical equilibrium and entropy generation of methane steam reforming overall reaction as a function of the reaction extent (Haseli, 2019). A similar approach applies to optimizing ammonia decomposition in hydrogen production (Ojelade and Zaman, 2021). It is important to note that little attention has been paid to including the reaction quotient in the thermodynamic study of chemical processes.

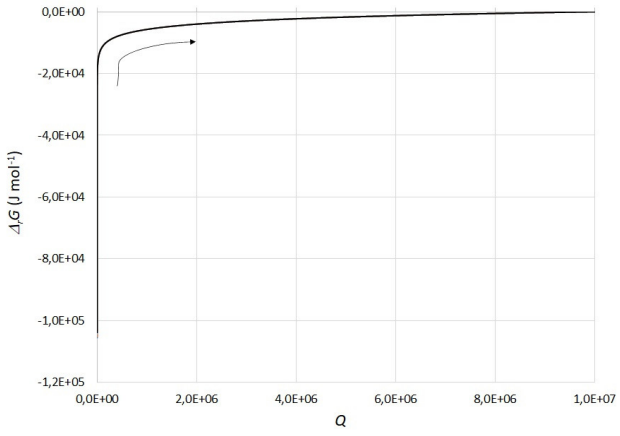


Figure 4. Change in the Gibbs function of the reaction $\Delta_r G$ as a function of the reaction quotient. In the calculations, the following initial conditions are used: $B_0 = 3\times10^{-3}$, $A_0 = 10^{-1}$, $Y_0 = 10^{-2}$, $P_0 = 10^{-4}$. The values for parameters f and k_1/k_{-1} are 0.48 and 10^7, respectively.
Source: Authors

The Oregonator: reactions O1 to O5

By integrating the set of Equations (10)-(14), the reaction extent of each step O1-O5 is obtained, and, from these, the change over time of each of the model variables is calculated: $A = A_0 - \xi_1 - \xi_3 + \xi_4$, $B = B_0 - \xi_5$, $X = X_0 + \xi_1 - \xi_2 + \xi_3 - \xi 2_4$, $Y = Y_0 - \xi_1 - \xi_2 + f\xi_5$, $Z = Z_0 + \xi_3 - \xi_5$, $P = P_0 + \xi_1 + 2\xi_2 + \xi_4$.

This study does not analyze the stability of fixed points in the Oregonator model. Instead, Figure 5 demonstrates that the model displays an oscillatory behavior for the initial variable values provided. Variable B serves as a control parameter for the model's dynamics, decreasing continuously and stepwise until it reaches a constant value, which is consistent with the oscillatory dynamics of the model. The net consumption rate of B depends on which set of reaction steps has a kinetic control on the dynamics: steps O1 to O4 or step O5. The curves in Figure 5 show that, within the constraints established in this study, the model exhibits transient oscillations that eventually disappear, after which the reaction proceeds monotonically. The continuous depletion of B during the reaction process satisfies the second law of thermodynamics, as the primary reactants remain constant, providing the free energy that drives the reaction towards equilibrium.

The variables X, Y, and Z in the Oregonator model exhibit oscillatory dynamics corresponding to intermediate species in the reaction mechanism. X drives the chemical instability in the reaction mechanism, and Figure 6 shows its oscillatory dynamics. The curves (a) and (b) demonstrate that the oscillations in X dampen and eventually vanish, exhibiting monotonic dynamics. These results agree with the well-known experimental evidence of the BZ reaction, which is explained with the FKN mechanism and modeled semi-quantitatively with the Oregonator (Field *et al.*, 1972; Field and Noyes, 1974). Curve (b) highlights the effect of the initial value of B as a control parameter of the reaction system's evolution. The Oregonator transitions from a cycle of damped oscillations with higher amplitude to one with lower amplitude and frequency. This finding is crucial because it indicates that, if the model is not studied for an extended period of time, the completion of an oscillatory cycle could be mistaken for the approach to equilibrium. However, for the constraints imposed in this work, all variables undergo sustained changes, implying that the path to chemical equilibrium is asymptotic and far from coinciding with the end of the oscillatory dynamics.

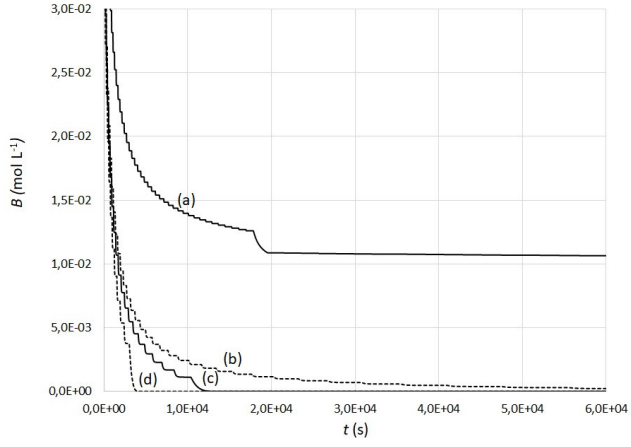


Figure 5. Dynamics of the variable B, the organic substrate in the Oregonator. Curves correspond to a different initial value for the variable B, as follows: $(a)3.8\times10^{-2}$, $(b)4.0\times10^{-2}$, $(c)4.1\times10^{-2}$, $(d)5.1\times10^{-2}$. Other initial conditions are $A_0 = 3.0\times10^{-2}$, $Y_0 = 10^{-6}$, $P_0 = 10^{-6}$, $Z_0 = 10^{-3}$, and $f = 0.48$.
Source: Authors

The Oregonator model displays oscillatory dynamics in the variables X, Y, and Z, corresponding to intermediate species in the reaction mechanism. Since each of the Oregonator steps O1-O5 has a reaction extent according to the system of Equations (10)-(14), it is not possible to define the overall extent of the reaction as a function of the model dynamics. To estimate the extent of the **Og** reaction, we calculated the dynamics of variables A, B, Y, and P, as shown in Figures 5 and 6. The large temporal dynamic of the Oregonator indicates that the extent of the Og reaction calculated on this time scale is also significant and cannot be related to the initial values of the variables.

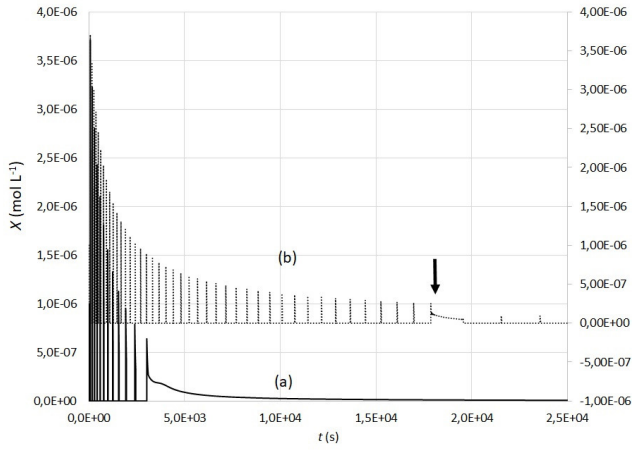


Figure 6. Dynamics of the variable X, the autocatalytic chemical species of the Oregonator. Curves correspond to a different initial value for the variable B, as follows: $(\mathbf{a})3.8\times10^{-2}$,$(\mathbf{b})4,0\times10^{-2}$. Other initial conditions are $A_0 = 3,0\times10^{-2}$, $Y_0 = 10^{-6}$, $P_0 = 10^{-6}$, $Z_0 = 10^{-3}$, and $f = 0.48$. The values for curve (b) are on the secondary axis.
Source: Authors

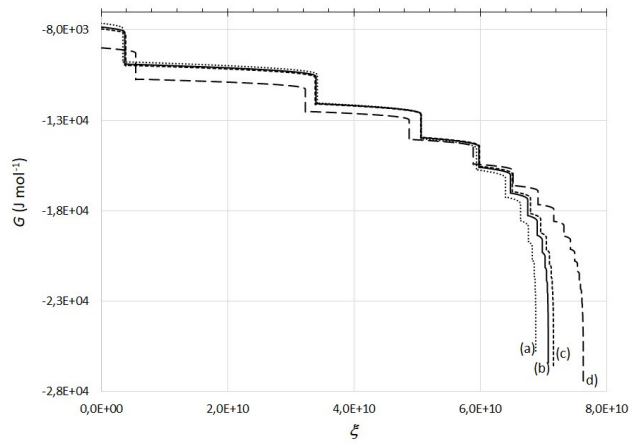


Figure 7. Change in the Gibbs function of the Oregonator as a function of the reaction extent of the overall reaction Og. Curves correspond to a different initial value for the variable B, as follows: $(\boldsymbol{a})3.8\times10^{-2}$,$(\boldsymbol{b})4.0\times10^{-2}$,$(\boldsymbol{c})4.1\times10^{-2}$,$(\boldsymbol{d})5.1\times10^{-2}$. Other initial conditions are $A_0 = 3.0\times10^{-2}$, $Y_0 = 10^{-6}$, $P_0 = 10^{-6}$, $Z_0 = 10^{-3}$, and $f = 0,48$.
Source: Authors

We analyzed the Gibbs function of the Oregonator from the sum of the Gibbs functions of each step O1-O5. Figure 7 shows the variation in the Gibbs function for different initial conditions of variable B as a function of the extent calculated for the overall reaction Og. This result is interesting because it shows how the Gibbs function decreases continuously but not smoothly, describing the oscillatory dynamics of the model. Furthermore, the Gibbs function describes different trajectories in the evolution to equilibrium, depending on the initial conditions. Figure 8 shows how the Gibbs function of the Oregonator reaction, $\Delta_r G$, varies using quantities calculated similarly to that indicated in Figure 7. We obtained $\Delta_r G$ by summing the quantities calculated for each step of the model and calculating the reaction quotient from the overall extent of the reaction. The behavior of $\Delta_r G$ as a function of the reaction quotient in Figure 8 is like that observed in Figure 4, despite the oscillatory dynamics of the Oregonator. However, even after integrating the model for a long time, $\Delta_r G$ does not approach zero, indicating that the system is not at chemical equilibrium. This behavior could be explained by the fact that the estimated reaction quotient has a magnitude that is far from the value of the global equilibrium constant: 4.8×10^{49}.

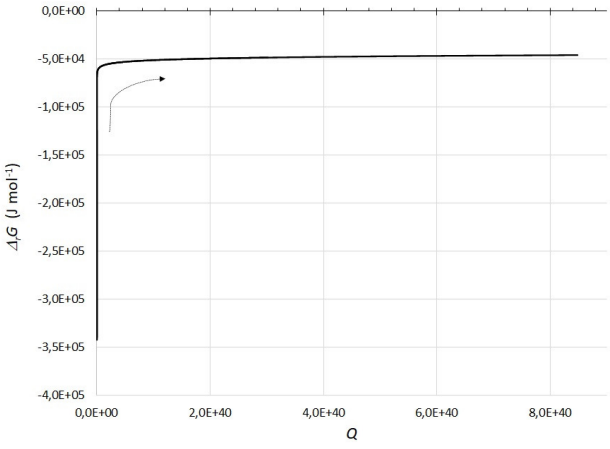


Figure 8. Change in the Gibbs function of Oregonator as a function of the reaction quotient of the overall reaction Og. The initial conditions for the variables and parameters are $\boldsymbol{B} = 4.0\times10^{-2}$, $A_0 = 3.0\times10^{-2}$, $Y_0 = 10^{-6}$, $P_0 = 10^{-6}$, $Z_0 = 10^{-3}$, and $f = 0.48$.
Source: Authors

Finally, we can calculate the entropy production rate of the Oregonator from the results obtained in Figures 7 and 8 by using Equation (8), as shown in Figure 9, where the curves for different initial conditions of variable *B* demonstrate good agreement between the thermodynamic dissipation of the Oregonator and the reaction quotient. The thermodynamic dissipation decreases continuously and smoothly, approaching zero and indicating equilibrium. However, as in Figure 8, the reaction quotient is not close to the expected value for the equilibrium constant.

These results highlight the challenge of applying classical thermodynamics fundamentals to complex reaction systems. While all the thermodynamic fundamentals hold, they cannot be generalized for each step of a reaction system or extended to analyze the overall reaction. As shown in Figures 5 and 6, studying complex processes is challenging from kinetic or energetic approaches, and different methodologies are considered to rationalize the process description (de Oliveira *et al.*, 2016). The traditional method used in this section describes physical, electrochemical, catalytic, and biochemical multistep processes evolving in time (Heimburg, 2021). However, little effort has been made to deepen the thermodynamic study of these systems, since most research focuses on kinetic and dynamic description

and analysis (Marin *et al*., 2019). In the thermodynamics of irreversible processes, entropy production rate is a valuable quantity to optimize the energetic performance of complex processes (Heimburg, 2017; Nieto-Villar, 2020; Arango-Restrepo *et al*., 2020). Given the formalism used herein, some applications of the entropy production rate show the utility of this quantity to describe multistep chemical systems (Dutt, 1985; Barragán *et al*., 2015; Barragán and Montoya, 2021).

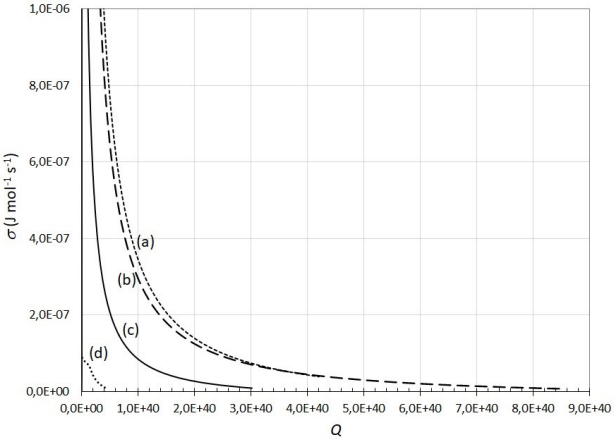


Figure 9. Changes in the entropy production rate as a function of the reaction quotient of the overall reaction Og. Curves correspond to a different initial value for the variable B, as follows: $(a)3.8\times10^{-2}$, $(b)4.0\times10^{-2}$, $(c)4.1\times10^{-2}$, $(d)5.1\times10^{-2}$. Other initial conditions are $A_0=3.0\times10^{-2}$, $Y_0=10^{-6}$, $P_0=10^{-6}$, $Z_0=10^{-3}$, and $f=0.48$.
Source: Authors

Conclusions

This study demonstrated that classical thermodynamics principles can be applied to single-step reactions with non-stoichiometric coefficients, as exemplified by the overall reaction Og of the Oregonator model. The thermodynamic affinity decreases continuously and smoothly for a single-step chemical reaction until reaching zero at equilibrium. In contrast, the Gibbs function decreases until the reaction quotient equals the equilibrium constant for a single-step reaction.

However, for complex coupled reaction systems like the Oregonator, the thermodynamic formalism could not be verified. The lack of knowledge of the reaction extent and the reaction quotient hinders the evaluation of thermodynamic functions and the verification of the state of chemical equilibrium. Non-stoichiometric coefficients in these models, such as parameter f, complicate the determination of the reaction extent because the system's dynamics can remain in non-equilibrium states, distorting thermodynamic quantities. In this context, the rate of entropy production is a valuable quantity. Further research must define and express the reaction extent for coupled chemical reactions.

Acknowledgements

The authors thank Universidad Nacional Colombia for supporting the research project no. 41121, titled *Physicochemical evaluation of the control exercised by the molecular structure of the organic substrate in the global dynamics of the Belousov-Zhabotinsky oscillating reaction* to support the development of postgraduate theses (2017-2018).

Author contributions

Jean P. Montoya carried out part of the programming of the models and participated in conceptualizing the problem and the discussions. Alexander Contreras-Payares participated in conceptualizing the problem and the discussions. Daniel Barragán was responsible for conceptualizing the problem, modeling, and writing the first version of the manuscript. All authors participated in discussing the results and revising this document.

Conflicts of interest

The authors of this paper declare that they have no conflicts of interest.

CRediT author statement

All authors: conceptualization, methodology, software, formal analysis, investigation, data curation, writing (original draft, review, and editing).

References

Arango-Restrepo, A., Barragán, D., and Rubi, J. M. (2020). Modelling non-equilibrium self-assembly from dissipation. *Molecular Physics*, *118*(9-10), e1761036. https://doi.org/10.1080/00268976.2020.1761036

Banerjee, A. C. (1995). *Teaching chemical equilibrium and thermodynamics in undergraduate general chemistry classes. Journal of Chemical Education, 72*(*10*), *879*. https://doi.org/10.1021/ed072p879

Barragán, D., Ágreda, J., and Parra, W. (2015). Entropy production in the Oregonator model perturbed in a calorimeter with a chemical pulse. *Journal of Thermal Analysis and Calorimetry*, *119*, 705-713. https://doi.org/10.1007/s10973-014-4129-0

Barragán, D., and Montoya, J. P. (2021). Producción de entropía en un modelo autocatalítico no-isotérmico perturbado con pulsos de analito. *Revista de la Facultad de Ciencias*, *10*(2), 9-27. https://doi.org/10.15446/rev.fac.cienc.v10n2.83839

Beretta, G. P., and Gyftopoulos, E. P. (2015). *What is a chemical equilibrium state. Journal of Energy Resources Technology, 137*(*2*), *4*. https://dx.doi.org/10.1115/1.4026384

de Oliveira, L. P., Hudebine, D., Guillaume, D., and Verstraete, J. J. (2016). A review of kinetic modeling methodologies for complex processes. *Oil & Gas Science and Techno-*

logy–Revue d'IFP energies nouvelles, *71*(3), 45. https://doi.org/10.2516/ogst/2016011

Dutt, A. K. (1985). Instabilities and oscillations in ''The reversible oregonator model'': A thermodynamic approach to calculate the excess entropy production (δ m P) and the corresponding antisymmetric function (δ m Π). *The Journal of Chemical Physics*, *82*(10), 4678-4682. http://dx.doi.org/10.1063/1.448972

Epstein, I. R., and Pojman, J. A. (1998). *An introduction to nonlinear chemical dynamics: oscillations, waves, patterns, and chaos*, Oxford University Press.

Eu, B. C., and Al-Ghoul, M. (2018). *Chemical thermodynamics: Reversible and irreversible thermodynamics*. World Scientific.

Field, R. J., Koros, E., and Noyes, R. M. (1972). Oscillations in chemical systems. II. Thorough analysis of temporal oscillation in the bromate-cerium-malonic acid system. *Journal of the American Chemical Society*, *94*(25), 8649-8664. https://doi.org/10.1021/ja00780a001

Field, R. J., and Noyes, R. M. (1974). Oscillations in chemical systems. IV. Limit cycle behavior in a model of a real chemical reaction. *The Journal of Chemical Physics, 60*(5), 1877-1884. https://doi.org/10.1063/1.1681288

Haseli, Y. (2019). Criteria for chemical equilibrium with application to methane steam reforming. *International Journal of Hydrogen Energy*, *44*(12), 5766-5772. https://doi.org/10.1016/j.ijhydene.2019.01.130

Heimburg, T. (2017). Linear nonequilibrium thermodynamics of reversible periodic processes and chemical oscillations. *Physical Chemistry Chemical Physics*, *19*(26), 17331-17341. https://doi.org/10.1039/C7CP02189E

Heimburg, T. (2021). Damped physical oscillators, temperature and chemical clocks. *arXiv preprint*. https://doi.org/10.48550/arXiv.2112.03083

Honig, J. M. (2020). *Thermodynamics: Principles characterizing physical and chemical processes.* Academic Press.

Kangas, P. (2015). *Modelling the super-equilibria in thermal biomass conversion: Applications and limitations of the constrained free energy method* [Technical Report]. VTT Technical Research Centre of Finland LTD. https://publications.vtt.fi/pdf/science/2015/S92.pdf

Kangas, P., Vázquez, F. V., Savolainen, J., Pajarre, R., and Koukkari, P. (2017). Thermodynamic modelling of the methanation process with affinity constraints. *Fuel*, *197*, 217-225. https://doi.org/10.1016/j.fuel.2017.02.029

Kiprijanov, K. S. (2016). Chaos and beauty in a beaker: The early history of the Belousov-Zhabotinsky reaction. *Annalen der Physik, 528*(3-4), 233-237. https://doi.org/10.1002/andp.201600025

Kondepudi, D., and Prigogine, I. (2014). *Modern thermodynamics: From heat engines to dissipative structures*, John Wiley & Sons.

Koukkari, P., Pajarre, R., and Kangas, P. (2018). Thermodynamic affinity in constrained free-energy systems. *Monatshefte für Chemie-Chemical Monthly*, *149*, 381-394. https://doi.org/10.1007/s00706-017-2095-5

Koukkari, P., and Pajarre, R. (2021). Phase diagrams with the driving force and extent of reaction as axis variables. *Calphad*, *74*, 102290. https://doi.org/10.1016/j.calphad.2021.102290

Marin, G. B., Yablonsky, G. S., and Constales, D. (2019). *Kinetics of chemical reactions: decoding complexity*. John Wiley & Sons.

Martínez-Grau, M., Solaz-Portolés, J. J., and Sanjosé, V. (2014). *Propuesta de un test para determinar el conocimiento conceptual de estudiantes universitarios sobre la constante de equilibrio químico y su aplicación en estudiantes españoles. Química Nova, 37,* 740-744. http://dx.doi.org/10.5935/0100-4042.20140118

Nieto-Villar, J. M. (2020). Una mirada a los sistemas complejos desde la termodinámica. *Suplemento de la Revista Mexicana de Física*, *1*(4), 17-24. https://doi.org/10.31349/SuplRevMexFis.1.4.17

Ojelade, O. A., and Zaman, S. F. (2021). Ammonia decomposition for hydrogen production: a thermodynamic study. *Chemical Papers*, *75*, 57-65. https://doi.org/10.1007/s11696-020-01278-z

Patiño-Sierra, D. F., and Barragán, D. (2022). Un método formal para la armonización conceptual del equilibrio químico. *Revista de la Facultad de Ciencias*, *11*(2), 148-161. https://doi.org/10.15446/rev.fac.cienc.v11n2.99977

Prausnitz, J. M., Lichtenthaler, R. N., and de Azevedo, E. G. (1998). *Molecular thermodynamics of fluid-phase equilibria*. Pearson Education.

Prigogine, I. (1961). *Introduction to irreversible thermodynamics*. Interscience.

Quílez-Pardo, J., and Solaz-Portolés, J. J. (1995). Students' and teachers' misapplication of Le Chatelier's principle: Implications for the teaching of chemical equilibrium. *Journal of Research in Science teaching, 32*(9), 939-957. https://doi.org/10.1002/tea.3660320906

Rogers, F., Huddle, P. A., and White, M. W. (2000). The complexity of teaching and learning chemical equilibrium. *Journal of Chemical Education, 76*(4), 554. https://doi.org/10.1021/ed076p554

Rosenberg, R. M., and Klotz, I. M. (2008). *Chemical thermodynamics: basic concepts and methods*. Wiley.

Saunders, N., and Miodownik, A. P. (Eds.) (1998). *CALPHAD (calculation of phase diagrams): A comprehensive guide*. Elsevier.

Solaz-Portolés, J. J. (2011). Variation of extent of reaction in closed chemical equilibrium when changing the temperature at constant volume. *Química Nova*, *34*, 710-713. https://doi.org/10.1590/S0100-40422011000400028

Tosun, I. (2021). *The thermodynamics of phase and reaction equilibria*. Elsevier.

Tyson, J. J. (1981). On scaling the oregonator equations. In C. Vidal and Pacault (Eds.), *Nonlinear Phenomena in Chemical Dynamics* (pp. 222-227). Springer. https://doi.org/10.1007/978-3-642-81778-6_33

Tyson, J. J. (1982). Scaling and reducing the Field-Koros-Noyes mechanism of the Belousov-Zhabotinskii reaction. *The Journal of Physical Chemistry, 86*(15), 3006-3012. https://doi.org/10.1021/j100212a039

Wang, C., Chen, L., Xia, S., and Sun, F. (2016). Maximum production rate optimization for sulphuric acid decomposition process in tubular plug-flow reactor. *Energy*, *99*, 152-158. https://doi.org/10.1016/j.energy.2016.01.040

Research Article/Civil and Sanitary Engineering

http://doi.org/10.15446/ing.investig.104078

Numerical Assessment of Bidirectional Roller Bearing Isolators under Near-Fault Earthquakes

Evaluación numérica de aisladores de soportes rodantes bidireccionales sometidos a sismos cercanos a la falla

Nelson Andrés Ortiz-Cano[1], Ricardo González-Olaya[2], Carlos Andrés Gaviria-Mendoza[3], Carlos Magluta[4], and Ney Roitman[5]

ABSTRACT

Base isolation with roller bearing systems has been widely studied in recent years due to its successful performance in the seismic protection of buildings and bridges. This paper numerically evaluates the effectiveness of a bi-directional roller bearing (RB) seismic isolation system composed of sloped bearing plates and multiple rollers arranged in both orthogonal-in-plane directions. Previous experimental results obtained with unidirectional and bidirectional RBs were used to validate the 3D numerical model of an isolated building with RBs. The model was used to obtain the nonlinear response of a four-story multi-column building when subjected to pairs of scaled near-fault earthquake records. The effects of the bearing plate inclination angle (ranging from 1.0 to 4.0°), the sliding friction force, and supplementary dissipation mechanisms (0.0-0.5 N/kg, *i.e.* friction force normalized with the structure mass) were evaluated. The results show that the proposed bidirectional RB system is suitable for reducing the seismic response of stiff and flexible multi-column structures. In particular, the RB system reduces the acceleration responses by 5-85% in flexible structures and by 86-96% in stiff ones. Furthermore, bearing plates with an inclination angle greater than or equal to 3.0° have significant benefits in terms of self-centering capacity.

Keywords: base isolation, nonlinear seismic response, supplementary dissipation mechanism, inclination angle

RESUMEN

El aislamiento de base con sistemas de rodamiento ha sido ampliamente estudiado en los últimos años debido a su exitoso desempeño en la protección sísmica de puentes y edificaciones. Este artículo evalúa numéricamente la efectividad de un sistema de aislamiento sísmico bidireccional de soportes rodantes (RBs) compuesto por superficies de rodamiento inclinadas y múltiples rodillos dispuestos en ambas direcciones ortogonales en el plano. Se utilizaron resultados obtenidos con RBs unidireccionales y bidireccionales para validar el modelo numérico 3D de un edificio aislado con RBs. Se utilizó el modelo para obtener la respuesta no lineal de un edificio multicolumna de cuatro pisos frente a pares de registros escalados de terremotos cercanos a la falla. Se evaluaron los efectos del ángulo de inclinación de la superficie de rodamiento (en el rango de 1.0 a 4.0°), de la fuerza de fricción por rodamiento y de mecanismos suplementarios de disipación (0.0-0.5 N/kg, *i.e.*, fuerza de fricción normalizada con la masa de la estructura). Los resultados muestran que el sistema de RBs bidireccionales propuesto es adecuado para reducir la respuesta sísmica de estructuras multicolumna rígidas y flexibles. En particular, el sistema de RBs reduce las respuestas de aceleración en 5-85 % en estructuras flexibles y en 86-96 % en estructuras rígidas. Además, las superficies de rodamiento con un ángulo de inclinación mayor o igual a 3.0° muestran beneficios significativos respecto a su capacidad de autocentrado.

Palabras clave: aislamiento de base, respuesta sísmica no lineal, mecanismo de disipación suplementario, ángulo de inclinación

Received: August 3rd 2022
Accepted: September 27th 2023

Introduction

Isolation systems have been successfully used since 1969 in 12 720 projects worldwide (Walters, 2015). Although the earliest applications of these seismic protection mechanisms were related to critical buildings such as hospitals and emergency facilities, they have recently been extended to ensure a better seismic response in lower and mid-rise residential building projects (Wang *et al.*, 2017).

This technology has been implemented in new construction projects, updating structures to current standards, and in the retrofitting of structures with insufficient earthquake (EQ) resistance (Naeim and Kelly, 1999; Tsai, Wu, Chang, and Lee, 2007; Matsagar and Jangid, 2008; Hosseini and Soroor, 2011, 2013; Erdik, Ulker, Şadan, and Tuzun, 2018; Ryan, Okazaki, Coria, Sato, and Sasaki, 2018). Nevertheless, the widespread application of these devices is a current challenge that requires improving existing methods and developing cost-effective alternative devices and other

[1]Civil engineer, Universidad Nacional de Colombia, Medellín Campus, Colombia. PhD in Civil Engineering, Universidade Federal do Rio de Janeiro, Brazil. Affiliation: Auxiliary professsor, Universidad Militar Nueva Granada, Colombia. E-mail: nelson.ortiz@unimilitar.edu.co
[2]Civil engineer, Universidad Militar Nueva Granada, Colombia. MSc in Civil Engineering, Universidad Militar Nueva Granada, Colombia. Affiliation: Young researcher, Universidad Militar Nueva Granada, Colombia. E-mail: ricardo.gonzalezo@unimilitar.edu.co
[3]Civil engineer, Universidad del Valle, Colombia. PhD in Civil Engineering, Universidad de Puerto Rico, Puerto Rico. Affiliation: Associate professsor, Universidad Militar Nueva Granada, Colombia. E-mail: carlos.gaviria@unimilitar.edu.co
[4]Civil engineer, Universidade Federal do Rio de Janeiro, Brazil. PhD in Civil Engineering, Universidade Federal do Rio de Janeiro, Brazil. Affiliation: Full professsor, Universidade Federal do Rio de Janeiro, Brazil. E-mail: magluta@coc.ufrj.br
[5]Civil engineer, Pontifícia Universidade Católica do Rio de Janeiro, Brazil. PhD in Civil Engineering, Universidade Federal do Rio de Janeiro, Brazil. Affiliation: Full professsor, Universidade Federal do Rio de Janeiro, Brazil. E-mail: roitman@coc.ufrj.br

isolation strategies (Bagerzadeh-Karimi and Geneş, 2019; Calhoun, Tehrani, and Harvey-Jr, 2019; Beirami-Shahabi, Zamani-Ahari, and Barghian, 2020; Zhang and Ali, 2021).

Seismic isolation systems are the most commonly used devices in earthquake-resistant structures. Natural rubber bearings (NRB), lead rubber bearings (LRB), high-damping rubber bearings (HDRB), slider bearings, and mono-, double-, and triple-friction pendulum (FP) systems stand out for their design simplicity and effectiveness. Some isolated structure projects and the development of new materials were described by (De Luca and Guidi, 2019) and (Nobari Azar, Karimzadeh Naghshineh, and Sen, 2022). An isolation seismic layer must provide (i) low lateral stiffness, (ii) re-centering capacity, and (iii) supplementary damping. Recently, roller seismic isolation bearings (RSIB) have gained attention, as (Lee, Liang, and Snyder, 2005) reported and showcased the advantages of their patent (WO 2005/031088 A2).

Several studies have shown that roller bearing (RB) isolators exhibit excessive displacements under near-fault ground records (Rawat, Ummer, and Matsagar, 2018), in addition to stress concentration and lower friction resistance (Beirami-Shahabi *et al.*, 2020). Moreover, an elementary RB isolator device does not offer integrated displacement control (Zhang and Ali, 2021). These issues can be mitigated via energy dissipation devices (Ortiz-Cano, Magluta, and Roitman, 2015), *e.g.*, by implementing shock absorbers with a particular bumper and gap configuration (Andreaus and De Angelis, 2020) or by introducing traditional springs (Zhang and Ali, 2021).

Numerous attempts have been made to implement RB devices with a low friction force in order to cover a wide range of earthquake intensities, but they have resulted in unnecessarily high isolator displacements (Rawat *et al.*, 2018; Rawat and Matsagar, 2021). Period-matching effects (tuned to the natural period of the building) may arise in RB systems, even when located on lower floors (Harvey-Jr and Gavin, 2015). Furthermore, ground motions with long-period components that reach the natural period of the isolator device may weaken its performance (Calhoun, 2018; P. Chen, Hsu, Zhong, and Wang, 2021). In this sense, the robustness of RB isolators could be enhanced through a tailored design based on seismic demand characteristics, such as ground motion period and intensity (Harvey-Jr and Gavin, 2015).

Sliding friction mechanisms are commonly integrated into RB systems to provide additional friction force while reducing the peak displacement (Lee, Ou, Niu, Song, and Liang, 2010). However, friction is deteriorated over time due to wear during recurrent loading cycles and weathering from external conditions (Lee *et al.*, 2010), in addition to the sliding velocity, surface temperature, and other physical phenomena (Lee *et al.*, 2010; Zhang and Ali, 2021). (Foti, Catalan Goni, and Vacca, 2013) proposed an alternative dissipation mechanism: a two-layer rubber cover attached to the lower and upper plates of RB devices, *i.e.*, a rubber-layer roller bearing (RLRB) isolation device for low-rise lightweight structures and equipment applications. This mechanism provides a higher damping capacity than typical steel-steel friction devices. However, given the potential resonance during seismic events with a predominance of low frequencies and the computational effort required to solve the contact problem, this type of mechanism has not been implemented in all building systems (Foti, 2019). In this vein, there is a research opportunity to develop reliable analytical and numerical models aimed at predicting the behavior of isolation systems (Beirami-Shahabi *et al.*, 2020).

This paper aims to study a bidirectional seismic isolation device composed of multiple rollers in the direction of the orthogonal plane. First, the model developed by (Ortiz-Cano *et al.*, 2015) is extended to couple the bidirectional nonlinear responses of the RB system when subjected to a pair of horizontal ground motion components. The numerical model is validated via several tests with a scaled building model, the RB system, and a combination of both. Then, the performance of low-rise and mid-rise buildings with the RB system is numerically evaluated against near-fault ground motions for several bearing plate slope angles and sliding friction forces. In the final section, some remarks and recommendations are provided.

Dynamic behavior of isolated buildings with RB systems

(Ortiz-Cano *et al.*, 2015) studied a sloped roller-type isolation device in which multiple rollers move between a V-shaped bearing plate and a flat surface in a single horizontal direction. This formulation was augmented to include two ground motion components. The dynamic behavior of isolated buildings with RBs under base excitations in a system with multiple degrees of freedom (MDoF) can be represented in the time domain via Equation (1).

$$M\ddot{u} + C\dot{u} + Ku + R\left(f_s + f_{dr} + f_{ds}\right) = -M\Gamma\ddot{u}_g \tag{1}$$

where M, C, and K are the mass, damping, and stiffness matrices of the MDoF system, respectively. In addition, $\ddot{u}$, $\dot{u}$, and u are the acceleration, velocity, and displacement vectors; R is a matrix that allocates the forces of restoration (f_s), rolling friction (f_{dr}), and sliding friction (f_{ds}) of the RB system in the j direction, into a matrix containing the forces of the MDoF system in the i direction. The elements that make up the R vector take values of 0 or 1 as follows: $R_{ij} = 1$ if, in DoF i, the forces f_s, f_{dr}, and f_{ds} are in the j direction; otherwise, $R_{ij} = 0$.

The forces f_s, f_{dr}, and f_{ds}, which characterize the RB system along one horizontal motion path, were extended for two orthogonal directions. These forces are defined in tensor notation through Equations (2), (3), and (4).

$$f_s{}^{\mathrm{T}} = \{\ \tfrac{1}{2}m_1 g\sin(\theta_1)\, f_{H_1}(u) \quad \tfrac{1}{2}m_2 g\sin(\theta_2)\, f_{H_2}(u)\ \} \tag{2}$$

$$f_{dr}{}^{\mathrm{T}} = \{\ \mu_{r_1} m_1 g f_{H_1}(\dot{u}) \quad \mu_{r_2} m_2 g f_{H_2}(\dot{u})\ \} \tag{3}$$

$$f_{ds}{}^{\mathrm{T}} = \{\ \mu_{s_1} N_1 f_{H_1}(\dot{u}) \quad \mu_{s_2} N_2 f_{H_2}(\dot{u})\ \} \tag{4}$$

where m is the mass supported by the bearing plates of the RB system; g is the gravitational acceleration; μ_r is the rolling friction coefficient; μ_s is the sliding friction coefficient; N is the normal force on the sliding interface; and θ is the angle

of the V-shaped surfaces. The subscripts 1 and 2 denote the direction of each quantity in the horizontal plane.

f_H is a suitable function that replaces the sign function used in earlier models of RB systems, and it allows to simulate a smoother transition between the V-shaped surfaces. In this work, the function proposed by (Ortiz-Cano, Magluta, and Roitman, 2014; Ortiz-Cano *et al.*, 2015) was used. The Ortiz-Cano function was written in terms of a parameter called *yield displacement* d and an auxiliary variable x associated with the displacement and speed of the system's base. This function is defined by

$$f_H(x) = \begin{cases} 1 & \text{if} \quad x \geq d \\ \frac{1}{d}x & \text{if} \quad -d < x < d \\ -1 & \text{if} \quad x \leq d \end{cases} \tag{5}$$

Finally, in Equation (1), $\ddot{u}_g$ is the seismic excitation vector containing the base accelerations for each direction in which the building can be excited, *i.e.*, three translational and three rotational DoF in the Cartesian space. The term Γ is an influence matrix that relates the excited DoF i with the direction of the seismic excitation j. The elements of the Γ matrix take values of 0 or 1 according to the following rule:

$$\Gamma_{ij} = \begin{cases} 0 & \text{if DoF } i \text{ is not excited in the } j \text{ direction} \\ 1 & \text{if DoF } i \text{ is excited in the } j \text{ direction} \end{cases} \tag{6}$$

To solve Equation (1), the *ode23t* numerical solver of ordinary differential equations (ODE) of MATLAB (The MathWorks Inc., 2019) was used, given its great efficiency and numerical stability when compared to other methods in the MATLAB's ODE pack. In general terms, the algorithm of this solver is an implementation of the trapezoidal rule with an adaptive step size associated with the use of a *free* interpolant (Shampine, Reichelt, and Kierzenka, 1999). The use of this solver requires that Equation (1), a second order equation, be rewritten as a first-order ODE system, as shown in the following expression:

$$\dot{v} = A_1 v(t) + A_2 H(u, \dot{u}) + A_3 \ddot{u}_g \tag{7}$$

with

$$A_1 = \begin{bmatrix} 0 & \mathsf{I} \\ -M^{-1}K & -M^{-1}C \end{bmatrix} \tag{8}$$

$$A_2 = \begin{bmatrix} 0 \\ -M^{-1}R \end{bmatrix} \tag{9}$$

$$A_3 = \begin{bmatrix} 0 \\ -\Gamma \end{bmatrix} \tag{10}$$

where, A_1 and $v(t)^{\mathrm{T}} (= \{u \quad \dot{u}\})$ denote the matrix of properties and the response vector of the system in a state-space form. The non-linearity of the system is represented by the $H(u, \dot{u})$ function, which includes the effects of the RB system.

Model validation

The results of an experimental evaluation and the numerical simulation of a building with and without RB isolation were compared in order to validate the numerical model. This section describes the characteristics of the building and the numerical model under study.

Building

The physical model of the frame building corresponds to the one studied by (Ortiz-Cano *et al.*, 2014, 2015). It consists of a four-story-one-bay frame structure supported by polymeric columns with a rectangular cross-section of 20.0×2.9 mm. The model has a total height of 803.4 mm and a floor area of 150×200 mm. The slabs and the story-column joints are constructed from aluminum plates. The base of the building has a thickness of 8,0 mm, and the stories and story-column joints have a thickness of 4.0 mm. In addition, the building has lumped additional masses of 0.50 kg on the second and third stories. A detailed outline of the experimental model's geometry is shown in Figure 1.

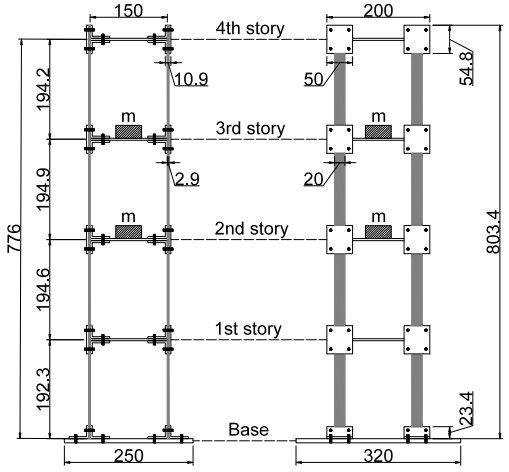


Figure 1. Physical model geometry. Measures in mm.
Source: Authors

According to the modal identification tests (*i.e.*, impulsive excitation on the first floor with a hammer) carried out by (Ortiz-Cano *et al.*, 2014, 2015) and the analysis of floor acceleration time series using the short-time Fourier transform (STFT) method, the modal parameters (mean values and standard deviations) identified for the fixed-base building model are summarized in Table 1.

Table 1. Identified dynamic properties of the building

Mode [Type-Direction]	f_{exp} [Hz]	ξ_{exp} [%]
1st flexural-weak	6.82±0.02	0.59±0.03
2nd flexural-weak	20.44±0.03	0.51±0.01
3rd torsional	21.55±0.02	2.88±0.03
4th flexural-weak	31.70±0.02	0.30±0.01

Source: Authors

In addition, Rayleigh damping constant values of $\alpha = 4.03 \cdot 10^{-1}$ and $\beta = 5.58 \cdot 10^{-5}$ were determined for the first two modes of vibration (which add up to a 95% modal participation of the total mass of the building), as the

Rayleigh damping model was used to assemble the damping matrix C within the numerical simulation algorithm.

RB system

We studied the seismic behavior of a bidirectional RB isolation system composed mainly of two surfaces and an array of rollers in an orthogonal arrangement. Previous works have shown that the use of these isolation systems enhances the dynamic response of structures subjected to seismic excitations (Ortiz-Cano *et al.*, 2014; Menga, Foti, and Carbone, 2017). However, this isolation scheme has some drawbacks involving the self-centering mechanism and roller bearing support stability (Sanchez-Torres, Rico-Caviedes, Ortiz-Cano, Nieto-Leal, and Gaviria-Mendoza, 2019).

Therefore, the bidirectional isolation system includes an intermediate V-shaped bearing plate inclined in both directions, as seen in Figure 2, which allows both attenuating the building's acceleration response and ensuring its return to the initial position (self-centering capacity) after a ground motion event. To avoid pounding between the rollers and the V-shaped intersection, a rolling arc zone with a fixed curvature radius higher than the roller radius is used, aiming to maintain a predominantly linear sloping surface (Wang *et al.*, 2014). In addition, energy dissipation occurs through lateral frictional plates.

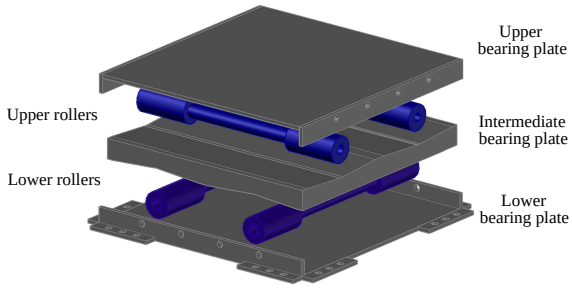


Figure 2. Bidirectional RB model
Source: Authors

Numerical model

3D frame elements were considered, which are characterized by the elasticity modulus E, the Poisson ratio ν that defines the shear modulus G, the density ρ, the area A, the polar moment of inertia J_x, and the moments of inertia I_y and I_z. The axis of the 3D frame elements in local coordinates coincides with the x direction in global coordinates according to the adopted coordinate system, which is implicitly shown in Figure 3. In the numerical representation of the physical model, three types of 3D frame elements were used. Table 2 lists their characteristics.

Table 2. Mechanical and geometric characteristics of the elements

Element type		1	2	3
E	[GPa]	28	70	70
ν	[-]	0.33	0.33	0.33
ρ	[kg/m^3]	1800	2700	0
A	[m^2]	$5.9 \cdot 10^{-5}$	$4.0 \cdot 10^{-4}$	$4.0 \cdot 10^{-4}$
J_x	[m^4]	$1.5 \cdot 10^{-10}$	$7.7 \cdot 10^{-9}$	$7.7 \cdot 10^{-9}$
I_y	[m^4]	$2.0 \cdot 10^{-9}$	$1.3 \cdot 10^{-7}$	$1.3 \cdot 10^{-7}$
I_z	[m^4]	$4.0 \cdot 10^{-11}$	$2.5 \cdot 10^{-9}$	$2.5 \cdot 10^{-9}$

Source: Authors

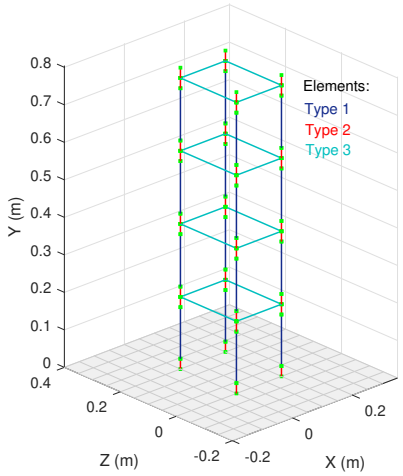


Figure 3. Model components
Source: Authors

Element type 1 represents the columns constructed in a polymeric material, while types 2 and 3 correspond to the aluminum connections and the slabs defined as beams in the numerical model (Figure 3). It is noteworthy that the density of component 3 takes the value of 0 kg/m^3 because the slab masses, as well as other factors that provide weight on each story, were modeled as masses concentrated at the nodes of each floor. Thus, concentrated masses of 539.3 , 196.6 , 321.2 , 321.1 , and 196.6 g were incorporated for the nodes of the base and the first, second, third, and fourth floors.

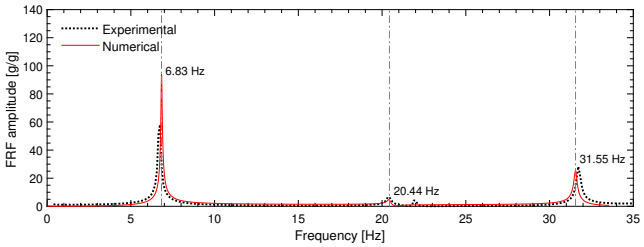
The area and the inertia values were established by geometric definition. For type 2 and 3 elements, the typical E and ν values for aluminum reported in the literature were assumed. For type 1 elements, an average ν value that is typical of polymeric materials was assumed, and E was tuned from the value used by (Ortiz-Cano, 2013) to obtain the best representation of the structure under fixed-base conditions.

Model assesment

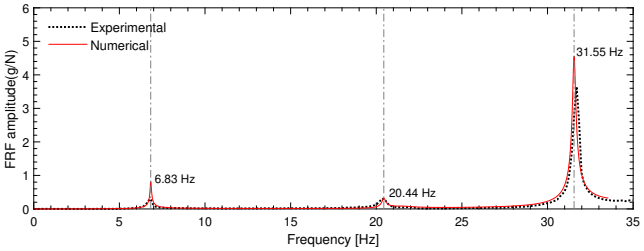
A comparative analysis of the numerical and experimental responses of the building, the RB system, and the building with the RB system is presented below.

The simulated building response was validated in the frequency and time domains for the fixed-base building model. To this effect, the frequency response functions (FRFs) and the accelerations of the floors were used.

Figure 4 shows that the natural frequencies, which were numerically obtained, correspond to experimental measures with a maximum difference of 0.4%. In general, the energy (FRF amplitude) follows the pattern of the experimental results for all the modes and natural frequencies of the numerical model. Thus, the fixed-base building model was evaluated in the frequency domain.



(a) With base excitation



(b) With impact excitation on the first floor

Figure 4. FRFs comparison for the third floor
Source: Authors.

The acceleration response of the proposed numerical model and the experimental measures under white noise base excitation are depicted in Figure 5. It can be observed that the acceleration estimated via the numerical simulation agrees with the experimental response regardless of the acceleration level reached.

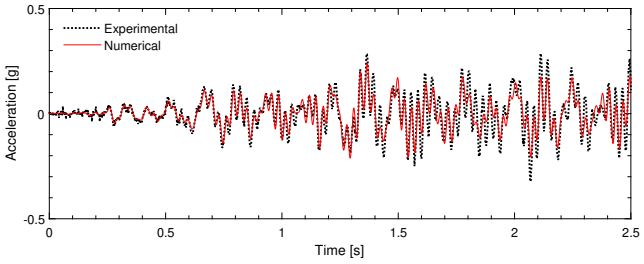
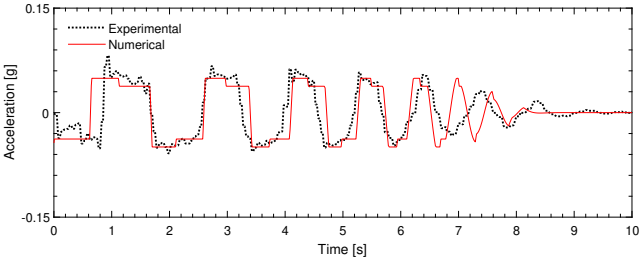


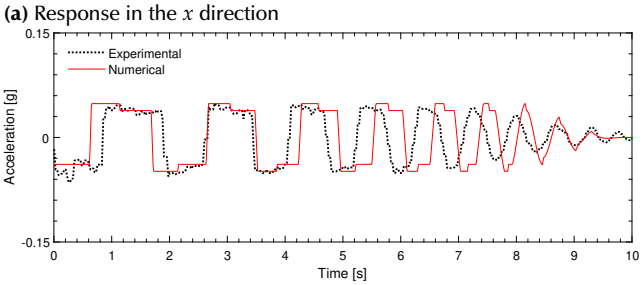
Figure 5. Time series of the acceleration response of the third floor under base excitation
Source: Authors.

In the RB system, the damping provided by the friction between the roller and the sloping surfaces was estimated from the experimental acceleration response of the RB system under free vibration. To this effect, an initial displacement for both horizontal directions was applied. In the numerical model, the roller friction parameter μ_r was tuned using an iterative manual scheme to match the RB response in both directions. A value of $\mu_r = 0.0057$ was found to be suitable for each motion direction. To validate the dynamical behavior of the bidirectional RB isolation system, masses of 1.90 and 2.78 kg were considered for the intermediate floors and the base plate, respectively. These values correspond to the physical properties of this structure.

Figure 6 compares the free vibration response of the physical and numerical models. It can be observed that the numerical representation follows the same pattern recorded in the experimental test in both directions.



(a) Response in the x direction

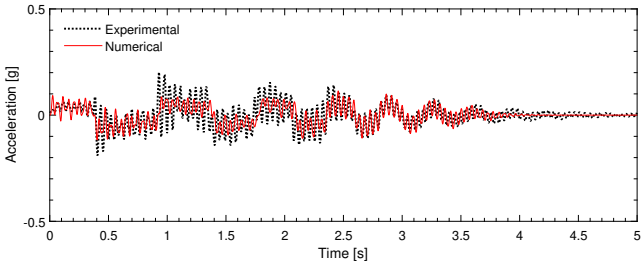


(b) Response in the y direction

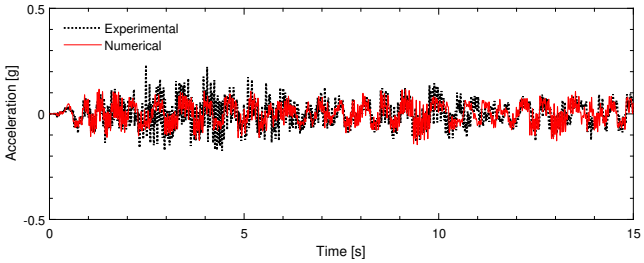
Figure 6. Time series of the acceleration response of the RBs under free vibration
Source: Authors

A time lag between the numerical and experimental acceleration is observed when the RBs are close to the equilibrium position, *i.e.*, the vertex of V-shape surfaces. This limitation in the model fit is associated with surface imperfections in the vertex, which increase friction and compromise the performance of the isolator. The repetitive path of numerically simulated response advanced the experimental acceleration record after 6.5 s, indicating that the damping of the RB device is reduced with time (Figure 6a). Furthermore, in the orthogonal direction, the response in the y-direction (Figure 6b) shows that the square form of the numerical acceleration has a time delay with the experimental acceleration after 4.0 s, indicating that the damping of the RB device increases as the displacement of the RBs decreases. Although the model does not capture the experimental measurements with absolute accuracy, particularly when the system is close to stopping, it can correctly represent the overall behavior of the system.

The acceleration response of the building coupled with the RB system was analyzed under free vibration and subjected to white noise excitation, as shown in Figure 7. It can be seen that the simulated response of the numerical model follows the trend of the target values (labeled as *experimental* in Figure 7) for this kind of dynamic load, which is particularly important for providing a reliable analysis in the following sections.



(a) With RBs under free vibration



(b) With RBs under base excitation

Figure 7. Time series of the acceleration response of the second floor
Source: Authors

RB system performance against near-fault earthquakes

This section evaluates the performance of the RB system in reducing the structural response of buildings when subjected to the horizontal components of two near-fault seismic records.

The two structures considered are the product of a partial modification to the validated numerical model, which aimed to obtain (i) an S1 model in which the vibration frequencies of the first two modes were less than 2.0 Hz and (ii) an S2 model in which the vibration frequencies of the first two modes were greater than 10.0 Hz, representing buildings with low and high lateral stiffness, respectively.

This modification was made by affecting the inertias I_y and I_z of type 1 components (*i.e.*, columns). For the S1 model, I_y was reduced to $3.0 \cdot 10^{-12}\,\mathrm{m}^4$, and I_z to $2.0 \cdot 10^{-12}\,\mathrm{m}^4$. Moreover, for the S2 model, I_y was reduced to $1.0 \cdot 10^{-10}\,\mathrm{m}^4$, and I_z was increased to $1.0 \cdot 10^{-10}\,\mathrm{m}^4$. Table 3 shows the predominant frequencies of the modified structures corresponding to the first three modes of vibration.

Table 3. Vibration modes and frequencies of models S1 and S2

Mode [type-direction]	Frequency [Hz]	
	S1	S2
1st flexural-weak	1.51	10.21
2nd flexural-strong	1.85	10.30
3rd torsional	3.09	10.73

Source: Authors

Near-fault earthquakes are generally characterized by long-duration, large-magnitude displacement pulse with large accelerations and a limited frequencies band compared to far-fault earthquakes. This form of seismic pulse (Mukhopadhyay and Gupta, 2013) creates a significant isolator displacement demand (Rawat *et al.*, 2018). Considering the above, we used near-fault seismic records obtained from PEER Ground Motion Database (Ancheta *et al.*, 2014), as shown in Table 4 were used. It should be noted that these records have been widely employed in the literature (Ou, Song, and Lee, 2010; Ortiz-Cano *et al.*, 2015; Rawat *et al.*, 2018), since they provide appreciably large ground displacement quantities induced by the hanging wall effect (Donahue and Abrahamson, 2014) and or the fling-step effect (X. Chen, Liu, Zhou, and Yang, 2020). Vertical ground motions components were not used because they have a negligible effect on the performance of different isolation devices (Ou *et al.*, 2010; Beirami-Shahabi, Zamani-Ahari, and Barghian, 2019).

Table 4. Near-fault earthquake records

Data	Earthquake			
Event name	Northridge - 01		Imperial Valley-06	
Date	January 17, 1994		October 15, 1979	
Record	Newhall - FS		El Centro - Array #5	
Component	90	360	140	230
PGD[1] [cm]	17.6	34.3	48.9	75.2
PGA[2] [g]	0.566	0.59	0.529	0.383

[1]Peak ground displacement
[2]Peak ground acceleration

Source: Authors

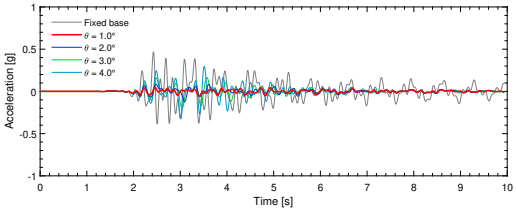
Since the S1 and S2 models were derived from a physical model, the displacement amplitude of the selected seismic records was scaled to 5% of their PGD. The ground motions were scaled using the Make Quake program software (Quanser Consulting Inc., 2010), which implements the scaling algorithm developed by (Kausel and Ushijima, 1979), allowing to preserve both the acceleration magnitude and the frequency content of the original seismic signal. This procedure incorporates a parabolic baseline correction over different integration schemes in the frequency domain, *i.e.*, the continuous (aliasing), pseudo-continuous, central difference, and linear acceleration schemes, obtaining consistent results between the time domain and the frequency domain.

First, the performance of the S1 and S2 models when subjected to excitations under fixed base conditions was studied. Then, the RB system was added, considering four slope angles in both directions of the V-shaped surfaces: 1.0, 2.0, 3.0, and 4.0°. The selected performance parameters were the root mean square (RMS) of the top floor's absolute acceleration of the structure ($\ddot{u}_{n=N}$, where N is the top floor) and the relative displacement of the isolator ($u_b = u_{n=0} - u_g$, where u_g is the induced displacement of the ground motion).

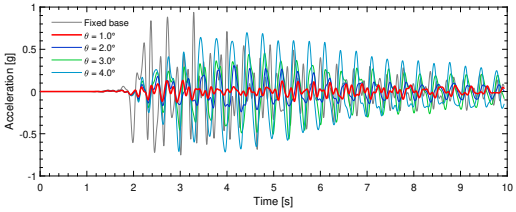
The absolute accelerations and base displacements provide a proportion of the lateral forces induced by the seismic excitation and the critical design point of the isolator devices, respectively (Jangig, 2000; Rawat *et al.*, 2018). Figures 8 and 9 present the acceleration response of the fourth floor in both orthogonal directions and the hysteresis diagram of the RB system for each structure when subjected to the most challenging ground motions.

Figures 8 and 9 show that all configurations reduce the structural response of models S1 and S2. The acceleration at the top floor is reduced as its slope angle decreases, achieving a peak acceleration reduction of 80-95% in the lowest slope angle configuration. As expected, the inclination of the plate induces additional forces (restoration

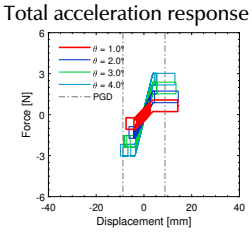
forces) that increase the horizontal acceleration of the structure. However, the application of the V-shape surface includes an energy dissipation system, where re-centering capabilities are required to avoid residual displacements, as described below.



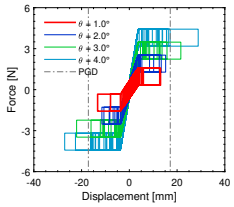
(a) Total acceleration response of the fourth floor in the strong direction



(b) Total acceleration response of the fourth floor in the weak direction



(c) RB hysteresis diagram in the strong direction

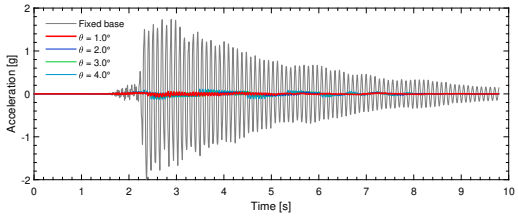


(d) RB hysteresis diagram in the weak direction

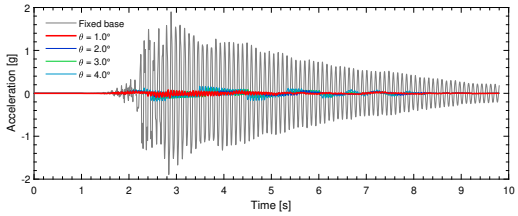
Figure 8. Response of model S1 when subjected to the scaled earthquake record of Northridge (1994) Newhall - FS
Source: Authors

Figure 8b shows that, even though the peak acceleration is reduced by the RB system in all cases, the maximum amplitude of the isolated structure's response to the excitation is close to that of the structure with the fixed base for a 3.0° inclination angle. To evaluate the changes in the total duration of the building's response records, the RMS was used as a suitable response variation indicator. Thus, the performance of the RB system in terms of the RMS was analyzed for each of the simulations carried out. To this effect, we determined the control effectiveness (CE) of the response of the building with the RB system on a particular story as follows:

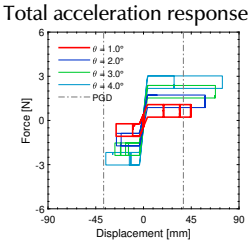
$$CE\,[\%] = \frac{\sigma_u - \sigma_c}{\sigma_u} \cdot 100 \qquad (11)$$



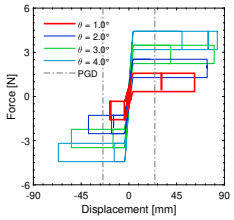
(a) Total acceleration response of the fourth floor in the strong direction



(b) Total acceleration response of the fourth floor in the weak direction



(c) RB hysteresis diagram in the strong direction

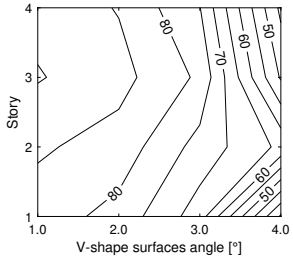


(d) RB hysteresis diagram in the weak direction

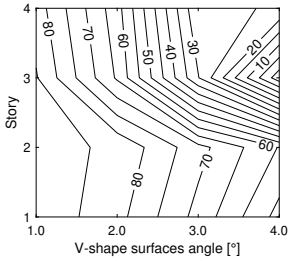
Figure 9. Response of model S2 when subjected to the scaled earthquake record of Imperial Valley (1979) El Centro - Array #5
Source: Authors

where σ_c and σ_u denote the directional combination of the RMS values corresponding to both orthogonal directions of the building's motions (*i.e.*, σ_x and σ_z) with and without the RB system. Thereby, σ_c and σ_u were calculated for each story of the structure by means of the square root of the sum of the squares (SRSS) of the RMS values in the x and z directions (*i.e.*, $\sqrt{\sigma_x^2 + \sigma_z^2}$).

Figures 10 and 11 summarize the CE of the response of the building with the RB system for different rolling surface slope angles per floor.

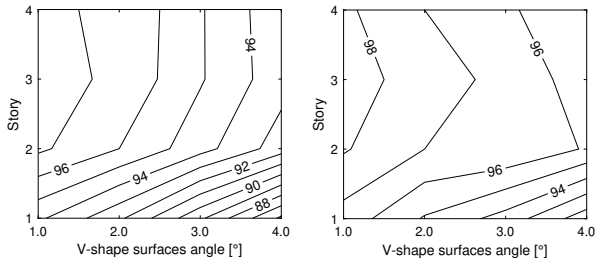


(a) Imperial Valley (1979) El Centro - Array #5



(b) Northridge (1994) Newhall - FS

Figure 10. Contour plot of acceleration CE *vs.* V-shape surface angle for the S1 structure
Source: Authors.



(a) Imperial Valley (1979) El Centro - Array #5 **(b)** Northridge (1994) Newhall - FS

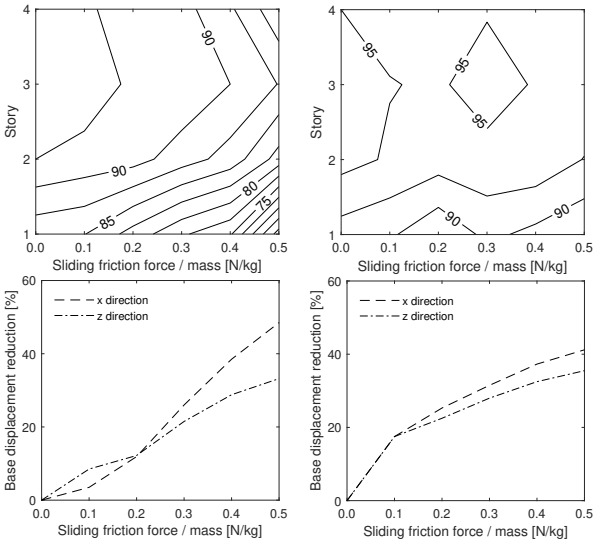
Figure 11. Contour plot of acceleration CE *vs.* V-shape surface angle for the S2 structure
Source: Authors

In the S1 model with the RB system, the response of the building can reach reductions of 5-85% (Figure 10). The accelerations at the top floor decreases with the slope angle, achieving reductions of 80-95% in peak acceleration with the lowest slope angle configuration. As expected, the inclination of the plate induced additional forces (restoration forces), increasing the horizontal acceleration of the structure. However, as mentioned before, the V-shape surface should have an energy dissipation system with re-centering capabilities.

As for the S2 model, with the RB system, the response of the building can achieve reductions close to 99% with a slope angle of 1.0° (Figure 11b). In particular, the acceleration CE of the RB device slightly decreases when the angle increases from 1.0 to 4.0º. As a result, the response of the structure reaches a minimum reduction of 86% on the critical story (*i.e.*, the first floor) when considering a slope angle of 4.0° and the Imperial Valley (1979) earthquake (Figure 11a).

Additionally, a set of simulations was performed, which included an energy dissipation system in the form of friction surfaces within the RB device, providing a sliding friction force of 0-0.5 N per kg of structure mass in both orthogonal directions in the horizontal plane (*i.e.*, the x and z directions). It was demonstrated, for the flexible building (S1 model) any combination of slope angle, bearing surfaces, and an energy dissipation force that significantly reduces the base displacement leads to an amplification of the building's response. This is stated upon the basis of the slope angles considered in this work (*i.e.*, 1.0, 2.0, 3.0, and 4.0°.

The results for the stiff building (S2 model) with a RB system and a slope angle of 4.0° are presented in Figure 12. For the sake of comparison, the force induced by the energy dissipation system was normalized with the mass supported by the RB system in each direction. This Figure shows that a significant reduction in the building response and its base displacement can be achieved by using a slope angle of 4.0° in the rolling plates, with a small sliding friction of 0.1 N/kg per supported mass.



(a) Imperial Valley (1979) El Centro - Array #5 **(b)** Northridge (1994) Newhall FS

Figure 12. Contour plot of acceleration CE *vs.* sliding friction per supported mass with $\theta = 4°$ for the S2 structure
Source: Authors

Figure 12 shows that the ability of the RB system to reduce the building's response decreases as the control of the base displacements increases, *i.e.*, as the force induced by the friction dissipation system increases. Thus, the greatest building response reduction with a minimal base displacement can be achieved through a proper combination of the the V-shaped surface's inclination angle and the energy dissipation force.

Conclusions

This paper presents a numerical study aimed at evaluating the performance of a bidirectional RB system in reducing the structural response of a multi-column system subjected to the horizontal components of two near-fault earthquake records. This reduction was evaluated via a direct comparison and the RMS of the simulation results obtained under fixed-base conditions and with the RB system. The following conclusions are drawn:

- The RB system manages to significantly reduce the response of the evaluated structures when subjected to near-fault earthquake excitations. A direct comparison shows reductions of up to 80% in peak acceleration for the flexible structure and 98% for the rigid structure. In terms of the RMS, acceleration reductions of 5-85% are achieved for the flexible structure, in addition to 86-96% for the rigid structure.

- One way to control the displacements generated at the base is the use of energy dissipation systems. In this numerical study, frictional energy dissipation systems were simulated, which could reduce approximately 35% of the base displacements and ensure a RMS reduction of at least 60% in the total acceleration of the rigid structure. As for the flexible structure, it was found that a reduction in the base displacements inevitably leads to the amplification of the building's response under fixed-base conditions, which indicates

that using this displacement control system is not suitable in this context.

- An adequate relationship between the angle of the inclined surfaces and the parameters associated with the energy dissipation system will result in maximal reduction in the total acceleration response and minimal base displacement for a particular building and a set of seismic excitations.

Despite that, the analyses yielded encouraging results. However, note that the conclusions reached are limited to scaled building models and seismic records. To generalize the implications of this study, additional research on full-scale numerical structural models should be performed, using several ground motion records that represent different characteristics in terms of frequency and amplitude.

Acknowledgements

Product derived from the project INV-ING-2982 funded by Research Vice-Principalship of Universidad Militar Nueva Granada - Year 2019. This support is gratefully acknowledged.

References

Ancheta, T., Darragh, R., Stewart, J., Seyhan, E., Silva, W., Chiou, B., ... Donahue, J. (2014). Earthquake spectra. *NGA-West2 Database*(30(3):989-1005. https://doi.org/10.1193/070913EQS197M).

Andreaus, U., and De Angelis, M. (2020). Influence of the characteristics of isolation and mitigation devices on the response of single-degree-of-freedom vibro-impact systems with two-sided bumpers and gaps via shaking table tests. *Structural Control and Health Monitoring*, *27*(5), e2517. https://doi.org/10.1002/stc.2517.

Bagerzadeh-Karimi, M., and Geneş, M. (2019). Probabilistic behavior assessment of base-isolated buildings and base isolation systems subjected to various earthquakes with different components. *Arabian Journal for Science and Engineering*, *44*(10), 8265-8288. https://doi.org/10.1007/s13369-019-03867-x.

Beirami-Shahabi, A., Zamani-Ahari, G., and Barghian, M. (2019). Suspended Columns for Seismic Isolation in Structures (SCSI): A preliminary analytical study. *Earthquakes and Structures*, *16*(6), 743-755. https://doi.org/10.12989/eas.2019.16.6.743.

Beirami-Shahabi, A., Zamani-Ahari, G., and Barghian, M. (2020). Base isolation systems: A state of the art review according to their mechanism. *Journal of Rehabilitation in Civil Engineering*, *8*(2), 37-61. https://doi.org/10.22075/jrce.2019.16186.1306.

Calhoun, S. (2018). *Evaluation of rolling-type isolation systems for seismic hazard mitigation* (Unpublished master's thesis). The University of Oklahoma.

Calhoun, S., Tehrani, M., and Harvey-Jr, P. (2019). On the performance of double rolling isolation systems. *Journal of Sound and Vibration*, *449*, 330-348. https://doi.org/10.1016/j.jsv.2019.02.030.

Chen, P., Hsu, S., Zhong, Y., and Wang, S. (2021). Real-time hybrid simulation of smart base-isolated raised floor systems for high-tech industry. *Smart Structures and Systems*, *23*(1), 91-106. https://doi.org/10.12989/sss.2019.23.1.091.

Chen, X., Liu, Y., Zhou, B., and Yang, D. (2020). Seismic response analysis of intake tower structure under near-fault ground motions with forward-directivity and fling-step effects. *Soil Dynamics and Earthquake Engineering*, *132*, 106098. https://doi.org/10.1016/j.soildyn.2020.106098.

De Luca, A., and Guidi, L. (2019). State of art in the worldwide evolution of base isolation design. *Soil Dynamics and Earthquake Engineering*, *125*, 105722 https://doi.org/10.1016/j.soildyn.2019.105722.

Donahue, J., and Abrahamson, N. (2014). Simulation-based hanging wall effects. *Earthquake Spectra*, *30*(3), 1269-1284 https://doi.org/10.1193/071113EQS200M.

Erdik, M., Ulker, O., Şadan, B., and Tuzun, C. (2018). Seismic isolation code developments and significant applications in Turkey. *Soil Dynamics and Earthquake Engineering*, *115*, 413-437 https://doi.org/10.1016/j.soildyn.2018.09.009.

Foti, D. (2019). Rolling devices for seismic isolation of lightweight structures and equipment. design and realization of a prototype. *Structural Control and Health Monitoring*, *26*(3), e2311 https://doi.org/10.1002/stc.2311.

Foti, D., Catalan Goni, A., and Vacca, S. (2013). On the dynamic response of rolling base isolation systems. *Structural Control and Health Monitoring*, *20*(4), 639-648 https://doi.org/10.1002/stc.1538.

Harvey-Jr, P., and Gavin, H. (2015). Assessment of a rolling isolation system using reduced order structural models. *Engineering Structures*, *99*, 708-725 https://doi.org/10.1016/j.engstruct.2015.05.022.

Hosseini, M., and Soroor, A. (2011). Using orthogonal pairs of rollers on concave beds (OPRCB) as a base isolation system Part I: Analytical, experimental and numerical studies of (OPRCB) isolators. *The Structural Design of Tall and Special Buildings*, *20*, 928-950. https://doi.org/10.1002/tal.568.

Hosseini, M., and Soroor, A. (2013). Using orthogonal pairs of rollers on concave beds (OPRCB) as a base isolation system Part II: Application to multi-story and tall buildings. *The Structural Design of Tall and Special Buildings*, *22*, 192-216. https://doi.org/10.1002/tal.671.

Jangig, R. (2000). Stochastic seismic response of structures isolated by rolling rods. *Journal of Structural Engineering*, *22*(8), 937-946 https://doi.org/10.1016/S0141-0296(99)00041-3.

Kausel, E., and Ushijima, R. (1979). *Baseline correction of earthquake records in the frequency domain* (Tech. Rep. No. R79-34). Cambridge, MA, USA: Massachusetts Institute of Technology.

Lee, G., Liang, Z., and Snyder, G. (2005). *Seismic isolation bearing, PCT WO 2005/031088 A2*.

Lee, G., Ou, Y., Niu, T., Song, J., and Liang, Z. (2010). Characterization of a roller seismic isolation bearing with supplemental energy dissipation for highway bridges. *Journal of Structural Engineering*, *136*(5), 502-510 https://doi.org/10.1061/(ASCE)ST.1943-541X.0000136.

Matsagar, V., and Jangid, R. (2008). Base isolation for seismic retrofitting of structures. *Practice periodical on structural design and construction*, *13*(14), 175-185 https://doi.org/10.1061/(ASCE)1084-0680(2008)13:4(175).

Menga, N., Foti, D., and Carbone, G. (2017). Viscoelastic frictional properties of rubber-layer roller bearings (RLRB) seismic isolators. *Meccanica*, *52*, 2807-2817 https://doi.org/10.1007/s11012-016-0612-y.

Mukhopadhyay, S., and Gupta, V. (2013). Directivity pulses in near-fault ground motions|i: Identification, extraction and modeling. *Soil Dynamics and Earthquake Engineering*, *50*, 1-15 https://doi.org/10.1016/j.soildyn.2013.02.017.

Naeim, F., and Kelly, J. (1999). *Design of seismic isolated structures: From theory to practice*. New York, USA: Wiley & Sons https://doi.org/10.1002/9780470172742.

Nobari Azar, F., Karimzadeh Naghshineh, A., and Sen, M. (2022). Preparation and characterization of natural rubber based new elastomers for high-damping base isolation systems. *Journal of Elastomers & Plastics*, *54*(6), 959-974. https://doi.org/10.1177/00952443221075505.

Ortiz-Cano, N. (2013). *Avaliacão de sistemas de isolamento para o controle de vibracões de edifícios submetidos a excitacões de base* (Unpublished doctoral dissertation). Universidade Federal do Rio de Janeiro.

Ortiz-Cano, N., Magluta, C., and Roitman, N. (2014). Numerical and experimental studies of a building with elastomeric and roller seismic isolation bearing. In *9th international conference on structural dynamics*.

Ortiz-Cano, N., Magluta, C., and Roitman, N. (2015). Numerical and experimental studies of a building with roller seismic isolation bearings. *Structural Engineering & Mechanics*, *54*(3), 475-489. https://doi.org/10.12989/sem.2015.54.3.475.

Ou, Y., Song, J., and Lee, G. (2010). A parametric study of seismic behavior of roller seismic isolation bearings for highway bridges. *Earthquake Engineering & Structural Dynamics*, *39*(5), 541-559. https://doi.org/10.1002/eqe.958.

Quanser Consulting Inc. (2010). *Make Quake Program*.

Rawat, A., and Matsagar, V. (2021). An oblate spheroid base isolator and floating surface diaphragm for seismic protection of liquid storage tank. *Journal of Earthquake Engineering*, *26*(10), 5447-5475. https://doi.org/10.1080/13632469.2021.1875939.

Rawat, A., Ummer, N., and Matsagar, V. (2018). Performance of bi-directional elliptical rolling rods for base isolation of buildings under near-fault earthquakes. *Advances in Structural Engineering*, *21*(5), 675-693. https://doi.org/10.1177/1369433217726896.

Ryan, K., Okazaki, T., Coria, C., Sato, E., and Sasaki, T. (2018). Response of hybrid isolation system during a shake table experiment of a full-scale isolated building. *Earthquake Engineering and Structural Dynamics*, *47*, 2214-2232. https://doi.org/10.1002/eqe.3065.

Sanchez-Torres, D., Rico-Caviedes, A., Ortiz-Cano, N., Nieto-Leal, A., and Gaviria-Mendoza, C. (2019). Assessment of a roller seismic isolation bearing for buildings under bidirectional excitations. In *XL Ibero-Latin-American Congress on Computational Methods in Engineering*.

Shampine, L., Reichelt, M., and Kierzenka, J. (1999). Solving Index-1 DAEs in MATLAB and Simulink. *SIAM Review*, *18*(3), 538-552. https://doi.org/10.1137/S003614459933425X.

The MathWorks Inc. (2019). *MATLAB R2019a*.

Tsai, M., Wu, S., Chang, K., and Lee, G. (2007). Shaking table tests of a scaled bridge model with rolling-type seismic isolation bearings. *Engineering Structures*, *29*, 694-702. https://doi.org/10.1016/j.engstruct.2006.05.025.

Walters, M. (2015). Seismic isolation: The gold standard of seismic protection. In *Structural performance*. Structure Magazine. https://www.structuremag.org/?p=8770:~:text=Thecurrentlyapplicableconceptof,astheearthmovesviolently.

Wang, S., Hwang, J., Chang, K., Shiau, C., Lin, W., Tsai, M., ... Yang, Y. (2014). Sloped multi-roller isolation devices for seismic protection of equipmente and facilities. *Earthquake Engineering & Structural Dynamics*, *43*, 1443-1461. https://doi.org/10.1002/eqe.2404.

Wang, S., Lin, W., Yang, C., Chang, Y., K. Huang, Hwang, J., and Hwang, J. (2017). Recent progress in Taiwan on seismic isolation, energy dissipation, and active vibration control. In *15th world conference on seismic isolation, energy dissipation, and active vibration control of structures*.

Zhang, C., and Ali, A. (2021). The advancement of seismic isolation and energy dissipation mechanisms based on friction. *Soil Dynamics and Earthquake Engineering*, *146*, 106746. https://doi.org/10.1016/j.soildyn.2021.106746.

Research Article/Electronics and Telecommunications Engineering

http://doi.org/10.15446/ing.investig.106310

Channel Operating Margin Exploration as a Complementary Transceiver Circuit Design Tool for 25 Gbps PAM4 Serial Links

Exploración del margen operativo del canal como herramienta complementaria para el diseño de circuitos transceptores con enlaces seriales PAM4 de 25 Gbps

Luisa F. Dovale-Vargas[1], Oscar M. Reyes-Torres[2], and Elkim F. Roa-Fuentes[3]

ABSTRACT

The design of high-speed serial links continues to attract the attention of the electronics industry due to the steady development of different telecommunications standards, generating a constantly growing data rate and new modulation schemes. However, conventional certification metrics can lead to sub-optimal transmit (Tx) and receive (Rx) circuit designs. Therefore, the Ethernet standard IEEE 802.3bj introduced a more effective evaluation method called *channel operating margin* (COM) to explore the design space at an early stage. Although the advantages of COM have been discussed in the literature and only a few works explore its potential as a backplane design tool, there are no reports on the use of COM as a complementary design tool for transceiver circuits. This work studies the use of COM as a complementary tool for transceiver design. COM performance is evaluated for four 100GBASE-KP4 backplanes and different equalization architectures. The impact of the metric and the challenges associated with incorporating new equalization structures into the COM flow are discussed. The results reveal a conventional Tx-Rx architecture that exceeds the COM threshold and an alternative one that improves the opening of the eye diagram but does not exceed the threshold.

Keywords: COM, equalizers, IEEE 802.3ck, eye diagram, single-bit response

RESUMEN

El diseño de enlaces seriales de alta velocidad continúa atrayendo la atención de la industria electrónica debido al desarrollo constante de diferentes estándares de telecomunicaciones, generando una tasa de datos en constante crecimiento y nuevos esquemas de modulación. Sin embargo, las métricas de certificación convencionales pueden llevar a diseños de circuitos de transmisión (Tx) y recepción (Rx) subóptimos. Por lo tanto, el estándar Ethernet IEEE 802.3bj introdujo un método de evaluación más efectivo llamado *margen de operación del canal* (COM, por sus siglas en inglés) para explorar el espacio de diseño en una etapa temprana. Aunque las ventajas de COM han sido discutidas en la literatura y solo unos pocos trabajos exploran su potencial como una herramienta de diseño de *backplane*, no hay informes sobre el uso de COM como una herramienta de diseño complementaria para circuitos transceptores. Este trabajo estudia el uso de COM como una herramienta complementaria para el diseño de transceptores. El rendimiento de COM se evalúa para cuatro *backplanes* 100GBASE-KP4 y diferentes arquitecturas de ecualización. Se discuten el impacto de la métrica y los desafíos asociados a la incorporación de nuevas estructuras de ecualización en el flujo de COM. Los resultados revelan una arquitectura Tx-Rx convencional que supera el umbral de COM y una alternativa que mejora la apertura del diagrama de ojo pero no supera el umbral de COM.

Palabras clave: COM, ecualizadores, IEEE 802.3ck, diagrama de ojo, respuesta a único bit

Received: May 1st 2023
Accepted: February 27th 2024

Introduction

Many high-speed serial links employ SerDes systems for transmission speeds of up to 112 Gb/s (Champion and Tracy, 2019). However, for rates exceeding 25 Gb/s, strict transceiver design is necessary to compensate for noise and crosstalk, ensuring compliance with certification metrics (Dong *et al.*, 2014). Design parameters within the metric (Figure 1) must be met for each transmission speed. Evaluating parameters individually leads to performance misinterpretation, and inaccurate channel loss estimation affects transceiver circuit design (Filip, 2018).

The IEEE 802.3bj Ethernet standard introduced channel operating margin (COM) as a global certification metric for high-speed serial channels (IEEE, 2014). COM simplifies evaluation by measuring a single signal-to-noise ratio, encompassing the entire SerDes system. It aims to address channel performance ambiguities and facilitate trade-off analysis, potentially reducing transceiver complexity. However, COM documentation is limited. Works such as those by de Paulis *et al.* (2020), de Paulis *et al.* (2022), and

[1]Electronics engineer, Universidad Industrial de Santander, Colombia. MSc in Telecommunications Engineering, Universidad Industrial de Santander, Colombia. E-mail: luisa.dovale@saber.uis.edu.co
[2]Electronics engineer, Universidad Industrial de Santander, Colombia. MSc in Electronics, Universidad Industrial de Santander, Colombia. Dr-Ing, Hamburg University of Technology, Germany. Affiliation: Full professor, Universidad Industrial de Santander, Colombia. E-mail: omreyes@uis.edu.co
[3]Electrical engineer, Universidad Industrial de Santander, Colombia. MSc in Electrical Engineering, University of São Paulo, Brazil. PhD in Electrical Engineering, Purdue University, United States. Affiliation: Full professor, Universidad Industrial de Santander, Colombia. E-mail: efroa@uis.edu.co

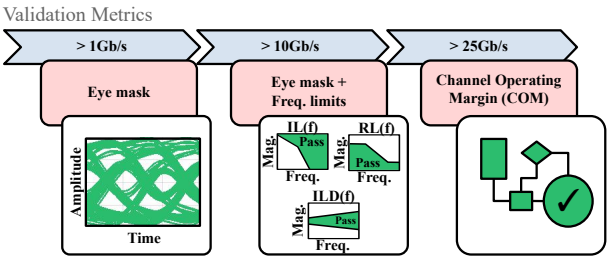


Figure 1. Evolution of validation metrics according to the data rate
Source: Authors

Gore and Mellitz (2014) explore the use of COM oriented to *backplanes* design, with the absence of any report that uses COM as a complementary tool for transceiver architecture design.

Our work evaluates the feasibility of using COM as a transceiver design mechanism, addressing compliance issues with traditional strategies. We focus on Tx/Rx modeling within the COM validation metric. Modifications to the COM flow include incorporating new receiver-side equalization architectures. Specifically, decision feedback equalizer (DFE) architectures are studied using infinite impulse response (IIR) taps instead of discrete taps. Previous studies have demonstrated the advantages of DFE with IIR filters in terms of power consumption and equalization efficiency (Shahramian and Chan Carusone, 2015; Chen *et al.*, 2017; Elhadidy and Palermo, 2013; Kim *et al.*, 2009; Shahramian *et al.*, 2016). DFE with IIR filters effectively cancels inter-symbol interference (ISI) with low circuit complexity. However, implementing DFE IIR within the COM computation flow introduces certain drawbacks, which are thoroughly discussed in this research.

The performance of various equalization architectures is evaluated using four 50 Gb/s channels from different companies, based on the parameters and specifications of the IEEE 802.3ck standard (?, ?). We discuss the performance of COM with different transceivers, identifying the one with the minimal complexity that achieves a final COM value greater than 3 dB for all cables. Additionally, we provide recommendations on using COM as a design mechanism.

This work explains COM and its impact on transceiver architecture, describing the utilization of COM as a tool for exploring transceiver architecture, and summarizes the COM results obtained for different equalization architectures. Additionally, it discusses considerations for incorporating different DFE architectures into the COM algorithm, presenting the results for different architecture variations. Finally, the main conclusions of this work are stated.

Channel operating margin (COM)

COM is a mathematical algorithm that determines a figure of merit (FoM) in decibels. The FoM is a signal-to-noise ratio (SNR) that compares the sampled signal against the total noise contribution. COM results offer valuable information regarding transmission quality for a specific transmission system (Tx-channel-Rx). For channels exceeding 25 Gb/s, the standard requires the SNR to exceed the 3 dB threshold for proper system functionality. As a time-domain metric, COM can also analyze interconnected systems, pinpointing strengths and weaknesses in Tx/Rx design. It serves as a communication bridge between signal integrity and SerDes designers.

Description and limitations of the COM algorithm

COM evaluates SerDes system performance under different conditions: transceiver circuit complexity, signal modulation, several channel impairments, and others, aiming to reduce the circuit design complexity. The user can import the s-parameter of the channel (these s-parameters can be downloaded from the repository of the 802.3ck IEEE standard) and the number of taps for the feed-forward equalizer (FFE) and DFE or gain for the continuous-time linear equalizer (CTLE).

COM transceiver architecture

The COM transceiver topology includes a FFE on Tx and a CTLE along with a DFE on Rx. These equalizers mitigate ISI. The FFE primarily reduces pre-cursor ISI, while the CTLE and DFE reduce post-cursor ISI.

The COM FFE has three pre-cursors (c(-3), c(-2), c(-1)), one post-cursor (c(1)), and a main cursor (c(0)). The user defines the range for each tap. The weight of the main cursor (c(0)) is calculated using the following equation: $c(0) = 1 - |c(1)| - |c(-1)| - |c(2)| - |c(-3)|$.

After FFE, COM calculates the CTLE parameters. CTLE consists of two stages: the first stage (H_{CTLE}) for high frequencies has one zero around the Nyquist frequency (f_z) and two poles (f_{p1} and f_{p2}). The second stage (H_{CTLE2}) has a pole and a zero at the same frequency (f_{LF}), approximately 1/3 of the Nyquist frequency.

$$H_{CTLE}(f) = \frac{10^{\frac{g_{DCHF}}{20}} + j\frac{f}{f_z}}{(1 + j\frac{f}{f_{p1}})(1 + j\frac{f}{f_{p2}})} \quad (1)$$

$$H_{CTLE2}(f) = \frac{10^{\frac{g_{DCLF}}{20}} + j\frac{f}{f_{LF}}}{(1 + j\frac{f}{f_{LF}})} \quad (2)$$

Figure 2 shows the behavior of the HF stage (Figure 2a) and the LF stage (Figure 2b) for the different DC gain values, marking f_{p2} for HF and f_{LF} for LF. The pole, zero, and DC gain parameters are defined externally by the user.

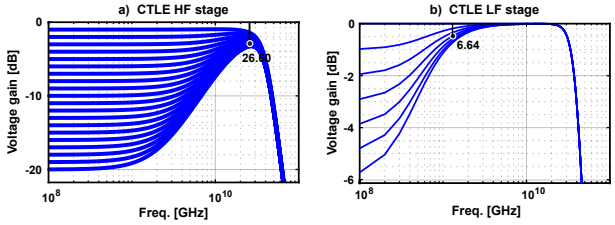


Figure 2. COM CTLE magnitude TF example. a) CTLE voltage gain for high frequencies; b) CTLE voltage gain for low frequencies.
Source: Authors

After CTLE, the next equalization stage is DFE. The COM DFE model is ideal and supports 1-35 post-cursor taps. The phase detector algorithm used in the CDR criterion is similar to that proposed by Mueller and Muller (1976), defining the

sample point from the channel equalized single-bit-response (SBR).

Transceiver optimization cycle

To maximize the FoM, the algorithm exhaustively searches all possible combinations of the TxFFE, CTLE, and DFE parameters, evaluating a unique SBR for each combination.

$$FoM = 10\log_{10}\left(\frac{A_s^2}{\sigma_{Tx}^2 + \sigma_{ISI}^2 + \sigma_J^2 + \sigma_{XT}^2 + \sigma_N^2}\right) \quad (3)$$

The numerator A_s corresponds to the value of the available signal, and the denominator parameters correspond to the noise contributions. The sum of denominator terms determines the amplitude of the noise interference.

The size of the COM exploration space depends on the number of parameters defined in each equalizer. For example, a transceiver architecture with two different values for HF DC gain, one value for LF DC gain, 3-FFE taps, and 3-DFE taps results in an eight-dimensional space that needs to be optimized.

Since COM was developed in the MATLAB sequential environment, to calculate the FoM, it implements a series of steps to optimize each parameter:

- First, the algorithm evaluates the Equations 1 and 2 for the first values of the HF and the LF DC gain to determine the transfer function of the CTLE.
- Then, the algorithm sequentially computes the amplitude of each FFE tap (c(-3), c(-2), c(-1), c(1)). The amplitude of c(0) must be must meet the condition that $c(0) = 1 - |c(1)| - |c(-1)| - |c(2)| - |c(-3)| > 0.5$ to ensure that all transmitters support the FFE architecture. If c(0) does not reach the threshold, the algorithm selects another combination of amplitudes until c(0) exceeds 0.5. The FFE function is constructed and applied to the impulse response along with the CTLE transfer function to calculate the SBR for each channel.
- Next, COM estimates the sampling time, the value of the available signal A_s, and the amplitude of the DFE tap from the SBR of the THRU channel. A_s is calculated using Equation 4, where t_s is the sample point and R_{LM} is the level mismatch ratio. Additionally, COM computes an approximation of noise from a normal distribution of each noise contribution.

$$A_s = \frac{R_{LM} * SBR(t_s)}{(level - 1)} \quad (4)$$

- Finally, COM calculates the FoM using A_s and the noise approximation through Equation 3, where the noise approximation corresponds to the denominator and A_s to the numerator.

This process repeats for all possible combinations of the equalization parameters, obtaining a different FoM for each cycle. All FoM results are compared, selecting the best result selected and extracting the equalization parameters.

COM as transceiver exploration

SerDes architectures comprise a pre-emphasis on TX, usually an FFE, along with CTLE and DFE on Rx, as shown in Figure 3. Traditional transceiver design methodology causes transceiver circuit over-design or incorrect system validation due to excessive design margins (Filip, 2018). This section explores how to avoid these problems, specifically examining the feasibility of using the specification itself to evaluate the impact of different equalizer settings. This evaluation consists of obtaining different COM values by varying equalizer settings such as the FFE coefficients, the DC gain of the CTLE, or the DFE coefficients. In addition to the variation of the FFE, CTLE, and DFE parameters, we also explore the impact of adding a second CTLE in the cascade connection on the optimization loop. The testbench used to evaluate the equalization architecture (EQA) performance is composed of four different channels from different companies, as provided by standard IEEE 802.3ck.

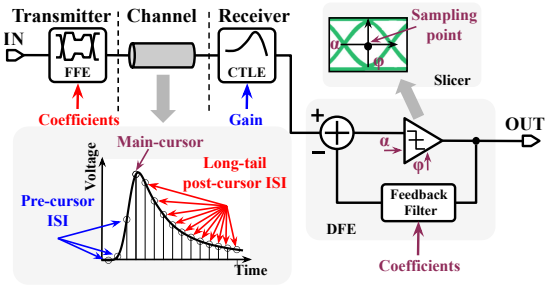


Figure 3. Typical equalization scheme of a serial-link transceiver
Source: Authors

COM exploration setup

To obtain the COM exploration space, we considered the conditions and parameters provided by the IEEE 802.3ck update, which are summarized in Table 1. The user can change the parameters defined in this Table. We can define the DC gain range of the HF stage (g_{DCHF}) and the LF stage (g_{DCLF}) for the CTLE, the number of activated taps for the FFE (from c(-3) to c(1)), as well as the amplitude range of each sample. Furthermore, with the N_b parameter, the user can activate up to 35-DFE taps. COM exploration space will be analyzed in two case studies: case study 1 corresponds to conventional equalization structures (FFE + CTLE +DFE), and case study 2 has an additional CTLE in its equalization structure.

Once the parameters are defined, the CTLE architecture is modified, adding an extra scale factor in Equation 1 as well as modifying the DC peaking gain from -3 to 3 dB (Figure 4). This change also applies to the additional CTLE used for case study 2.

The setup defined in Table 1 evaluates two different package length tests (z_p), one with a long package (length = 30 mm) and another one with a short package (length = 12 mm). Each package case results in a different COM value. The long package exhibits a lower COM due to more IL. Therefore, the reported results corresponded to a long package test. Additionally, Table 1 describes other important parameters, such as the signaling rate (f_b), the target detector error ratio (DER_0), and the level separation mismatch ratio (R_{LM}).

Table 1. Simulation parameters setup

	Parameter	Setting	Units
Signaling rate	f_b	53.125	Gb/s
Transmission line length	Z_p	[12 30]	mm
FFE main cursor	c(0)	0.5	
FFE pre-cursor coefficients	c(-1)	[-0.3:0.02:0]	
	c(-2)	[-0.04:0.02:0.12]	
	c(-3)	[-0.06:0.02:0.04]	
FFE post-cursor coefficient	c(1)	[-0.2:0.05:0.05]	
DFE length	N_b	15	UI
CTLE, HF DC gain	g_{DC}	[-16:1:0]	dB
CTLE, LF DC gain	g_{DC}	[-6:1:0]	dB
Number of samples per-unit interval	M	32	
Target detector error ratio	DER_0	1.00e-04	
Level separation mismatch ratio	R_{LM}	0.95	

Source: Authors

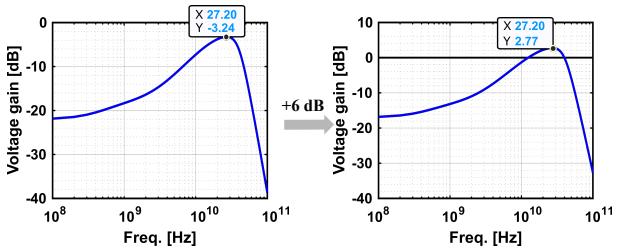


Figure 4. CTLE DC peaking gain modification
Source: Authors

Once COM exploration settings are defined, the four backplane channels from the IEEE 802.3ck study group (?, ?) are selected to test the EQAs' performance. The channel responses for channels Intel (Heck and Intel, 2018), Molex (Palkert and Molex, 2020), Samtec Op1, and Samtec Op2 (Mellitz *et al.*, 2018) are shown in Figure 5. All channels have an insertion loss of approximately 28 dB@26 GHz (Figure 5a). Figure 5b shows that the unequalized SBRs are normalized and centered by unit interval (UI) at f_b = 53.125 Gb/s, highlighting the three sample moments: the pre-cursor in green, the cursor as a black line, and the post-cursor in pink. In Figure 5a, the first two pre-cursor samples have the highest ISI contribution, suggesting that at least two taps of the Tx FFE are needed. In addition, the samples after the cursor determine the CTLE gains and the amplitude of the DFE leads.

Table 2. Characteristics of the test channel

Company	Molex	Intel	Samtec	
			Op1	Op2
Channel Length (inch)	16	22.63	39.3	39.3
IL target (dB) @26.5 GHz	**-28.97**	-28.84	-27.15	-26.35
ERL (dB)	**14.82**	16.58	21.17	17.54

Source: Authors

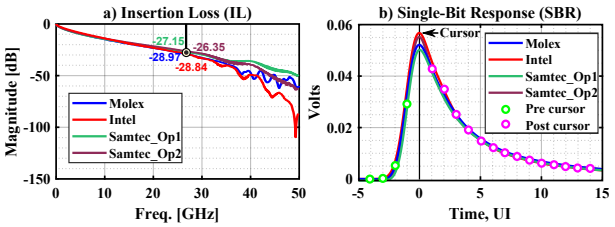


Figure 5. Test channel responses: a) insertion loss of all channels with an IL around 28 dB@26.56 GHz; b) SBR of the four backplane channels
Source: Authors

Table 2 summarizes the main characteristics of each channel. Highlighted in red, the Molex is the critical channel of the group, with the highest loss estimate (the highest insertion loss and the lowest effective return loss values). With this information, and using the traditional validation metric based on frequency limits, we can predict that the Molex channel might need a more complex transceiver circuit and perhaps report the lowest COM results relative to the other channels.

Simulations results and comparison

Figures 6 present four EQAs. The first EQA (Figure 6a) evaluates the impact of the variation of the FFE taps *vs.* the final value of the COM. This test performs a parametric sweep of the FFE taps, where 0 represents the absence of an FFE. The other three EQAs evaluate the impact of the variation of the DFE taps *vs.* the COM results via a parametric sweep of the DFE taps from 1 to 15 samples. The first two EQAs (6a and 6b) include all equalizers, while 6c and 6d do not contain one of the equalizers. The results show that, for all the cases that perform a parametric DFE sweep, COM saturation is achieved after a 3-tap DFE, suggesting that no more than 3-DFE taps are necessary. The minimal EQA reporting a COM result with an up to 3 dB margin is enclosed in a gray box (Figures 6a and 6b), where case 6a uses 5-FFE taps, the CTLE, and 1-DFE tap, and case 6b uses 3-FFE taps, the CTLE, and 3-DFE taps.

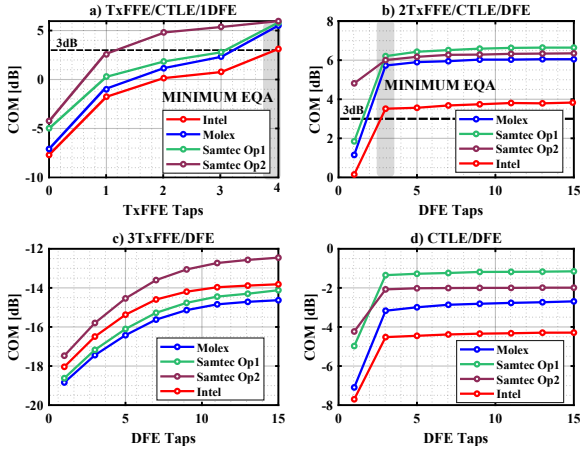


Figure 6. COM value for four common EQAs with a parametric sweep of TxFFE and DFE taps, including all contributions from the crosstalk channel. The complexity of the equalizer circuit increases from left to right on the x-axis.
Source: Authors

Similarly to case study 1, four different EQAs are selected for case study 2. However, we assumed that adding a second

CTLE in the cascade connection might decrease the number of DFE taps. This variation of the CTLE architecture makes no modifications to the predefined CTLE parameters. Figure 7 shows the COM results for the four EQAs. The first three EQAs (7a, 7b, and 7c) perform a parametric sweep of the FFE taps, but case 7c does not use any CTLE, and case 7d exhibit COM saturation after 3-DFE taps, as in case study 1. Note that, although the COM results behave similarly to those of case study 1, the former minimum architecture comprises two equal CTLEs. In addition, as in case study 1, the COM value does not meet the threshold if one of the equalizers is removed.

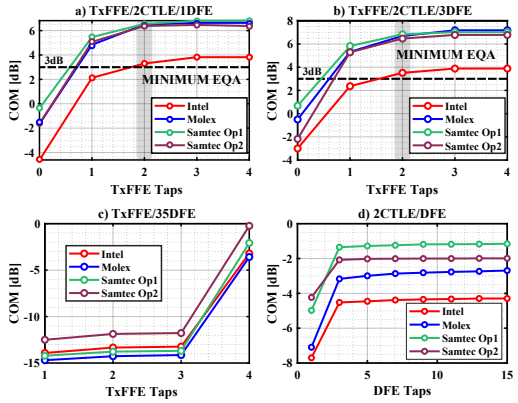


Figure 7. COM values for four common EQAs with a parametric sweep of TxFFE and DFE taps, including all contributions from the crosstalk channel and two CTLEs. The complexity of the equalizer circuit increases from left to right on the x-axis.
Source: Authors

In both case studies, the results confirm our initial hypothesis that the Molex channel requires a more complex transceiver circuit and yields the lowest COM results. Figures 6 and 7 demonstrate that, for all EQAs in both case studies, the COM results for the Molex channel resemble those obtained using the Samtec channels with the lowest estimated losses. This example highlights how COM can prevent the transceiver over-design resulting from a misinterpretation of channel losses using traditional validation metrics. Unlike validation metrics that drive transceiver design methodologies, COM can serve as a parallel design tool to validate design decisions and experimental EQAs, thus reducing chip over-design.

DFE IIR strategy in COM

Figure 8 shows the block diagram and the equalization process for a discrete 1-tap DFE applied 1/2 UI before and after each post-cursor. A DFE comprises a slicer that makes a symbol decision without amplifying the noise and a feedback loop with weighted derivation coefficients (H1, H2, through HN). Those coefficients are added to the input, canceling the post-cursor ISI. DFE implementations often employ up to ten feedback taps to achieve high-loss channel equalization. Unfortunately, the large number of feedback circuits used in a multi-tap DFE consumes significant power and chip area.

Some works (Shahramian and Chan Carusone, 2015; Chen *et al.*, 2017; Elhadidy and Palermo, 2013; Kim *et al.*, 2009; Shahramian *et al.*, 2016) explore power reduction in DFEs with multiple taps. These works propose employing IIR filters to create continuous-time taps, enabling the

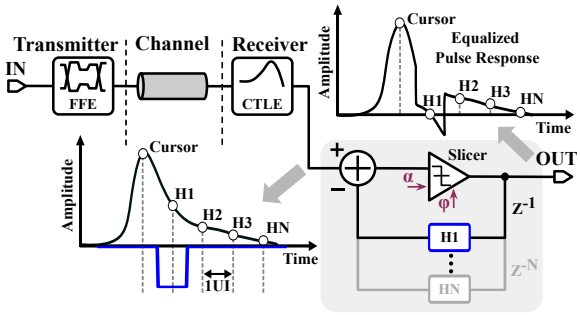


Figure 8. A discrete DFE architecture of N taps
Source: Authors

simultaneous cancellation of several post-cursor ISI taps and achieving a reduction in power consumption and circuit complexity. EQAs that utilize DFEs with IIR taps demonstrate enhanced equalization efficiency for backplane channels (Elhadidy and Palermo, 2013)

This section reviews state-of-the-art DFE architectures with IIR filters and analyzes their impact on the COM algorithm. Furthermore, a discussion of COM results for various EQAs is presented.

DFE IIR prior art

Channels with high post-cursor ISI require multi-tap DFEs, increasing chip area and power consumption. However, the ISI can be represented by a decreasing exponential in the SBR. Some studies present DFE architectures using IIR filters in the feedback loop, canceling multiple post-cursor samples with a single tap and reducing power and area consumption (Shahramian and Chan Carusone, 2015; Chen *et al.*, 2017; Elhadidy and Palermo, 2013; Kim *et al.*, 2009; Shahramian *et al.*, 2016). Like a traditional DFE, IIR DFE cancels post-cursor ISI without noise amplification. Figure 9 shows the diagram of a DFE IIR with a single feedback loop that efficiently eliminates post-cursor ISI of up to three UI compared to a traditional discrete DFE.

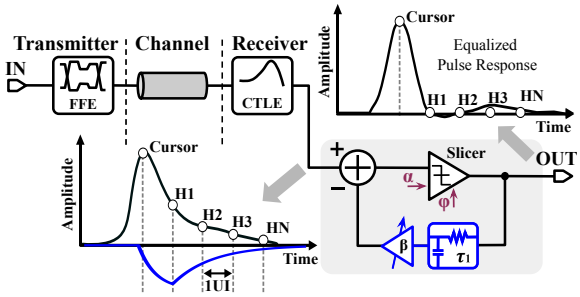


Figure 9. A DFE IIR architecture
Source: Authors

Figure 10 compares three DFE architectures using IIR filters. The first two are pure DFE IIR architectures, while the third combines one discrete tap with two IIR filters. The DFE IIR architecture in Figure 10a compensates for up to two post-cursor samples with low complexity. However, one IIR filter is insufficient for complete post-cursor ISI elimination, as ISI is only partially eliminated after the third tap. The DFE with two IIR architectures in Figure 10b provides better equalization by covering fast and slow decaying ISI tails. However, implementing a DFE IIR can degrade signal integrity as the feedback loop delay increases.

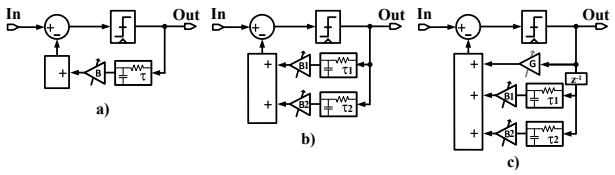


Figure 10. Architectures for link DFE IIR feedback equalization: a) one IIR in the first tap; b) two IIRs in the first tap; c) one discrete tap and two IIRs in the second tap **Source:** Authors

As a solution to delay degradation, Shahramian and Chan Carusone (2015) propose a DFE architecture that combines one discrete-time tap with IIR filters, as depicted in Figure 10c. This architecture achieves a better fit for ISI because the first discrete tap decays faster than the other post-cursor samples. However, it is important to carefully select the number of implemented IIR filters to avoid excessive power and area consumption.

This study integrates two decision feedback equalizers into the COM algorithm (Figure 10), the first with 1-IIR and the other with 2-IIR.

Challenges of implementing DFE IIR in COM

Figure 11 compares different DFE architectures: Figure 11a shows a DFE with only IIR filters in the first tap, while Figure 11b combines a discrete branch with IIR filters. The two filters, VIIR1 and VIIR2, have two different time constants, one for the fast-decaying (VIIR1 in blue) and the other for the slow-decaying (VIIR2 in red).

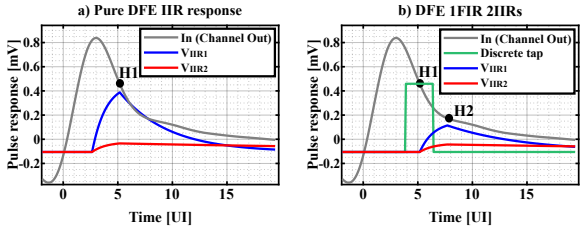


Figure 11. Behavior of DFE architectures implemented in COM: a) DFE with 2IIR in the first taps; b) DFE with one discrete tap along with two IIRs in the second tap
Source: Authors

To implement a DFE architecture with IIR filters in the COM algorithm, the impulse response of the filter was defined using Equation 5. This equation captures the decaying exponential behavior characteristic of the post-cursor ISI. The IIR filter is represented by two terms, the first term ($0 \leq t \leq t1$) models the increasing behavior of the impulse response in one UI. The second term captures the decreasing behavior of the post-cursor long-tail ISI. The constants RC and RC2 correspond to the increasing and decreasing time constants, while A and B represent the SBR values at t = 0 and t = t1, respectively.

$$V_{IIR}(t) = \begin{cases} A(1 - e^{-t/RC}) & 0 \leq t \leq t1 \\ B(e^{-t/RC2}) & t > t1 \end{cases} \quad (5)$$

Using Equation 5, the COM algorithm was modified regarding two key points:

- **FoM maximization loop.** In the FoM loop, we replaced the process of DFE coefficients calculation with Equation 5, operating only the post-cursor samples of the SBR. We also modified the computation of sigma ISI to consider the residual post-cursor taps from SBR after DFE IIR effects instead of the direct subtraction of traditional DFE taps to SBR.
- **Total noise contribution.** Once the FoM optimization process ends, the total noise contribution A_{ni} used for the COM calculation is estimated as was shown in the previous sections. Figure 12 presents the process of obtaining the ISI noise contribution with a DFE with one discrete tap. The DFE response in blue is applied to the SBR in gray, resulting in the residual response used to compute the total ISI noise. Nevertheless, the DFE implemented in COM includes floating DFE taps highlighted in gray, which add extra compensation for the other post-cursor samples. In addition to extra compensation, the cursor sample is eliminated to consider only the residual response after the complete equalization process. Due to the high impact of floating taps in the COM results (up to 1 dB difference, as presented in Figure 11), the COM results obtained with traditional discrete DFE are recalculated to properly compare the different new DFE architectures.

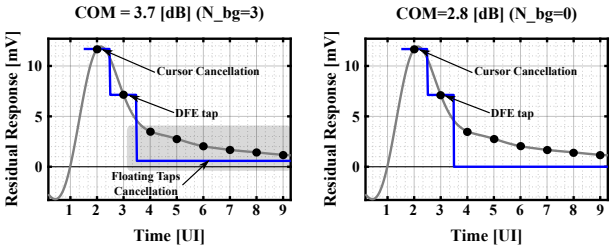


Figure 12. Residual response and COM results using a) DFE floating taps b) no DFE floating taps **Source:** Authors

The results are classified into two different groups. Each group comprises four EQAs with a variation of the DFE architecture, along with an FFE with five taps (c(1), c(0), c(-1), c(-2), c(-3)) and a CTLE without peaking gain modifications. The first group includes a pure DFE IIR with up to two IIR filters on the first tap. The second group uses a DFE with one discrete tap along with one and two IIR filters in the second tap. The COM results were obtained by implementing the same testbench and the setup defined in Table 1. Additionally, we compared the results obtained with the different DFE modifications against an equivalent architecture that uses the traditional discrete DFE. However, in this section, the COM results are recalculated with a traditional DFE architecture due to, now the DFE does not include floating taps, causing a significant COM reduction regarding the previous results.

Simulation results

We selected twelve different EQA combinations and grouped them, as previously mentioned. The main objective of this Section is to determine the least complex solution in which COM achieves the margin. Figure 13a presents the impact of a pure DFE IIR along with different FFE and CTLE

combinations *vs.* the final value of the COM. Since the IIRs are in the first tap, the COM results obtained from equivalent EQAs with a 1-tap DFE are presented in Figure 13b. The x-axis corresponds to different channel models, showing all the results for each channel.

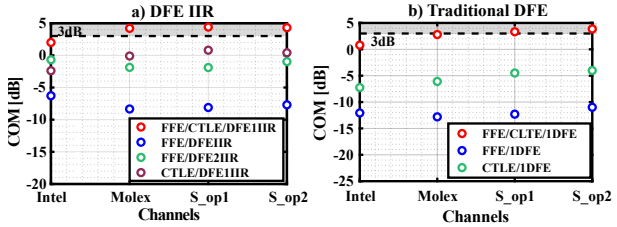


Figure 13. COM performance of four different EQAs, including all crosstalk channel contributions: a) EQA with a DFE with one and two IIR taps; b) commonly used EAQs
Source: Authors

It is relevant to note that, in Figure 13a, the EQA in green is the only one that uses two IIR filters. This consideration is since most of the long ISI tail is compensated by the CTLE, requiring a DFE with few taps. Therefore, for EQA with CTLE, the use of more than one IIR is unnecessary. This effect was evidenced in Figures 6 and 7, where the COM result showed saturation after 3-taps DFE for EQA including the CTLE.

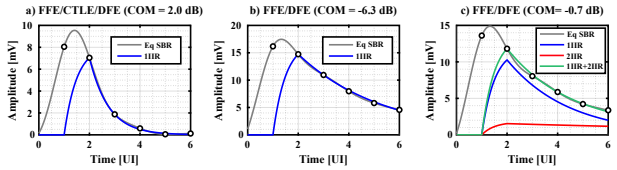


Figure 14. DFE IIR response for different Intel SBRs: a) DFE with an IIR for a fast-decaying SBR; b) DFE with an IIR for a slow-decay SBR; c) DFE with two IIR fast IIR (blue) and slow IIR (red)
Source: Authors

Replacing a traditional DFE with a pure DFE yields an improvement in the COM value, particularly when EQAs have only two equalizers. However, the transceiver architecture that achieves a 3 dB COM is equally complex to the one discussed in the previous section. Moreover, none of the EQAs can compensate for the high losses of the Intel channel. Figure 14 illustrates the designed DFE IIR function, which perfectly fits the ISI of the SBR. In the case with two IIRs (Figure 14c), the DFE response corresponds to the sum of the fast-decaying and the slow-decaying IIRs, canceling up to six post-cursor taps.

In a second equalization experiment, we selected four different EQAs, combining a discrete tap with IIR taps in the DFE architecture. It has been demonstrated (Shahramian and Chan Carusone, 2015) that this combination achieves a better fit of ISI. Therefore, using this DFE architecture, the COM results might exceed the 3 dB threshold with just two equalizers.

Figure 15 presents the COM results, including crosstalk channel contributions. The EQA without CTLE and with a DFE comprising one and two IIR taps is the only one with two IIRs. The other EQAs in Figure 15b include a 2-tap DFE. While the DFE architecture combining discrete and continuous taps promises to better mitigate delay degradations and adjust post-cursor ISI, the COM results in

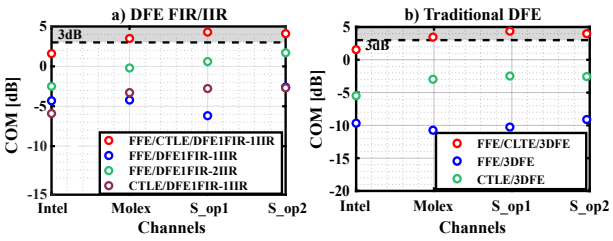


Figure 15. COM performance of four different EQAs, including all crosstalk channel contributions: a) EQA with a DFE with one and two IIR taps; b) commonly used EAQs
Source: Authors

Figure 15a were not significantly better than those obtained with a pure DFE IIR.

By comparing the DFE responses in Figures 14 and 16, a perfect fit canceling more than six post-cursor taps can be observed in both cases. Still, similarly to the results presented in the previous chapter, only one of the equalization architectures surpasses the 3 dB COM threshold. In Figures 13 and 15, however, there is an improvement in the COM with the DFE IIR. By comparing the results obtained using the architecture with CTLE+DFE IIR and CTLE+1DFE in Figure 13, we observed an improvement of around 5 dB in the COM result using DFE IIR *vs.* that with the traditional DFE.

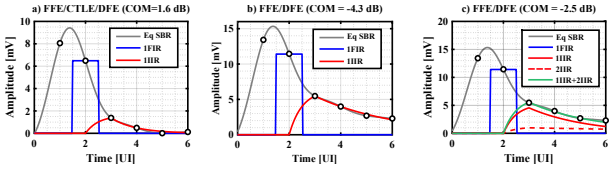


Figure 16. Discrete DFE and DFE IIR response for different Intel SBRs: a) DFE with one FIR tap and one IIR for a fast-decaying SBR; b) DFE with one FIR tap and one IIR for a slow decay SBR; c) DFE with one FIR tap and one IIR, fast IIR (red), and slow IIR (dotted red)
Source: Authors

Figures 17 and 18 show the eye diagrams of the Intel channel equalized by different EQAs. While increasing the number of IIR filters in the DFE leads to a drastic improvement in the eye pattern, using two IIR filters achieves near-optimal performance while maintaining a low system complexity.

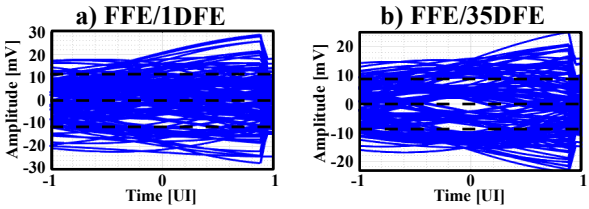


Figure 17. Simulated eye diagrams for an EQA with an FFE and a discrete DFE: a) 1-tap, b) 35-taps.
Source: Authors

It is also clear that, in some cases, the additional discrete DFE tap produces minuscule improvements in the eye-opening, and it can be avoided to reduce complexity. However, canceling the first ISI after the cursor when using only IIR filters can become more complex as the data rate increases due to delays in the feedback loop (Kim *et al.*, 2009; Shahramian *et al.*, 2016). Therefore, adding one discrete

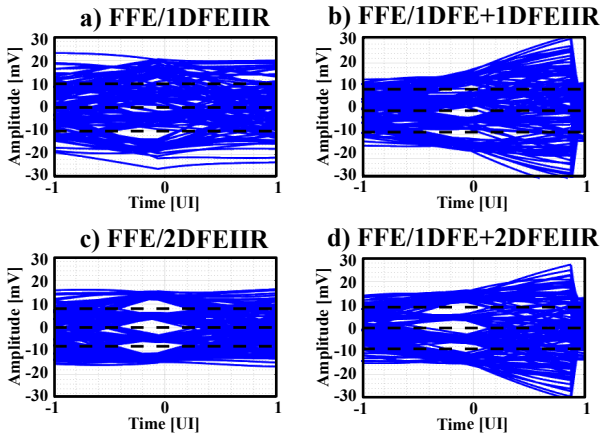


Figure 18. Simulated eye diagrams for an EQA with an FFE and a DFE: a) 1-IIR tap, b) 1-discrete tap along with a 1-IIR tap, c) 2-IIR, d) one discrete tap along with 2-IIR taps
Source: Authors

tap to the first feedback loop cancels the first post-course sample and makes it more robust to timing issues.

Conclusions

The COM is a channel validation metric that provides an estimate of the minimum complexity of a SerDes system. However, the equalization and data recovery circuit (CDR) models used in COM are idealized, disregarding noise sources such as parasitic capacitances and sampler jitter in the calculation. When these contributions are included, the total noise can significantly increase, potentially preventing COM from achieving the 3 dB margin for the defined equalizer parameters and IEEE 182.3ck conditions. This necessitates a reevaluation of the COM algorithm and the equalization models.

COM results demonstrate that traditional equalization architectures (FFE in Tx and CTLE and DFE in Rx) can achieve a COM > 3 dB, aligning with the findings of Peng *et al.* (2022) and Lee *et al.* (2015). However, integrating new equalization schemes into the COM framework allows for exploring multiple equalization proposals at an early design stage. Although the COM results for the DFE architecture with IIR taps were not favorable in this study, the eye diagram exhibits an ISI-free region, supporting the results of Roshan-Zamir *et al.* (2017).

In conclusion, COM has demonstrated its flexibility and efficiency, albeit with implementation and modification challenges. While it offers insights into transceiver architecture complexity, its sequential algorithm structure hinders the seamless integration and modification of equalization architectures, resulting in increased time costs. Therefore, as a transceiver design tool, COM is still in its early stages, indicating the need to restructure the FoM optimization cycle and incorporate EQAs with more complex models into the algorithm.

Author Contributions

All authors made significant contributions to the research and provided critical comments. Luisa Fernanda Dovale-Vargas contributed to the background research, developing the method, obtaining the simulation results, writing the manuscript, and discussing the results. Oscar Mauricio Reyes-Torres contributed to obtaining the simulation results and collecting data. He supervised the research and led it throughout the process, including the background research and the writing process. Élkim Felipe Roa-Fuentes conceived the idea, provided guidance during the research and writing of the article, and supervised the entire research process.

References

Anritsu Company. (2016). *White Paper: Measuring Channel Operating Margin*. https://dl.cdn-anritsu.com/en-us/test-measurement/files/Technical-Notes/White-Paper/11410-00989A.pdf.

Champion, B., and Tracy, N. (2019). *Delivering 100 Gbps Solutions for Chip-to-Module and Direct Attach Copper (DAC) Cable Implementations*. https://www.te.com/content/dam/te-com/documents/datacomm/global/delivering-100-gbps-solutions-for-chip-paper.pdf.

Chen, K.-Y., Chen, W.-Y., and Liu, S.-I. (2017). A 0.31-pJ/bit 20-Gb/s DFE With 1 Discrete Tap and 2 IIR Filters Feedback in 40-nm-LP CMOS. *IEEE Transactions on Circuits and Systems II: Express Briefs*, *64*(11), 1282-1286. https://doi.org/10.1109/TCSII.2016.2645698.

de Paulis, F., Rabinovich, R., Mellitz, R., and Resso, M. (2022). Feasibility analysis of chip-to-module channel operating at 212 gbps. In *2022 ieee 26th workshop on signal and power integrity (spi)* (p. 1-4). https://doi.org/10.1109/SPI54345.2022.9874926.

de Paulis, F., Wang-Lee, T., Mellitz, R., Resso, M., Rabinovich, R., and Danzy, O. (2020). Backplane channel design exploration at 112 gbps using channel operating margin (com). In *2020 ieee international symposium on electromagnetic compatibility and signal/power integrity (emcsi)* (p. 158-163). https://doi.org/10.1109/EMCSI38923.2020.9191569.

Dong, X., Mo, M., Raoand W. Jin., F., and G, Z. (2014). Relating COM to Familiar S-parameter Parametric to Assist 25Gbps System Design. *DesignCon 2014*. https://www.xilinx.com/content/dam/xilinx/publications/events/designcon/2014/8_TU2Slides_RelatingCOMtoFamiliarS_ParameterParametrictoAssist25GbpsSystemDesign.pdf.

Dovale, L., and Roa, E. (2021). In *2021 IEEE International Symposium on Circuits and Systems (ISCAS), title=Channel Operating Margin as Transceiver Architecture Design Tool* (p. 1-5). https://doi.org/10.1109/ISCAS51556.2021.9401322.

Elhadidy, O., and Palermo, S. (2013). A 10 gb/s 2-iir-tap dfe receiver with 35 db loss compensation in 65-nm cmos. In *2013 symposium on vlsi circuits* (p. C272-C273).

Filip, C. (2018). Effective return loss, a new metric for serdes channel or package characterization. *Mentor, A Siemens Business*.

Gore, B., and Mellitz, R. (2014). An excerise in applying channel operating margin (COM) for 10GBASE-KR channel design. In *2014 IEEE International Symposium on Electromagnetic Compatibility (EMC)* (p. 648-653). https://doi.org/10.1109/ISEMC.2014.6899050.

Heck, H., and Intel. (2018). *Server Backplane Channels*. https://www.ieee802.org/3/ck/public/18_11/heck_3ck_01_1118.pdf.

IEEE. (2014). IEEE Standard for Ethernet Amendment 2: Physical Layer Specifications and Management Parameters for 100 Gb/s Operation Over Backplanes and Copper Cables.

IEEE Std 802.3bj-2014 (*Amendment to IEEE Std 802.3-2012 as amended by IEEE Std 802.3bk-2013*), 1-368.

Kim, B., Liu, Y., Dickson, T. O., Bulzacchelli, J. F., and Friedman, D. J. (2009). A 10-gb/s compact low-power serial i/o with dfe-iir equalization in 65-nm cmos. *IEEE Journal of Solid-State Circuits*, *44*(12), 3526-3538. https://doi.org/10.1109/JSSC.2009.2031015.

Lee, J., Chiang, P.-C., Peng, P.-J., Chen, L.-Y., and Weng, C.-C. (2015). Design of 56 gb/s nrz and pam4 serdes transceivers in cmos technologies. *IEEE Journal of Solid-State Circuits*, *50*(9), 2061-2073. https://doi.org/10.1109/JSSC.2015.2433269.

Mellitz, R., Abbott, J., and Samtec. (2018). *112G Cabled Backplane Channel and PCB Design Impact Using 112G Ready Connectors*. https://www.ieee802.org/3/ck/public/adhoc/aug15_18/mellitz_3ck_adhoc_02_081518.pdf.

Mueller, K., and Muller, M. (1976). Timing recovery in digital synchronous data receivers. *IEEE Transactions on Communications*, *24*(5), 516-531. https://doi.org/10.1109/TCOM.1976.1093326.

Palkert, T., and Molex. (2020). *OD Backplane Channel Analysis*. https://www.ieee802.org/3/ck/public/20_01/palkert_3ck_01a_0120.pdf.

Peng, P.-J., Lee, P.-L., Huang, H.-E., Huang, W.-J., Lin, M.-W., Juang, Y.-Z., and Tseng, S.-H. (2022). A 56-gb/s pam-4 transmitter/receiver chipset with nonlinear ffe for vcsel-based optical links in 40-nm cmos. *IEEE Journal of Solid-State Circuits*, *57*(10), 3025-3035. https://doi.org/10.1109/JSSC.2022.3192711.

Roshan-Zamir, A., Elhadidy, O., Yang, H.-W., and Palermo, S. (2017). A reconfigurable 16/32 gb/s dual-mode nrz/pam4 serdes in 65-nm cmos. *IEEE Journal of Solid-State Circuits*, *52*(9), 2430-2447. https://doi.org/10.1109/JSSC.2017.2705070.

Shahramian, S., and Chan Carusone, A. (2015). A 0.41 pj/bit 10 gb/s hybrid 2 iir and 1 discrete-time dfe tap in 28 nm-lp cmos. *IEEE Journal of Solid-State Circuits*, *50*(7), 1722-1735. https://doi.org/10.1109/JSSC.2015.2402218.

Shahramian, S., Dehlaghi, B., and Chan Carusone, A. (2016). Edge-based adaptation for a 1 iir + 1 discrete-time tap dfe converging in 5 μ s. *IEEE Journal of Solid-State Circuits*, *51*(12), 3192-3203. https://doi.org/10.1109/JSSC.2016.2594209.

Research Article/Electrical Engineering

http://doi.org/10.15446/ing.investig.110310

Applying the Sine-Cosine Optimization Algorithm to the Parametric Estimation Problem in Three-Phase Induction Motors

Aplicación del algoritmo de optimización por senos y cosenos al problema de estimación paramétrica en motores de inducción trifásicos

Santos Daniel Niño-Callejas[1], Juan Camilo Palombi-Gómez[2], and Oscar Danilo Montoya[3]

ABSTRACT

The steady-state analysis of electrical machines requires a detailed characterization of their equivalent electrical circuit, which adequately represents the transformation and interaction between electrical and mechanical energy. This research aims to characterize the equivalent circuit of three-phase induction motors by minimizing the mean square error between the measured and calculated torque variables. These torques are obtained from data provided by the manufacturer, including starting, peak, and full-load torques. A metaheuristic optimization technique is applied to solve the resulting nonlinear programming model based on the interactions between the sine and cosine functions. The numerical results obtained with this algorithm demonstrate its efficiency in terms of response quality, reaching objective function values of less than 1×10^{-8} with regard to the measured and calculated variables. Simulation results in two test systems allow concluding that the parametric estimation problem in three-phase induction motors is a multimodal optimization problem. This implies a potentially infinite set of solutions that minimize the root mean square error and adequately represent the behavior of the motor's output torque under various probable operating conditions.

Keywords: metaheuristic optimization, electrical circuit characterization, multimodal optimization problem, manufacturer data

RESUMEN

El análisis del estado estacionario de las máquinas eléctricas requiere una caracterización detallada de su circuito eléctrico equivalente que represente adecuadamente la transformación y la interacción entre energía eléctrica y mecánica. El objetivo de esta investigación es caracterizar el circuito equivalente de motores de inducción trifásicos mediante la minimización del error cuadrático medio entre variables de torque medidas y calculadas. Estos torques se obtienen de datos suministrados por el fabricante, incluyendo los torques inicial, máximo y de carga plena. Se aplica una técnica de optimización metaheurística para resolver el modelo de programación no linear resultante, que se basa en las interacciones entre las funciones de seno y coseno. Los resultados numéricos obtenidos con este algoritmo demuestran su eficiencia en términos de calidad de la respuesta, alcanzando valores de función objetivo de menos de 1×10^{-8} respecto a las variables medidas y calculadas. Los resultados de simulaciones realizadas en dos sistemas de prueba permiten concluir que el problema de estimación paramétrica en motores de inducción trifásicos es un problema de optimización multimodal. Esto implica un conjunto de soluciones potencialmente infinitas que minimizan el eror cuadrático medio y representan adecuadamente el torque de salida del motor en varias condiciones probables de operación.

Palabras clave: optimización metaheurística, caracterización de circuitos eléctricos, problema de optimización multimodal, datos del fabricante

Received: July 14th 2023
Accepted: October 24th 2023

Nomenclature

Acronyms

GA	Genetic algorithm
HGAPSO	Hybrid between the GA and the PSO
PSO	Particle swarm optimizer
SCA	Sine-cosine algorithm

Functions

$\tau_{\max}$	Maximum torque (N.m)
τ_{fl}	full-load torque (N.m)
τ_{ind}	Induced torque (N.m)
τ_{st}	Starting torque (N.m)
E_f	Objective function aimed at minimizing the mean square error
R_{th}	Thevenin resistance (Ω)
V_{th}	Thevenin voltage (V)
X_{th}	Thevenin inductance (Ω)

Parameters

[1]Electrical engineering student, Universidad Distrital Francisco José de Caldas, Colombia. E-mail: sadninoc@udistrital.edu.co
[2]Electrical engineering student, Universidad Distrital Francisco José de Caldas, Colombia. E-mail: jcpalombig@udistrital.edu.co
[3]Electical engineer, Universidad Tecnológica de Pereira, Colombia. MSc in Electrical Enginerring, Universidad Tecnológica de Pereira, Colombia. PhD in Enginering (Electrical Engineering), Universidad Tecnológica de Pereira, Colombia. Affiliation: Assitant professor, Universidad Distrital Francisco José de Caldas, Colombia. E-mail: odmontoyag@udistrital.edu.co

ω_{sinc} **Synchronous speed (rad/s)**

τ^m_{max} **Manufacturer data: maximum torque (N.m)**

τ^m_{fl} **Manufacturer data: full-load torque (N.m)**

τ_{ind} **Induced torque (N.m)**

τ^m_{st} **Manufacturer data: starting torque (N.m)**

j **Complex unit, *i.e.*,** $j = \sqrt{-1}$

k_{max} **Parameter that stops the search if the objective function does not consecutively improve during** 20% **of the iterations**

N_s **Population size for the SCA**

R_1^{max} **Maximum value for the stator resistance** (Ω)

R_1^{min} **Minimum value for the stator resistance** (Ω)

r_2 **Decreasing coefficient associated with the importance of the sine and cosine functions during the exploration and exploitation stages**

R_2^{max} **Maximum value for the rotor resistance** (Ω)

R_2^{min} **Minimum value for the rotor resistance** (Ω)

r_3 **Random number between** 0 **and** 2π **with a uniform distribution**

r_4 **Random number between** 0 **and** 1 **with a uniform distribution**

s **Sliding coefficient (%)**

t **Iterative counter**

t_{max} **Maximum number of iterations**

V_{ph} **Single-phase voltage applied to the induction motor (V)**

X_1^{max} **Maximum value for the stator reactance** (Ω)

X_1^{min} **Minimum value for the stator reactance** (Ω)

X_2^{max} **Maximum value for the rotor reactance** (Ω)

X_2^{min} **Minimum value for the rotor reactance** (Ω)

X_M^{max} **Maximum value for the magnetizing reactance** (Ω)

X_M^{min} **Minimum value for the magnetizing reactance** (Ω)

Sets and indices

i **Position of the solution in the matrix of potential solutions**

Variables

R_1 **Resistive effect assigned to the windings of the stator in the induction motor (V)**

R_2 **Resistive effect assigned to the windings of the rotor in the induction motor** (Ω)

$s_{i,1}^{t+1}$ **Potential solution for the next iteration (*i.e.*,** $t+1$**), using the sine rule**

$s_{i,2}^{t+1}$ **Potential solution for the next iteration (*i.e.*,** $t+1$**), using the cosine rule**

s_i^t **Vector containing the** i^{th} **potential solution at iteration** t

X_1 **Reactance effect assigned to the windings of the stator in the induction motor** (Ω)

X_2 **Reactance effect assigned to the windings of the rotor in the induction motor** (Ω)

X_M **Equivalent magnetizing reactance** (Ω)

Introduction

General context

The processes related to the final use of electrical energy have gained significant relevance in terms of both operation and costs (Chauhan, Chauhan, and Badar, 2022). Therefore, energy efficiency plays a fundamental role in new industrial developments (Bouakkaz, Mena, Haddad, and Ferrari, 2021). This has been demonstrated multiple times through the increasing rate of vehicles that require less fuel and the growing efficiency of appliances (lower electricity consumption), *e.g.*, lamps that consume a quarter of the energy compared to classic lighting with incandescent bulbs (Nota, Nota, Peluso, and Lazo, 2020; Abo-Khalil *et al.*, 2022).

At the industrial level, one of the most common elements corresponds to induction motors, which are present in all economic sectors and are considered to be the cornerstone of the modern industry (Sengamalai *et al.*, 2022). It is estimated that engines use 65% of the electricity generated in the world, so they can contribute greatly to reducing energy consumption and CO_2 emissions (Payán, Fernandez, Ortega, and Santos, 2019). On the other hand, environmental policies based on energy efficiency, together with those related to the energy transition, comprise a large number of actions aimed at reducing global warming; the less energy is used, the fewer pollutants associated with the energy sector will be produced (Friederici, 2021).

Motivation

In order to contribute to the study of induction motors, large-scale industries with hundreds of engines must be analyzed. Here, due to intensive use, multiple internal parameters related to efficiency calculations can change over time (Trisha, Gupta, and Kumar, 2021). To update the internal parameters of induction motors (*i.e.*, series and parallel reactances), the specialized literature has proposed multiple optimization approaches that avoid physical interventions and favor classical laboratory tests (Véliz-Tejo, Travieso-Torres, Peters, Mora, and Leiva-Silva, 2022). These optimization algorithms focus on torque measurements at the terminals of the induction motor for different load conditions. With these measurements, a nonlinear non-convex optimization model has been proposed which minimizes the expected error between the calculated and the measured torques (Mohammadi and Akhavan, 2014).

This research proposes the application of the sine-cosine algorithm (SCA) to obtain the internal parameters of induction motors by solving the equivalent optimization model, since metaheuristic optimizers have demonstrated to be efficient and robust in solving parametric estimation problems associated with electrical devices, *i.e.*, induction machines, distribution transformers, and photovoltaic (PV) modules, among others.

Literature review

Although there are many methods to estimate the parameters of induction motors, iterative methods based on least squares are the most widespread, given their simplicity and reasonable convergence times (Lindenmeyer, Dommel, Moshref, and Kundur, 2001; Toliyat, Levi, and Raina, 2003; Pedra and Corcoles, 2004; Gupta, Wadhwani, and Kapoor, 2011). In addition, multiple metaheuristic optimization algorithms have been applied to solve the nonlinear non-convex optimization problem regarding parametric estimation in electrical machines, as they are easily programmable and require low computational efforts. These algorithms include bee colony optimization (Aminu, 2019), particle swarm optimization (Huynh and Dunnigan, 2010), the gravitational search algorithm (Avalos, Cuevas, and Gálvez, 2016), and the water cycle algorithm (Ćalasan, Micev, Ali, Zobaa, and Aleem, 2020), among others. Some of the most recent applications belonging to the family of metaheuristic optimization algorithms for parametric estimation in induction motors are discussed below.

(Mohammadi and Akhavan, 2014) combined the classical genetic algorithm and the particle swarm optimizer to obtain a hybrid optimization approach aimed at determining the electrical parameters of three-phase induction machines. The optimization problem was formulated as a nonlinear programming model, where the mean square error between manufacturer data and the calculated starting, maximum, and full-load torques was considered as the objective function. Numerical results confirmed that this hybrid approach provides better numerical results regarding the objective function value when compared to the genetic algorithm and particle swarm optimization.

The work by (Wu, Tseng, and Chen, 2018) employed the polynomial regression approach to determine the electrical parameters of induction machines. Numerical results provided accurate estimations when compared to the experimental motor curve and the procedure established by the IEEE Standard 112 Test.

In (Guedes, Castoldi, Goedtel, Agulhari, and Sanches, 2018), the authors applied the differential evolution algorithm to estimate the electrical parameters of three-phase induction motors via dynamic simulation. Theoretical results with the proposed optimization algorithm and experimental validations confirmed the accuracy of this approach. However, the authors provided no comparative analysis with additional optimization algorithms.

The study by (Rezk, Elghany, Al-Dhaifallah, Sayed, and Ibrahim, 2019) presented an effective optimization method to estimate parameters in three-phase induction motors. The particle swarm optimization algorithm was used in combination with experimental validations, with the purpose of reducing the error between the theoretical model and the experimental setup. Numerical results confirmed that the particle swarm optimizer yields acceptable results. However, comparative analyses with other metaheuristic optimizers were not presented to demonstrate the effectiveness of the proposed approach.

(Joodaki, Shojaei, and Lotfi, 2020) applied the cuckoo search algorithm to estimate parameters in three-phase induction motors while considering manufacturer torque data. Numerical comparisons in two machines with a genetic algorithm, the water cycle algorithm, and the bacterial proliferation approach, among others, confirmed the effectiveness of the cuckoo search algorithm in minimizing the error between the measured and the calculated torques.

The main characteristics of the approaches mentioned above are the following: (i) the manufacturer data regarding the starting, maximum, and full-load torques are typically used to formulate a minimization problem aimed at finding the electrical parameters of the motor, which makes the theoretical and the calculated torques equal; and (ii), given the complexity of the optimization model, most studies focus on the application of metaheuristic optimization algorithms to obtain an efficient solution.

Contributions and scope

In light of the above, this research presents the application of the sine-cosine algorithm (SCA) as an efficient optimization technique to estimate the parameters of the electrical circuits in induction motors. The SCA is implemented to solve the exact nonlinear non-convex optimization model for minimizing the error between measured and calculated torques under different load conditions (*i.e.*, starting, full, and maximum load torques). This algorithm was identified as a potential solution methodology because it has already demonstrated its effectiveness and robustness in solving similar problems, as is the case of parametric estimation in single-phase transformers and PV modules (Bocanegra, Montoya, and Molina, 2021; Montoya, Gil-González, and Grisales-Noreña, 2020).

Regarding the scope of this research, note that all measurements of the induction machine have been taken with specialized torque measurement systems, which have been reviewed and conditioned (filtered) prior to their evaluation in the proposed SCA. These data have also been provided by the manufacturer of the induction machine. In addition, a comparison between different metaheuristic optimizers (the genetic algorithm, the particle swarm optimizer, and a combination of the two) has been performed to validate our proposal.

Document structure

The remainder of this document is structured as follows. Section Mathematical formulation presents the general mathematical formulation associated with parametric estimation in three-phase induction motors while considering torque measurements. Section Solution methodology describes the general characteristics of the SCA and its application to the analyzed problem. Section Test systems outlines the main characteristics of the two induction machines examined with regard to their

manufacturer data and the upper and lower bounds imposed on the decision variables. Section Numerical results and discussion shows the numerical validations and a comparison between the SCA, the genetic algorithm (GA), the particle swarm optimizer (PSO), and a GA-PSO hybrid in the first test system. In addition, the multimodal nature of the optimization model is shown via three solutions provided by the SCA in the second test system. Finally, Section Conclusions and future work lists the main concluding remarks of this research and possible future studies using metaheuristic optimization methods.

Mathematical formulation

The parametric estimation problem in three-phase induction motors is formulated as a nonlinear non-convex optimization problem using the Thevenin equivalent of these machines under steady-state conditions. Figure 1 presents the equivalent circuit representation of the induction motor.

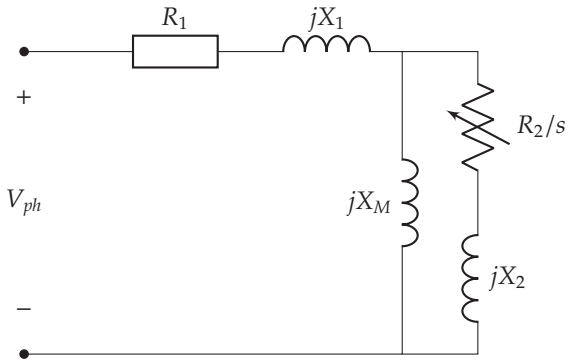


Figure 1. Single-phase equivalent of the induction motor
Source: Authors

In Figure 1, R_1 and R_2 denote the resistive effects on the stator and rotor elements, X_1 and X_2 correspond to the reactance equivalent values of the stator and rotor windings, X_M is the equivalent magnetization reactance, V_{ph} means the single-phase voltage applied to the induction motor, and s is the sliding under normal operating conditions. Note that j is the imaginary unit.

To obtain the equivalent Thevenin representation of the induction motor in Figure 1, consider that the motor load (*i.e.*, R_2/s) is removed. Under this condition, the equivalent V_{th}, R_{th}, and X_{th} are obtained.

$$V_{th} = \frac{X_M}{X_1 + X_M} V_{ph}, \tag{1}$$

$$R_{th} = \frac{X_M R_1}{X_1 + X_M}, \tag{2}$$

$$X_{th} = \frac{X_M X_1}{X_1 + X_M}, \tag{3}$$

Remark 1. *Note that, in order to obtain* (1)*, it was assumed that* $X_M >> R_1$*, which implies that the effect of* R_1 *can be neglected in the Thevenin voltage calculation and equivalent impedance* (*Mohammadi and Akhavan, 2014*).

Figure 2 depicts the equivalent Thevenin circuit representing the induction motor per phase shown in Figure 1.

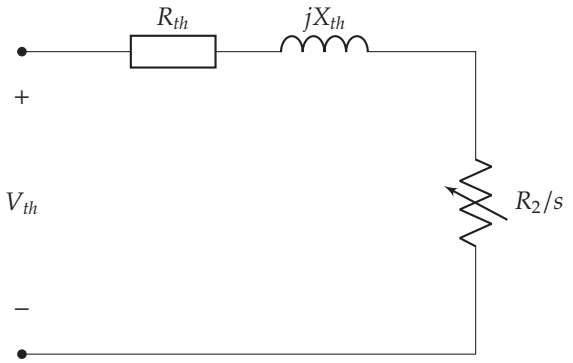


Figure 2. Thevenin equivalent of the single-phase circuit of the induction motor
Source: Authors

In Equation (4), the general induced torque is defined.

$$\tau_{ind} = \frac{3V_{th}^2 R_2}{s\omega_{\text{sinc}}\left[\left(R_{th} + \frac{R_2}{s}\right)^2 + (X_{th} + X_2)^2\right]}, \tag{4}$$

where ω_{sinc} is the synchronous speed.

In addition, the starting torque (τ_{st}) occurs when $s = 1$, which implies that it can be defined from (4) as follows:

$$\tau_{st} = \frac{3V_{th}^2 R_2}{\omega_{\text{sinc}}\left[(R_{th} + R_2)^2 + (X_{th} + X_2)^2\right]}, \tag{5}$$

The maximum torque ($\tau_{\max}$) is reached at maximum converted power, which occurs when $\frac{R_2}{s} = \sqrt{R_{th}^2 + (X_{th} + X_2)^2}$. This maximum torque is defined in (6).

$$\tau_{\max} = \frac{3V_{th}^2}{2\omega_{\text{sinc}}\left[R_{th} + \sqrt{R_{th}^2 + (X_{th} + X_2)^2}\right]}, \tag{6}$$

Finally, the full-load torque (τ_{fl}) is reached when $s = s_{fl}$ in (4), which yields

$$\tau_{fl} = \frac{3V_{th}^2 R_2}{s_{fl}\omega_{\text{sinc}}\left[\left(R_{th} + \frac{R_2}{s_{fl}}\right)^2 + (X_{th} + X_2)^2\right]}. \tag{7}$$

Considering the starting, maximum, and full-load torques defined from (5) to (7), the optimization model that includes the calculated and measured torques of an induction motor is formulated, using the sum of the errors as the objective function, which is defined in (8).

$$\min E_f = \left(\frac{\tau_{st} - \tau_{st}^m}{\tau_{st}^m}\right)^2 + \left(\frac{\tau_{\max} - \tau_{\max}^m}{\tau_{\max}^m}\right)^2 + \left(\frac{\tau_{fl} - \tau_{fl}^m}{\tau_{fl}^m}\right)^2, \tag{8}$$

where E_f is the objective function, and τ_{st}^m, $\tau_{\max}^m$, and τ_{fl}^m represent the measured values associated with the starting, maximum, and full-load torques, respectively.

Note that, in order to ensure that the solution to the objective function in (8) is subject to the torque equality constraints (5)–(7), some typical bounds are imposed. These

are associated with the stator and rotor resistances and reactances, including the magnetization reactance. These box-type constraints are listed below.

$$R_1^{\min} \leq R_1 \leq R_1^{\max}, \tag{9}$$
$$X_1^{\min} \leq X_1 \leq X_1^{\max}, \tag{10}$$
$$R_2^{\min} \leq R_2 \leq R_2^{\max}, \tag{11}$$
$$X_2^{\min} \leq X_2 \leq X_2^{\max}, \tag{12}$$
$$X_M^{\min} \leq X_M \leq X_M^{\max}, \tag{13}$$

where $Y^{\min}$ and $Y^{\max}$ represent the lower and upper bounds assigned to the decision variables.

Remark 2. *The problem regarding parametric estimation in three-phase induction motors corresponds to a nonlinear programming model aimed at minimizing the average square error defined in* (8) *and constrained by equalities* (5)–(7)*, which are related to the starting, maximum, and full-load torques* (*calculated*)*, as well as by the box-type constraints* (9)–(13)*, which define the solution space where the decision variables exist* (*i.e., reactances and resistances*).

Solution methodology

This study applies the SCA to solve the optimization problem expressed in (5)–(13). The SCA is a metaheuristic optimization algorithm that belongs to the family of mathematics-inspired metaheuristics. It works from an initial (feasible) population that evolves during the iteration process while following trigonometric rules based on the sine and cosine functions (Attia, Sehiemy, and Hasanien, 2018). The main characteristics of the SCA approach are presented below.

Initial population

The SCA is a population-based optimizer that explores and exploits the solution space from an initial population generated upon the basis of the upper and lower bounds of the decision variables in order to make it 100% feasible. The structure of the solution individual i at iteration t (with $t = 0$) that conforms to the initial population is presented below.

$$s_i^t = [\overbrace{0.1242}^{R_1} \quad \overbrace{0.3589}^{R_2} \quad \overbrace{1.1614}^{X_1} \quad \overbrace{1.3221}^{X_2} \quad \overbrace{39.4618}^{X_M}]$$

where each vector component is obtained as a random value between the lower and upper bounds defined in (9)–(13), following a uniform distribution.

Evolution of the population

The SCA applies evolution rules based on the trigonometric sine and cosine functions in order to explore the solution space. To determine whether an individual s_i^{t+1} will be part of the population, the following evolution rule is applied:

$$s_{i,1}^{t+1} = s_i^t + r_2 \sin(r_3)\,|r_4 s_{\text{best}}^t - s_i^t|, \;\; i = 1, 2, ..., N_s \tag{14}$$
$$s_{i,2}^{t+1} = s_i^t + r_2 \cos(r_3)\,|r_4 s_{\text{best}}^t - s_i^t|, \;\; i = 1, 2, ..., N_s \tag{15}$$

where $s_{i,1}^{t+1}$ and $s_{i,2}^{t+1}$ are two potential candidate solutions derived from the sine and cosine rules in (14) and (15), r_3 is a random number with a uniform distribution between 0 and 2π, r_4 is a random number between 0 and 1 with a uniform distribution, and N_s is the number of potential individuals that make up the population. Note that s_{best}^t represents the best current solution in the population. In addition, r_1 controls the exploration and exploitation stages of the optimization algorithm by following a linear rule defined below.

$$r_2 = 2\left(1 - \frac{t}{t_{\max}}\right). \tag{16}$$

To determine whether one of the individuals $s_{i,1}^{t+1}$ or $s_{i,2}^{t+1}$ will be part of the next population, the following criteria are applied.

i. If $E_f\left(s_{i,1}^{t+1}\right) < E_f\left(s_{i,2}^{t+1}\right)$ and $E_f\left(s_{i,1}^{t+1}\right) < E_f\left(s_i^t\right)$, then $s_i^{t+1} = s_{i,1}^{t+1}$.

ii. If $E_f\left(s_{i,2}^{t+1}\right) \leq E_f\left(s_{i,1}^{t+1}\right)$ and $E_f\left(s_{i,2}^{t+1}\right) < E_f\left(s_i^t\right)$, then $s_i^{t+1} = s_{i,2}^{t+1}$.

iii. Otherwise, $s_i^{t+1} = s_i^t$.

Remark 3. *To ensure that each potential solution* ($s_{i,1}^{t+1}$ *and* $s_{i,2}^{t+1}$) *is feasible, each of its components is reviewed/corrected, with the purpose of maintaining their values between the lower and upper bounds* (*see box-type constraints* (9)–(13)).

Stopping criteria

One of these stopping criteria must be met to determine whether the SCA has finished exploring and exploiting the solution space.

i. If the number of iterations $t_{\max}$ has been reached, or

ii. if, during $k_{\max}$ iterations, the value of the objective function (*i.e.*, $E_f\left(s_{\text{best}}^t\right)$) has not improved.

Note that $k_{\max}$ is set as 20% of the total number of iterations.

Summary of the SCA

The application of the SCA to the parametric estimation problem in three-phase induction motors while considering torque measurements is summarized in Algorithm 1.

Test systems

Two induction machines were considered to validate the proposed SCA with regard to the studied problem. The first system corresponds to a three-phase induction motor with 5 hp, 460 V, and 60 Hz. Its main characteristics are reported in Table 1, which have been adapted from (Mohammadi and Akhavan, 2014). The second machine is a three-phase induction motor with 25 hp, 460 V, and 60 Hz, whose main characteristics are reported in Table 2.

Algorithm 1: Application of the SCA to estimate parameters in three-phase induction motors

Data: Read data of the three-phase induction motor under analysis
Define the lower and upper bounds of the decision variables;
Define N_s, $t_{\max}$, and $k_{\max}$;
Generate the initial population while observing the lower and upper bounds in (9)–(13);
Evaluate the starting, maximum, and full-load torques in (5), (6), and (7);
Evaluate the objective function in (8);
Find the best current solution s^t_{best};
Make $k = 0$;
for $t = 1 : t_{\max}$ **do**
  Obtain the value of r_1 using (16);
  for $i = 1 : N_s$ **do**
    Obtain the random numbers for r_3 and r_4;
    Generate the potential solutions $s^{t+1}_{i,1}$ and $s^{t+1}_{i,2}$ using Equations (14) and (15);
    Check the feasibility of $s^{t+1}_{i,1}$ and $s^{t+1}_{i,2}$ and correct if necessary;
    Evaluate the starting, maximum, and full-load torques for $s^{t+1}_{i,1}$ and $s^{t+1}_{i,2}$ in (5), (6), and (7);
    Evaluate the objective functions $E_f\left(s^{t+1}_{i,1}\right)$ and $E_f\left(s^{t+1}_{i,2}\right)$;
    if $E_f\left(s^{t+1}_{i,1}\right) \leq E_f\left(s^{t+1}_{i,2}\right)$ & $E_f\left(s^{t+1}_{i,1}\right) \leq E_f\left(s^{t}_{i}\right)$ **then**
      Make $s^{t+1}_{i} = s^{t+1}_{i,1}$;
    else
      if $E_f\left(s^{t+1}_{i,2}\right) \leq E_f\left(s^{t+1}_{i,1}\right)$ & $E_f\left(s^{t+1}_{i,2}\right) \leq E_f\left(s^{t}_{i}\right)$ **then**
        Make $s^{t+1}_{i} = s^{t+1}_{i,2}$;
      else
        Make $s^{t+1}_{i} = s^{t}_{i}$;
      end
    end
  end
  Update the value of the best current solution s^{t+1}_{best};
  if $E_f\left(s^{t+1}_{best}\right) < E_f\left(s^{t}_{best}\right)$ **then**
    Make $k = k + 1$;
  end
  Make $k = 0$;
  if $k \geq k_{\max}$ **then**
    Report the best current solution in s^{t+1}_{best};
    break;
  end
end
Result: Return the best solution found

Table 1. First test system (5 hp, 460 V, and 60 Hz)

Characteristic	Value	Unit
Capacity	5	HP
Rate power	3.7285	kW
Nominal voltage	460	V
Frequency	60	Hz
Number of poles	4	–
Full-load sliding (s_{fl})	0.0210	–
Starting torque (τ_{st})	119.2629	Nm
Maximum torque ($\tau_{\max}$)	149.0820	Nm
Full-load torque (τ_{fl})	19.6730	Nm

Source: Authors

Table 2. Second test system (25 hp, 460 V, and 60 Hz)

Characteristic	Value	Unit
Capacity	25	HP
Rate power	15.54	kW
Nominal voltage	460	V
Frequency	60	Hz
Number of poles	4	–
Full-load sliding (s_{fl})	0.030	–
Starting torque (τ_{st})	106.46	Nm
Maximum torque ($\tau_{\max}$)	82.43	Nm
Full-load torque (τ_{fl})	228.73	Nm

Source: Authors

The lower and upper limits of the decision variables for both test systems are listed in Table 3.

Table 3. Upper and lower bounds of the decision variables in both test systems

	First test system		Second test system	
Par.	**Min.** (Ω)	**Max.** (Ω)	**Min.** (Ω)	**Max.** (Ω)
R_1	1.0	1.20	0.40	0.80
R_2	1.0	1.20	0.20	0.50
X_1	1.0	1.20	0.80	1.40
X_1	1.0	1.20	0.20	0.60
X_M	30	50	20	40

Source: Authors

Numerical results and discussion

For the computational implementation of the proposed SCA, the MATLAB software (version 2021*b*) was employed on a PC with an AMD Ryzen 7 3700 2.3 GHz processor and 16.0 GB RAM, running a 64-bit version of Microsoft Windows 10 Single Language. To implement the SCA, a population size (*i.e.*, N_s) of 100, 1000 iterations, and 100 repetitions were employed.

Results obtained for the first test system

The first test system was taken from (Mohammadi and Akhavan, 2014), where three metaheuristic optimizers were applied: the PSO (Gulbahçe and Karaaslan, 2021), the classical GA (Fortes, Ferreira, and Coelho, 2013), and a hybrid between the two (HGAPSO) (Mohammadi and Akhavan, 2014). The numerical results obtained with each comparison method were contrasted with those of the proposed SCA.

Table 4 presents a comparative analysis between the measured torques (provided by the manufacturer) and the values calculated via the optimal solution reported by the methods used for comparison (Rengifo-Santana, Benzaquen-Suñe, Aller-Castro, Bueno-Montilla, and Restrepo-Zambrano, 2015). Note that the errors in Table 4 correspond to the absolute error, which was calculated as presented below:

$$\text{Error} = 100\% \left| \frac{z_m - z_c}{z_m} \right|, \tag{17}$$

where z_m corresponds to the measured data and z_c to the calculated value for each variable.

Table 4. Comparative torque analysis for the first test system

Method	τ_{st} (Nm)		$\tau_{\max}$ (Nm)		τ_{fl} (Nm)	
	Calculated	**Error** (%)	**Calculated**	**Error** (%)	**Calculated**	**Error** (%)
Manuf.	119.2629	–	149.0820	–	19.6730	–
GA	124.0018	3.97	154.0410	3.33	19.8227	0.76
PSO	121.0186	1.47	151.1170	1.37	19.8136	0.71
HGAPSO	119.2300	0.03	149.1226	0.03	19.7877	0.58
SCA	119.2639	8.38×10^{-4}	149.0827	4.58×10^{-4}	19.6734	1.90×10^{-3}

Source: Authors

In comparison with those reported in the literature (Mohammadi and Akhavan, 2014), the results in Table 4 show that:

i. The SCA provides the best estimates for the starting, maximum, and full-load torques. Regarding the starting torque, the estimation error was about 8.38×10^{-4}%, while the HGAPSO approach reported about 3×10^{-2}%, *i.e.*, about 35 times higher than the SCA. As for the maximum torque, the SCA reached an estimation error of about 4.58×10^{-4}%, which is 65 times lower than the result reported for the HGAPSO approach. In the case of the full-load torque, the SCA reported an estimation error of about 1.90×10^{-3}, followed by the HGAPSO with a value of 0.58%, *i.e.*, the precision of the SCA is about 305 times better than that of the HGAPSO for this torque estimation.

ii. The GA and the PSO, when implemented independently, exhibit higher estimation errors in comparison with the HGAPSO approach, which is evident in the case of the starting and maximum torques. However, in the case of the full-load torque, these three methods have a similar behavior, with values between 0.58 and 0.76%.

iii. As for the objective function value, it is worth mentioning that the GA showed a value of about 2.75×10^{-4}, the PSO reported about $3,48 \times 10^{-5}$, the HGAPSO found about 3.41×10^{-5}, and the SCA reached about 4.63×10^{-10}. These results confirm that the proposed SCA is the best optimization method to estimate parameters in three-phase induction machines while considering manufacturer data, as the objective function value is more than 70 000 times lower than that of the HGAPSO as reported by (Mohammadi and Akhavan, 2014).

Table 5 compares the estimated parameters with respect to the manufacturer data and the results obtained with the metaheuristic optimizers.

In comparison with the specialized literature (Mohammadi and Akhavan, 2014), these numerical results show the following:

i. Regarding the resistive and reactance parameters in the stator and rotor (*i.e.*, R_1, R_2, X_1, and X_2), the SCA exhibited the best numerical estimations, with errors lower than 0.30%. Meanwhile, the HGAPSO approach reported errors higher than 0.70% for the same parameters, which confirms the effectiveness of the SCA in dealing with the analyzed optimization problem.

ii. The magnetization reactance, as estimated by the SCA, was 36.5475 Ω, with an error of about 4.82% compared to the manufacturer data. The HGAPSO approach reported 36.8888 Ω, with an estimation error of about 3.94%. However, even though the SCA showed a higher estimation error in this parameter, given its magnitude in comparison with the stator and rotor resistances and reactances, this has no significant effects on the final objective function value, which is much better for the SCA when compared to the HGAPSO approach. Furthermore, these results confirm that the problem under study is in fact a multimodal optimization problem, *i.e.*, it involves multiple combinations of variables to minimize the objective function (Bocanegra *et al.*, 2021).

Results obtained for the second test system

The second test system is a three-phase induction motor that has not previously been reported in the specialized literature on parameter estimation via metaheuristic optimization. Therefore, considering the effective, efficient, and robust performance of the proposed SCA in the first test system, its best three results for this system are presented. Table 6 presents a comparison between the manufacturer torque data and the calculated values for each of these solutions (Pedra and Corcoles, 2004). In contrast, Table 7 presents the estimated resistances and reactances, as well as their comparison against manufacturer data.

These results confirm that the best three solutions reached with the SCA report objective function values lower than 2.17×10^{-8}, with excellent performance regarding maximum torque, and slight deviations for the starting and full-load torques.

The numerical results in Table 7 show that:

i. The parameter with the lowest estimation error regarding the manufacturer data is the stator resistance (R_1).

ii. Solution 1 is the best one regarding the objective function value and the estimation errors in each parameter, with values lower than 2% in the case of stator and rotor resistances and reactances and about 15.43% in the case of the magnetization reactance.

iii The parameter with the highest estimation error is the magnetization reactance, which varies from 15 to

Table 5. Comparative analysis for the first test system

Method	R_1 (Ω)		R_2 (Ω)		X_1, X_2 (Ω)		X_M (Ω)	
	Calculated	**Error** (%)	**Calculated**	**Error** (%)	**Calculated**	**Error** (%)	**Calculated**	**Error** (%)
Manuf.	1.1150	–	1.0830	–	1.1260	–	38.4000	–
GA	1.0291	7.70	1.0304	4.86	1.0739	4.63	20.4139	46.84
PSO	1.1029	1.09	1.0351	4.42	1.0858	3.57	21.9514	42.80
HGAPSO	1.1229	0.71	1.0741	0.82	1.1169	0.81	36.8888	3.94
SCA	1.1135	0.13	1.0800	0.28	1.1237	0.20	36.5475	4.82

Source: Authors

Table 6. Comparative torque analysis for the second test system

Sol.	τ_{st} (Nm)		$\tau_{\max}$ (Nm)		τ_{fl} (Nm)		
	Calculated	**Error** (%)	**Calculated**	**Error** (%)	**Calculated**	**Error** (%)	E_f
Manuf.	106.46	–	228.73	–	82.43	–	–
Sol. 1	106.46	0	228.73	0	82.43	0	2.84×10^{-9}
Sol. 2	106.47	9.39×10^{-3}	228.73	0	82.42	1.21×10^{-2}	1.84×10^{-8}
Sol. 3	106.45	9.39×10^{-3}	228.73	0	82.44	1.21×10^{-2}	2.17×10^{-8}

Source: Authors

47% in all solutions. However, these variations did not affect the torque calculations reported in Table 6.

To demonstrate that the best solutions reached with the SCA (Tables 6 and 7) adequately model the torque curve in the analyzed three-phase induction machine, the torque behavior for all solutions, including the benchmark case, is depicted in Figure 3.

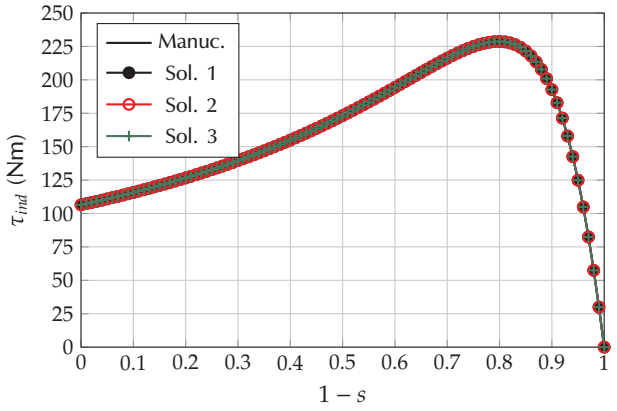


Figure 3. Torque behavior in the second test system for the best three solutions reached with the SCA
Source: Authors

Note that, regardless of the estimation difference exhibited by this particular parameter in Table 7 with respect to the manufacturer data, the best three solutions reached with the SCA allow adequately reproducing the induced torque defined in Equation (4), which confirms the results regarding the starting, maximum, and full-load torques in Table 6. These results confirm the multimodal nature of the parametric estimation problem in three-phase induction motors.

It is worth mentioning that an optimization problem is considered to be multimodal when multiple combinations of variables provide the same numerical objective function value, *i.e.*, the solution is not unique. However, in the case of parametric estimation in electrical machines, this behavior is typical, given the nonlinearities between parameters and electrical variables. Notwithstanding, if one solution is selected for numerical simulations and physical implementations, the motor's expected dynamic and static behavior will exhibit slight variations with respect to other potential solutions.

Conclusions and future work

This paper proposed an efficient solution methodology to determine electrical parameters in three-phase induction machines while considering measurement data provided by the manufacturer regarding starting, maximum, and full-load torques. This problem was formulated as a nonlinear programming model, with the aim of minimizing the sum of the errors between the measured and the calculated torques.

According to a comparative analysis with three metaheuristic optimization algorithms (*i.e.*, GA, PSO, and HGAPSO) in the first test system analyzed, the proposed SCA was the most effective optimization algorithm for the studied objective function, with a final value of about 4.63×10^{-10}, while the best literature report (the HGAPSO) found an objective function value of 3.41×10^{-5}. In addition, regarding each particular parameter (*i.e.*, stator and rotor resistances and reactances), the SCA exhibited the best numerical performance, with the lowest estimation errors. However, in the case of the magnetization reactance, the HGAPSO approach reported a better estimation, which could be attributed to the multimodal nature of the optimization problem under study.

Numerical results in the second test system confirmed that there are multiple combinations of resistances and reactances that allow for an adequate reproduction of the induced torque in the entirety of the operating range. In addition, the parameter with the highest deviation with respect to the manufacturer data was the magnetization reactance, which can also be attributed to the multimodal nature of the optimization model and the lack of information regarding the active power behavior of the induction motor.

Table 7. Comparative analysis for the second test system

Solution	R_1 (Ω)		R_2 (Ω)		X_1 (Ω)		X_2 (Ω)		X_M (Ω)	
	Calc.	**Error** (%)	**Calc.**	**Error** (%)	**Calc.**	**Error** (%)	**Calc.**	**Error** (%)	**Calc.**	**Error** (%)
Manuf.	0.6410	–	0.3320	–	1.1060	–	0.4640	–	26.3000	–
Sol. 1	0.6444	0.53	0.3356	1.08	1.1076	0.14	0.4733	2.00	30.3571	15.43
Sol. 2	0.6507	1.51	0.3422	3.07	1.0235	7.46	0.5750	23.92	38.6619	47.00
Sol. 3	0.6455	0.70	0.3367	1.42	1.2458	12.64	0.3432	26.03	35.8999	36.50

Source: Authors

In this study on parametric estimation, the highest errors were reported for the magnetizing reactance (*i.e.*, X_M). However, these errors can be explained by the fact that this is the largest parameter in the induction motor, with a parallel connection to the equivalent circuit motor. According to circuit theory, the sum of two or more parallel elements is always lower than the small parameter, which confirms that, for variations in the largest parameter, this effect is minimal or negligible in the final solution.

As future work, the following studies could be conducted: (i) the application of new metaheuristic optimization algorithms to solve the nonlinear programming model that represents the studied problem; (ii) the use of the exact model of the induction machine without simplifications in the Thevenin equivalent impedance; and (iii) the development of a formulation that includes efficiency, reactive power, and the power factor, among other parameters.

Acknowledgements

This research received support from the Ibero-American Science and Technology Development Program (CYTED), through thematic network 723RT0150, *Red para la integración a gran escala de energías renovables en sistemas eléctricos* (RIBIERSE-CYTED).

CRedit author statement

ODM conceived the idea, conducted the background research, supervised the study, and provided critical feedback. *SDN-C.* and *JCP-G.* collected the data, developed the workflow, and performed assessments. *SDN-C* and *JCP-G* wrote the main part of the manuscript, to which all authors contributed.

Conflicts of interest

The authors declare no conflict of interest.

Implementation of the first motor in MATLAB using the SCA

This appendix presents the commands used to implement the proposed optimization method. In the following MATLAB code, three essential things should be noted.

1. The first part of the algorithm focuses on parameterizing the test motor under analysis.
2. The main characteristics of the SCA are defined in order to carry out the optimization process.
3. A function named `[···] = PCM(···)` is recursively called to determine all the electromagnetic torques required in evaluating the objective function. This function is defined in the final part of the MATLAB code.

```
clc;clear;close;
tStart = tic;
Neval = 1;
ResultadosF = zeros(Neval,5);
for cant = 1:Neval
    Vph = 460*sqrt(2)/sqrt(3);
    f = 60;
    Xm = 38.4000; Xs = 1.1260; Xr = 1.1260;
    Rr = 1.0830; Rs = 1.1150;Tst = 119.2629;
    Tfl = 19.6730;Tmx = 149.0820;Sfl = 0.0210;
    [Tflo,f1,Tsta,f2,Tmax,f3,FF] = ...
    PCM(Vph,f,Xm,Xs,Rr,Rs,Tst,Tfl,Tmx,Sfl);
    tmax = 1000;
    xmin = [30.0000 1.0000 1.0000 1.0000];
    xmax = [46.0000 1.2000 1.2000 1.2000];
    NV = size(xmin,2);
    Ns = 100;
    x = (xmin) + rand(Ns,1).*(xmax - xmin);
    for i = 1:Ns
        Xm = x(i,1); Xs = x(i,2);
        Rs = x(i,3); Rr = x(i,4);
        [Tflo,f1,Tsta,f2,Tmax,f3,FF] = ...
        PCM(Vph,f,Xm,Xs,Rr,Rs,Tst,Tfl,Tmx,Sfl);
        x(i,NV+1) = f1^2 + f2^2 + f3^2;
    end
    x = sortrows(x,NV+1);
    for t = 0:tmax
        r1 = 1 - t/tmax; r2 = -pi + rand()*(2*pi);
        r3 = rand(); xbest = x(1,:);
        for i = 1:Ns
            if rand(1) >= 1/2
                xd = x(i,:) + r1*sin(r2)*...
                    abs(r3*xbest - x(i,:));
            else
                xd = x(i,:) + r1*cos(r2)*...
                    abs(r3*xbest - x(i,:));
            end
            for j = 1:NV
                if  xd(1,j) < xmin(1,j) ||...
                        xd(1,j) > xmin(1,j)
                    xd(1,j)  = xmin(1,j) + ...
                        rand()*(xmax(1,j) - ...
                        xmin(1,j));
                end
            end
            Xm  = xd(1,1); Xs  = xd(1,2);
            Rs= xd(1,3); Rr  = xd(1,4);
            [Tflo,f1,Tsta,f2,Tmax,f3,FF] = ...
            PCM(Vph,f,Xm,Xs,Rr,Rs,Tst,Tfl,Tmx,Sfl);
            xd(1,NV+1) = f1^2 + f2^2 + f3^2;
            if xd(1,end) < x(i,end)
                x(i,:) = xd;
            end
        end
```

```
        x = sortrows(x,NV+1);
        fprintf('Iteration: %d\n',t);
    end
    disp(xbest)
    ResultadosF(cant,:) = xbest;
end
tEnd = toc(tStart);
function [Tflo,f1,Tsta,f2,Tmax,f3,FF] = ...
  PCM(Vpha,f,Xmd,Xse,Rrg,Rsh,Tsti,Tflj,Tmxk,Sflm)
  Vth = Vpha*Xmd/(Xmd + Xse);
  Rth = Rsh*Xmd/(Xmd + Xse);
  Xth = Xse*Xmd/(Xmd + Xse);
  Kt = 3*(Vth^2)/(2*pi*f );
  Tflo = (Kt*Rrg)/(Sflm*((Rth + Rrg/Sflm)^2 + ...
        (Xth + Xse)^2));
  f1 = (Tflj- Tflo)/(Tflj);
  Tsta = (Kt*Rrg)/((Rth + Rrg)^2 + ...
        (Xth + Xse)^2);
  f2 = (Tsti - Tsta)/(Tsti);
  Tmax = (Kt)/(2*(Rth + sqrt((Rth)^2 + ...
        (Xth + Xse)^2)));
  f3 = (Tmxk -  Tmax)/(Tmxk);
  FF = f1^2 + f2^2 + f3^2;
end
```

References

Abo-Khalil, A. G., Abdelkareem, M. A., Sayed, E. T., Maghrabie, H. M., Radwan, A., Rezk, H., and Olabi, A. G. (2022). Electric vehicle impact on energy industry, policy, technical barriers, and power systems. *International Journal of Thermofluids*, *13*, 100134. Retrieved from https://doi.org/10.1016/j.ijft.2022.100134

Aminu, M. (2019). A parameter estimation algorithm for induction machines using artificial bee colony (ABC) optimization. *Nigerian Journal of Technology*, *38*(1), 193. Retrieved from https://doi.org/10.4314/njt.v38i1.24

Attia, A.-F., Sehiemy, R. A. E., and Hasanien, H. M. (2018). Optimal power flow solution in power systems using a novel Sine-Cosine algorithm. *International Journal of Electrical Power & Energy Systems*, *99*, 331–343. Retrieved from https://doi.org/10.1016/j.ijepes.2018.01.024

Avalos, O., Cuevas, E., and Gálvez, J. (2016). Induction motor parameter identification Using a gravitational search algorithm. *Computers*, *5*(2), 6. Retrieved from https://doi.org/10.3390/computers5020006

Bocanegra, S. Y., Montoya, O. D., and Molina, A. (2021). Sine-cosine optimization approach applied to the parametric estimation in single-phase transformers by considering voltage and current measures. *DYNA*, *88*(219), 19–27. Retrieved from https://doi.org/10.15446/dyna.v88n219.93670

Bouakkaz, A., Mena, A. J. G., Haddad, S., and Ferrari, M. L. (2021). Efficient energy scheduling considering cost reduction and energy saving in hybrid energy system with energy storage. *Journal of Energy Storage*, *33*, 101887. Retrieved from https://doi.org/10.1016/j.est.2020.101887

Ćalasan, M., Micev, M., Ali, Z. M., Zobaa, A. F., and Aleem, S. H. E. A. (2020). Parameter estimation of induction machine single-cage and double-cage models using a hybrid simulated annealing-evaporation rate water cycle algorithm. *Mathematics*, *8*(6), 1024. Retrieved from https://doi.org/10.3390/math8061024

Chauhan, R. K., Chauhan, K., and Badar, A. Q. (2022). Optimization of electrical energy waste in house using smart appliances management system { a case study. *Journal of Building Engineering*, *46*, 103595. Retrieved from https://doi.org/10.1016/j.jobe.2021.103595

Fortes, M. Z., Ferreira, V. H., and Coelho, A. P. F. (2013). The induction motor parameter estimation using genetic algorithm. *IEEE Latin America Transactions*, *11*(5), 1273–1278. Retrieved from https://doi.org/10.1109/tla.2013.6684404

Friederici, P. (2021). In germany, the energy transition continues. *Bulletin of the Atomic Scientists*, *77*(2), 82–85. Retrieved from https://doi.org/10.1080/00963402.2021.1885851

Guedes, J. J., Castoldi, M. F., Goedtel, A., Agulhari, C. M., and Sanches, D. S. (2018). Parameters estimation of three-phase induction motors using differential evolution. *Electric Power Systems Research*, *154*, 204–212. Retrieved from https://doi.org/10.1016/j.epsr.2017.08.033

Gupta, R. A., Wadhwani, A. K., and Kapoor, S. R. (2011). Early estimation of faults in induction motors using symbolic dynamic-based analysis of stator current samples. *IEEE Transactions on Energy Conversion*, *26*(1), 102–114. Retrieved from https://doi.org/10.1109/tec.2010.2062514

Gulbahçe, M. O., and Karaaslan, M. E. (2021). Estimation of induction motor equivalent circuit parameters from manufacturer's datasheet by particle swarm optimization algorithm for variable frequency drives. *Electrica*, *22*(1), 16–26. Retrieved from https://doi.org/10.5152/electrica.2021.21122

Huynh, D. C., and Dunnigan, M. W. (2010). Parameter estimation of an induction machine using a dynamic particle swarm optimization algorithm. In IEEE (Ed.), *2010 IEEE International Symposium on Industrial Electronics*. IEEE. Retrieved from https://doi.org/10.1109/isie.2010.5637818

Joodaki, A., Shojaei, A. A., and Lotfi, H. (2020). Estimation of induction motor parameters using COA algorithm. *Journal of Advances in Computer Research*, *11*(4), 117–130.

Lindenmeyer, D., Dommel, H., Moshref, A., and Kundur, P. (2001). An induction motor parameter estimation method. *International Journal of Electrical Power & Energy Systems*, *23*(4), 251–262. Retrieved from https://doi.org/10.1016/s0142-0615(00)00060-0

Mohammadi, H. R., and Akhavan, A. (2014). Parameter estimation of three-phase induction motor using hybrid of genetic algorithm and particle swarm optimization. *Journal of Engineering*, *2014*, 1–6. Retrieved from https://doi.org/10.1155/2014/148204

Montoya, O. D., Gil-González, W., and Grisales-Noreña, L. F. (2020). Sine-cosine algorithm for parameters' estimation in solar cells using datasheet information. *Journal of Physics: Conference Series*, *1671*(1), 012008. Retrieved from https://doi.org/10.1088/1742-6596/1671/1/012008

Nota, G., Nota, F. D., Peluso, D., and Lazo, A. T. (2020). Energy efficiency in industry 4.0: The case of batch production processes. *Sustainability*, *12*(16), 6631. Retrieved from https://doi.org/10.3390/su12166631

Payán, M. B., Fernandez, J. M. R., Ortega, J. M. M., and Santos, J. M. R. (2019). Techno-economic optimal power rating of induction motors. *Applied Energy*, *240*, 1031–1048. Retrieved from https://doi.org/10.1016/j.apenergy.2019.02.016

Pedra, J., and Corcoles, F. (2004). Estimation of induction

motor double-cage model parameters from manufacturer data. *IEEE Transactions on Energy Conversion*, *19*(2), 310–317. Retrieved from https://doi.org/10.1109/tec.2003.822314

Rengifo-Santana, J. W., Benzaquen-Suñe, J., Aller-Castro, J. M., Bueno-Montilla, A. A., and Restrepo-Zambrano, J. A. (2015). Parameter estimation method for induction machines using instantaneous voltage and current measurements. *Revista Facultad de Ingeniería Universidad de Antioquia*, *75*, 57–66. Retrieved from https://doi.org/10.17533/udea.redin.n75a07

Rezk, H., Elghany, A. A., Al-Dhaifallah, M., Sayed, A. H. M. E., and Ibrahim, M. N. (2019). Numerical estimation and experimental verification of optimal parameter identification based on modern optimization of a three phase induction motor. *Mathematics*, *7*(12), 1135. Retrieved from https://doi.org/10.3390/math7121135

Sengamalai, U., Anbazhagan, G., Thentral, T. M. T., Vishnuram, P., Khurshaid, T., and Kamel, S. (2022). Three phase induction motor drive: A systematic review on dynamic modeling, parameter estimation, and control schemes. *Energies*, *15*(21), 8260. Retrieved from https://doi.org/10.3390/en15218260

Toliyat, H., Levi, E., and Raina, M. (2003). A review of RFO induction motor parameter estimation techniques. *IEEE Transactions on Energy Conversion*, *18*(2), 271–283. Retrieved from https://doi.org/10.1109/tec.2003.811719

Trisha, Gupta, G. S., and Kumar, S. S. (2021). Review of the parameter estimation and transient analysis of three-phase induction motor. In M. J. B. Reddy, D. K. Mohanta, D. Kumar, and D. Ghosh (Eds.), *Advances in smart grid automation and Industry 4.0* (pp. 223–232). Springer Singapore. Retrieved from https://doi.org/10.1007/978-981-15-7675-1_21

Véliz-Tejo, A., Travieso-Torres, J. C., Peters, A. A., Mora, A., and Leiva-Silva, F. (2022). Normalized-model reference system for parameter estimation of induction motors. *Energies*, *15*(13), 4542. Retrieved from https://doi.org/10.3390/en15134542

Wu, R.-C., Tseng, Y.-W., and Chen, C.-Y. (2018). Estimating parameters of the induction machine by the polynomial regression. *Applied Sciences*, *8*(7), 1073. Retrieved from https://doi.org/10.3390/app8071073

Research Article/Industrial Engineering

http://doi.org/10.15446/ing.investig.104181

A Novel Global Probabilistic Fuzzy System for Occupational Risk Assessment (GPFSORA)

Un novedoso sistema probabilístico difuso global para la evaluación de riesgos laborales (GPFSORA)

Roberto Baeza-Serrato[1]

ABSTRACT

Occupational risk assessment is the process of estimating the magnitude of risks that cannot be avoided. Then, the corresponding assessment is carried out, using comparative tables with different evaluation methods. Current risk assessment techniques enable the individual assessment of each potential risk, but there is no method to globally assess potential risks in an organization. The motivation of this research was to develop an objective and quantitative risk assessment system through a diffuse probabilistic model integrating stochastic and non-stochastic uncertainty. To this effect, an empirical collective record was used, whose attribute of interest was the occurrence of different accident types over a period of 52 weeks. Here, each of the collectives represented a linguistic input variable. In the probabilistic fuzzification stage, the frequentist probability of the occurrence of accidents was determined. One of our most important contributions to probabilistic fuzzy systems lies in our classification of language labels based on the linguistic projection of frequentist probabilities via a projection membership function determined by experts. The use of the total probability theorem in the implication stage is also proposed. The output of the system determines the type of risk, its evaluation, and the probability of its occurrence, vital factors to be considered in prevention work. The system's stages are explicitly described and applied to real data corresponding to construction materials distribution company. One of the relevant conclusions of this research is that the integration of stochastic and imprecise uncertainty allows for a more reliable risk assessment system.

Keywords: probabilistic fuzzy system, frequentist probability, total probability theorem

RESUMEN

La evaluación de riesgos laborales es el proceso de estimar la magnitud de los riesgos que no se pueden evitar. Luego, se lleva a cabo la evaluación correspondiente, utilizando tablas comparativas con diferentes métodos de evaluación. Las técnicas actuales de evaluación de riesgos permiten la evaluación individual de cada riesgo potencial, pero no hay un método para evaluar globalmente los riesgos potenciales en una organización. La motivación de esta investigación fue desarrollar un sistema objetivo y cuantitativo de evaluación de riesgos a través de un modelo probabilístico difuso que integrara la incertidumbre estocástica y no estocástica. Para ello, se utilizó un registro colectivo empírico, cuyo atributo de interés fue la ocurrencia de diferentes tipos de accidentes durante un período de 52 semanas. Aquí, cada uno de los colectivos representaba una variable de entrada lingüística. En la etapa de difusión probabilística, se determinó la probabilidad frecuentista de la ocurrencia de accidentes. Una de nuestras contribuciones más importantes a los sistemas difusos probabilísticos radica en la clasificación de etiquetas de lenguaje con base en la proyección lingüística de probabilidades frecuentistas a través de una función de membresía de proyección determinada por expertos. También se propone el uso del teorema de probabilidad total en la etapa de implicación. La salida del sistema determina el tipo de riesgo, su evaluación y la probabilidad de su ocurrencia, factores vitales a tener en cuenta en el trabajo de prevención. Las etapas del sistema se describen explícitamente y se aplican a datos reales de una empresa de distribución de materiales de construcción. Una de las conclusiones relevantes de esta investigación es que integrar incertidumbre estocástica e imprecisa permite un sistema de evaluación de riesgos más confiable.

Palabras clave: sistema probabilístico difuso, probabilidad frecuentista, teorema de probabilidad total

Received: August 09th, 2022
Accepted: October 11th, 2023

Introduction

The fuzzy logic system cannot handle various uncertainties in practical applications and is notoriously not the best option to process stochastic uncertainties, since it can control, classify, or evaluate the process in two-dimensional natures (Sozhamadevi *et al*., 2012). Moreover, their linguistic expression cannot handle stochastic uncertainty, and the Mamdani and Takagi-Sugeno-type inference engine may not be suitable for working under a stochastic environment with incomplete dynamics (Li and Liu, 2009). Due to the complexity and diversity of big data, it is difficult to deal with uncertain and fuzzy information (Xue *et al*., 2020). Fuzzy systems emulate human reasoning. However, decision-making is currently based on all the information available from systems, including stochastic uncertainty conditions. This represents an area of opportunity to strengthen the

[1] PhD in Industrial Engineering, CIATEC, Mexico. Affiliation: Associate professor, Universidad de Guanajuato, Yuriria campus, México. Email: r.baeza@ugto.mx

inference results of fuzzy systems, which constitutes the motivation of our research. To date, researchers have paid more attention to multi-attribute decision-making problems in imprecise and uncertain environments (De *et al*., 2019). Fuzzy entropy provides a measure of fuzziness (ambiguity). Therefore, it is important in decision-making applications with imprecise (diffuse) values (Aggarwal, 2021). In multi-criteria group decision-making, those in charge often do not share the same concepts (Farhadinia *et al*., 2020). The fuzzy probabilistic set is designed to handle uncertainties that are blurred and stochastic in nature, so the its logic system can handle more complex uncertainties (Huang *et al*., 2019). Additionally, a fuzzy probability inference system provides a measure of dispersion and seems to be essential to designing systems that incorporate linguistic and/or numerical uncertainties, which are transformed into uncertainty rules, analogous to how variance relates to the mean (Sozhamadevi and Sathiyamoorty, 2015). The approach proposed in this research integrates non-stochastic uncertainty conditions, emulating the human cognitive process, and stochastic uncertainty conditions, using the empirical collectives of the number of accidents associated with different types of occupational risks.

In general, occupational health and safety (OSH) risk assessment models are established upon the basis of four basic steps: hazard identification, hazard characterization, exposure assessment, and risk characterization. Currently, there is little guidance for choosing the most suitable model for a given application, which is based on the individual judgment of the expert and can therefore lead to highly variable, experience-dependent results (Tian *et al*., 2018). One of the most important stages of a risk assessment system is the categorization of risk levels for the operations under study, *i.e.*, a characterization of risk levels, which are subjectively assigned according to the heuristic knowledge of the experts based on specific criteria. Then, the corresponding evaluations are conducted upon the bases of comparative tables. Boyaci (2022) described OSH risk assessment techniques in two categories: qualitative and quantitative. In qualitative techniques, a numerical value is assigned to each parameter, such as the probability and severity of the hazard. These values can be processed using mathematical and logical methods and a calculated risk value.

Gül (2020) stated that some of the most common methods for risk assessment are based on the subjective estimation of two or three criteria. The matrix method has two parameters called *probability* and *gravity*, and, in the Fine-Kinney method, there are three parameters: *gravity*, *probability*, and *frequency*. Moreover, in the failure modes and effects analysis (FMEA) method there are three parameters: *severity*, *probability*, and *detectability*.

These assessment methods are used to individually analyze and assess each potential risk in an organization, based on the knowledge and experience of experts. However, they are not useful for making a comprehensive assessment of all potential risks in the organization or for predicting their probability of occurrence. A comparison is often made in a table showing possible combinations, and the reference criteria are qualitatively described to carry out a subjective evaluation based on the interpretation of the experts, which may be imprecise and vague. Nevertheless, in each organization, there are registers of different types of risks and accidents, which can be used to design a global prediction, classification, and evaluation system according to specific needs and priorities, rather than an individual and subjective approach. This constitutes the motivation of our research, which develops a probabilistic system based on fuzzy logic that objectively integrates the calculated probabilities of each type of accident, analyzed by adding stochastic uncertainty. Fuzzy logic is used to integrate non-stochastic uncertainty and obtain the results in a language that is simple and clear to categorize, assess, and estimate. Labor-, material-, and site condition-related accident records are analyzed as an empirical collective. Frequentist probability is determined, which is projected on language labels through weighted percentages determined by experts according to their cognitive process. For the implication stage, we propose using the total probability theorem, where each of the linguistic variables has an important weighting factor. Finally, with the aggregation stage, there is an estimate of the occurrence of the three types of accidents. The novelty of this study lies in four points:

1. The proposal of a comprehensive method for assessing occupational risks in an organization

2. The integration of stochastic and non-stochastic uncertainty into a robust prediction, categorization, and evaluation system

3. The linguistic projection of frequentist probability as a mechanism to integrate the imprecise uncertainty of cognitive processes and emulate human reasoning, proposing new membership functions for projection on language labels

4. The use of the total probability theorem as the trigger level for each of the fuzzy rules in the implication stage

This manuscript is organized as follows. The introduction presents the limitations of diffuse systems and the area of opportunity to integrate stochastic uncertainty conditions to diffuse systems, aiming for a more robust system. A literature review is presented in the second section. The methodology is presented thereafter, and its application to a case study related to the construction industry is presented at the end.

Literature review

Below are the results of a comprehensive state of the art review regarding qualitative and quantitative risk assessment methods in various industries, which are aimed at analyzing, planning for, and preventing accidents.

Gül (2020) proposed a quantitative assessment of occupational risk via a technique based on classification for order performance by similarity to ideal solutions, in order to manage risks in the aluminum extrusion industry. This quantitative method, like qualitative and subjective methods, makes it possible to assess each of the risks individually and does not present an alternative for assessing risks jointly or globally in order to predict, classify, and evaluate the risks in an organization.

Xu *et al*. (2020) explored the quantitative differences between common occupational health risk assesment (OHRA) models. Risk ratios in five typical industries (leather manufacturing, wood furniture, printing and dyeing, and printing and clothing) were analyzed by using six OHRA models. The consistency, correlation, and reliability of quantitative differences between the models were evaluated. Additionally, Tian *et al*. (2018) identified the essential models of the Environmental Protection Agency, Singapore, Australia, Romania, the International Council of Mining and Metals, and Health Hazardous Substances Control as the six most common based on a literature review. Each OHRA qualitative comparison model had its own strengths and limitations and exhibited a diverse distribution at different levels regarding each evaluation indicator. Note that risk ratios are defined as the relationship between the risk level of a given risk factor (obtained through the given model) and the maximum risk level of that model. Both studies, which compared the most commonly used methodologies for risk assessment, highlighted the absence of a global method for assessing potential risks in an organization.

Fuzzy systems have been developed in several productive sectors to create stronger systems in risk assessment, as shown below.

Boyaci (2022) introduced a new approach to OSH risk assessment, combining the Fine-Kinney method and multi-criteria group decision-making to identify hazards and assess, prioritize, and mitigate risks. The proposed approach was applied to the operating theater of a public hospital in Turkey. The analysis involved three experts with occupational health and safety responsibilities in the unit.

Tang *et al*. (2021) developed a hybrid risk prioritization approach to the Fine-Kinney method through the widespread use of interactive and multi-criteria, better-worse decision-making and the diffuse type II set. First, language terms and type II fuzzy range numbers were employed to handle the problem of expressing uncertain evaluation information on team members. Second, in order to obtain the objective weights of the risk parameters, the best-worst method was proposed for determining their degrees of materiality. Third, blurred range II numbers were incorporated into the generalized method for determining hazard risk priority orders, which can simulate the limited rational behavior of experts under uncertain environment.

Fattahi *et al*. (2020) proposed a novel fuzzy multi-criteria decision-making model based on FMEA to more accurately assess the risks of different failure modes. Fattahi *et al*. (2020) used a method called *Modified-SIRA* (enhanced security risk assessment). The criteria and alternatives were prioritized based on the risk priority number and the methodology called *order preference technique by similarity to the ideal solution*, respectively. The results obtained regarding the risk priority number show that most accidents in the Pakistani construction industry are due to the deficiency of PPEs, followed by electrocution, the improper use of the available PPE, and falling from elevated platforms.

Gül and Celik (2018) proposed a combination of the Fine-Kinney method and a fuzzy rule-based expert system. This approach captures nonlinear causal relationships between Fine-Kinney parameters. Since there is a high level of vagueness involved in the OHS risk assessment data, the rule-based expert system was developed for probability, exposure, and consequence in evaluating risk scores.

Each of the studies proposing the integration of diffuse systems use the same approach as traditional systems, only aiming to assess each individual hazard and not predict, evaluate, and categorize the whole. None of the studies with diffuse logic integrate the stochastic uncertainty condition. So far, no research has been found in the literature which uses traditional models and modifies them with diffuse logic to assess the risks of a company collectively or integrates stochastic and non-stochastic conditions, which constitutes the motivation of this research.

Some recent studies integrating stochastic and non-stochastic conditions in applications different from those involving a risk assessment system are mentioned below.

Wang *et al*. (2021) proposed a probabilistic fuzzy inference method to improve the accuracy of evaluation results by considering the uncertainty of indicators. They elaborated a prediction model based on the Gaussian distribution function. The outputs of the model represent the inputs to the system, and the fuzzy membership functions are made probabilistic by multiplying their value by their corresponding probability.

Jiang and Liao (2021) proposed a network consensus analysis of probabilistic linguistic preference relations based on a new measure of probabilistic distance by Kolmogorov-Smirnov, introducing the cumulative probability distribution of sets of probabilistic linguistic terms. They used the quantitative terms of a probability label and the qualitative terms of a language label.

Modares and Desch (2021) discussed and compared various probabilistic and possibilistic methods based on finite elements, presenting the results of case studies on structures using static and dynamic uncertainty. These results suggest that incorporating uncertainty into the analysis procedure provides a higher level of confidence.

De Ridder *et al*. (2020) presented a novel approach based on machine learning for uncertainty quantification problems involving random and epistemic variables. They employed Bayesian optimization to efficiently propagate this hybrid uncertainty throughout the performance of the system under study.

Liang *et al*. (2020) proposed a multi-attribute group decision-making method under an uncertain and diffuse environment that considers the psychological state of decision makers. A weight determination model was constructed, incorporating objective information on WED decision-making and subjective preferences based on a scoring function and the principle of minimal relative entropy.

Gupta *et al*. (2020) extracted reference values using a controller based on probabilistic and diffuse set theory, and Liang *et al*. (2020) solved group decision-making problems containing inconsistent probabilistic preferred language relations and unknown expert weights.

Each of the above-presented studies integrates stochastic and non-stochastic uncertainty conditions, making decision-making more robust. The use of this approach is aimed at the design and development of diffuse probabilistic risk assessment systems by integrating the stochastic uncertainty state, using the frequency probabilities of different types of occupational risks as input variables, as well as their disaggregation based on the interpretation and knowledge of experts of the vagueness in their values, categorizing them into language labels.

Compared to existing diffuse probabilistic approaches, the novelty and contributions of our proposal can be summarized as follows:

1. An effective method based on probabilistic fuzzy inference is proposed, which can obtain better evaluation results under uncertain conditions.

2. Frequentist probability is broken down into diffuse labels based on the knowledge of the experts, who determine reliability percentages.

3. Based on the total probability theorem, we propose determining the stage of implication of the diffuse probabilistic system.

Fuzzy systems use fuzzy operators such as *maximum*, *minimum*, and *product* to define the compositional rules of inference in the stages of implication and aggregation, employing fuzzified values in the rules of knowledge. In this research, probabilistic values are used. We propose integrating the probabilistic values of each input variable through the total probability theorem in the stage of implication, based on the fact that each of the input variables is an independent event and the conditional probability for each language label is known through the projection of its relative frequencies, *i.e.*, the system's inference mechanism. This, instead of using a compositional rule with diffuse operators.

Methodology

The GPFSORA methodology consists of seven stages (Figure 1). The first stage of the methodology is to identify accident records as empirical groups, where the attribute to be analyzed is the number of accidents per week in a construction company over a period of 52 weeks. In the second stage, the frequentist probability of each group is determined. The third stage identifies the categories of importance or relevance according to the frequency of the number of accidents [a; b], and the frequentist probability is projected on three language labels based on the knowledge of the experts, as well as determined through the product of the frequentist probability of the type of accident and the proposed weighted projection membership function (top middle of Figure 1). In the fourth stage, diffuse rules are established, creating the knowledge base. One of the main contributions of this research lies in the fifth stage of implication, where the use of the total probability theorem is proposed. Finally, in the aggregation phase, the maximum operator is employed, and the probability, evaluation, and classification of accidents is estimated. The corresponding assessment is performed using the central limit theorem.

This research was validated with real data in the southern region of the state of Guanajuato, Mexico. It is worth mentioning that fuzzy systems are based on the fuzzification of input variables in an estimated domain, analyzing the different possible combinations in the established domains. In our research proposal, real data on accidents are used, which determine the domain of the system. They are based on evidence from historical records of accidents, as opposed to defining a domain estimate based on accidents that can happen. In summary, the definition of domains in the input variables is based on accidents that actually occurred, rather than estimating those that might occur.

Empirical collectives

The first stage consists of analyzing the records of the main types of accidents as empirical groups based on objective evidence of the reality of their occurrence over a specific period (Table 1).

Table 1. Accidents records

Week	Type of accident 1	Type of accident 2	.....	Type of accident m
1	X11	X21		Xm1
2	X12	X22		Xm2
.				
.				
n	X1n	X2n		Xmn

Where: m = type of accident, n = week number
Source: Authors

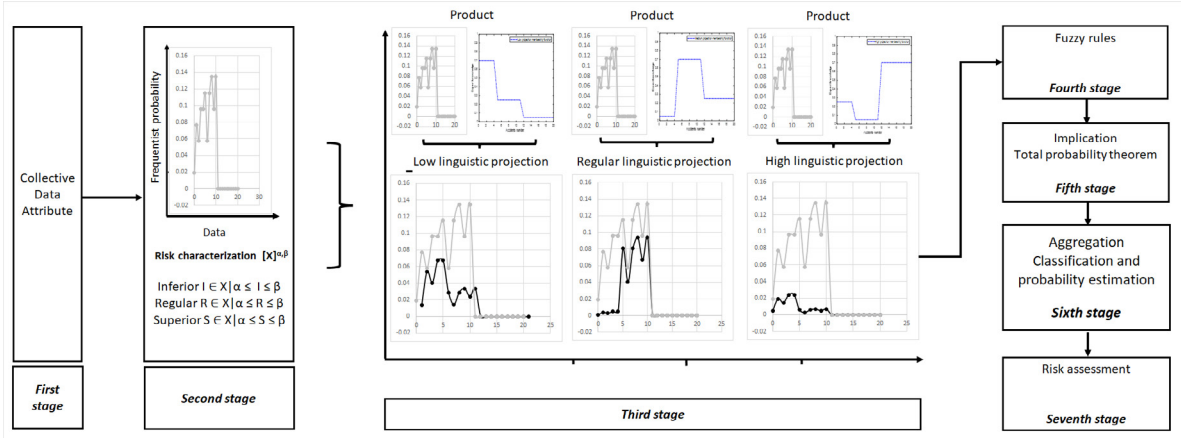


Figure 1. Research methodology
Source: Authors

Frequentist probability and risk characterization

The second stage consists of determining the frequentist probability or relative frequency, as expressed in Equation (1).

$$Rf = \frac{Absolute\ frecuency}{Total} \quad (1)$$

Based on records of the main types of accidents as empirical groups of the organization, integrating stochastic uncertainty into the system as the probability of the associated event occurring, the random variable X is the number of accidents per week in the organization, which can only take integer values between 0 and 20, based on historical records (Table 2).

Table 2. Frequentist probability records

Number of accidents X	Absolute frequency of number of accidents occurred per week	Relative frequency (Rf) of number of accidents	Accumulated relative frequency
0	1	1/52	1/52
1	4	4/52	5/52
2	3	3/52	8/52
3	5	5/52	13/52
4	5	5/52	18/52
5	6	6/52	24/52
6	3	3/52	27/52
7	6	6/52	33/52
8	7	7/52	40/52
9	5	5/52	45/52
10	7	7/52	52/52
11	0	0/52	52/52
12	0	0/52	52/52
13	0	0/52	52/52
14	0	0/52	52/52
15	0	0/52	52/52
16	0	0/52	52/52
17	0	0/52	52/52
18	0	0/52	52/52
19	0	0/52	52/52
20	0	0/52	52/52
Total of weeks	52	52/52	

Source: Authors

Experts determine an α-level set of crisp sets of X to classify them as inferior, regular, or superior for each type of accident in a clear and simple language based on subjective interpretation. In an α-level set of a crisp set I, R and S of X are a crisp set denoted by $[I]^{\alpha,\beta}$, $[R]^{\alpha,\beta}$, and $[S]^{\alpha,\beta}$ (Table 3). The characterization of the α-level sets is obtained by analyzing the history of accident records at work meetings by three plant industrial safety staff experts and two company managers who reached a consensus.

Table 3. Risk characterization

Risk characterization	$[X]^{\alpha,\beta}$	α, β
Inferior (I)	$I \in X \mid \alpha \leq I \leq \beta$	0.4
Regular (R)	$R \in X \mid \alpha \leq R \leq \beta$	5.11
Superior (S)	$S \in X \mid \alpha \leq S \leq \beta$	12.20

Source: Authors

Linguistic projection

Each of the frequentist probabilities is projected on three language labels as a mechanism to integrate the knowledge of experts to the interpretation of relative frequencies. A robust evaluation system is established by integrating stochastic uncertainty, using frequentist probability as input variable as well as vague uncertainty by projecting frequentist probability on labels based on expert knowledge.

One of the most important contributions of our research is the proposal of weighted projection membership functions based on risk categorization level sets to interpret and project the relative frequency of accident types into three categories or language tags, which are shown below. Each of these functions must be multiplied by the frequentist probability of each type of accident (Table 4).

The relative frequency value is projected on three membership functions. In each membership function projection, weightings of 70, 25, and 5% are given to each risk category. 70% of the value is assigned to the similar risk category with its projection.

Table 4. Weighted membership functions of linguistic projection

Low $[Rf]^{\lambda Rf}$	Medium $[Rf]^{\lambda Rf}$	High $[Rf]^{\lambda Rf}$
$I \in Rf \mid I = 0.7\ Rf$	$I \in Rf \mid I = 0.05\ Rf$	$I \in Rf \mid I = 0.25\ Rf$
$R \in Rf \mid R = 0.25\ Rf$	$R \in Rf \mid R = 0.7\ Rf$	$R \in Rf \mid R = 0.05\ Rf$
$S \in Rf \mid S = 0.05\ Rf$	$S \in Rf \mid S = 0.25\ Rf$	$S \in Rf \mid S = 0.7\ Rf$

Source: Authors

The relative frequency value is projected on three membership functions. In each membership function projection, weightings of 70, 25, and 5% are given to each risk category. 70% of the value is assigned to the similar risk category with its projection.

For a low linguistic projection, the weighting 70% is assigned to the inferior risk category (Figure 2). For the medium linguistic projection, the relevant weight 70% is assigned to the regular risk category (Figure 3). Finally, for the high linguistic projection, the weighting 70% is assigned to the superior risk category (Figure 4).

In this vein, the frequentist probability of each type of accident is projected to a language label called *low*, with weights of 70, 25, and 5%. The value of the relative frequency within the lower risk category of zero to four accidents is propagated while maintaining 70% of its value for a low linguistic projection (Figure 2).

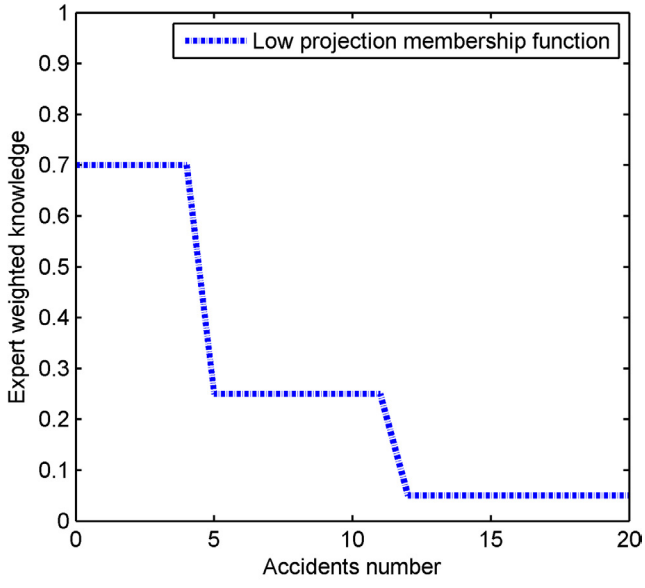


Figure 2. Low membership function projection
Source: Authors

The frequentist probability of each type of accident is projected to a language label called *medium*, with weights of 5, 70, and 25%. The value of the relative frequency within the regular risk category of five to 11 accidents is propagated while maintaining 70% of its value for medium linguistic projection (Figure 3).

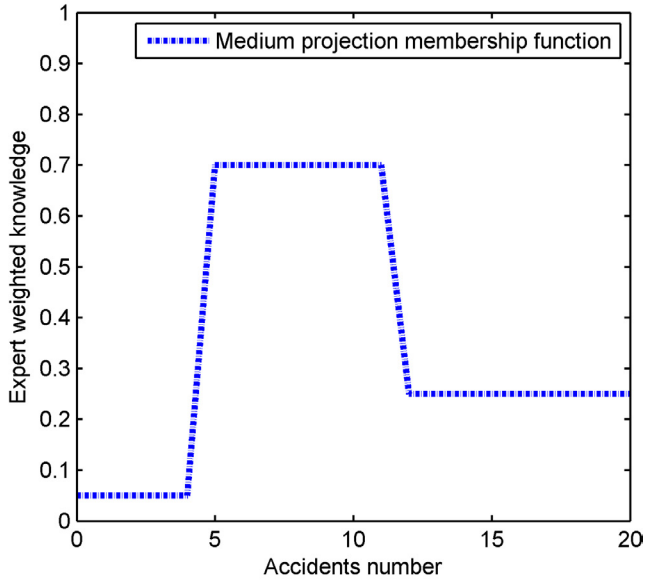


Figure 3. Medium membership function projection
Source: Authors

The frequentist probability of each type of accident is projected to a language label called *high*, with weights of 25, 5, and 70%. The value of the relative frequency within the superior risk category of 12 to 20 accidents is propagated while maintaining 70% of its value for the high linguistic projection (Figure 4).

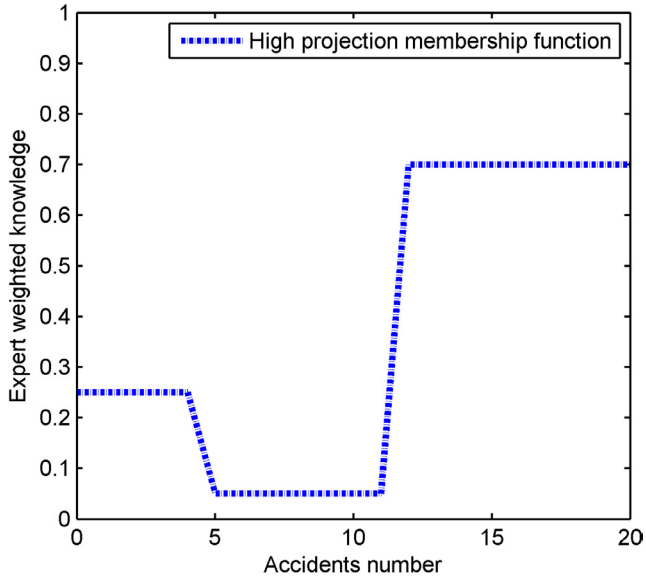


Figure 4. High membership function projection
Source: Authors

These percentages are based on the experts' interpretation of relative frequencies, which are projected on language labels as a mechanism to emulate the experts' reasoning when analyzing and categorizing the relative frequencies of accidents. If the category is lower, a higher percentage is given to the lower risk level. If the category is regular, a higher percentage is given to the average risk level, and,

finally, if the category is higher, a higher percentage is given to the higher risk level. The percentage supplements are distributed in the other two categories to meet the weighted projection (Table 5).

Table 5. Projection of frequentist probability

Risk category	Number of accidents of type 1	Relative frequency	Frequentist probability projection		
			Low	Medium	High
Inferior	0	1/52	70%	5%	25%
	1	4/52	70%	5%	25%
	2	3/52	70%	5%	25%
	3	5/52	70%	5%	25%
	4	5/52	70%	5%	25%
Regular	5	6/52	25%	70%	5%
	6	3/52	25%	70%	5%
	7	6/52	25%	70%	5%
	8	7/52	25%	70%	5%
	9	5/52	25%	70%	5%
	10	7/52	25%	70%	5%
	11	0/52	25%	70%	5%
Superior	12	0/52	5%	25%	70%
	13	0/52	5%	25%	70%
	14	0/52	5%	25%	70%
	15	0/52	5%	25%	70%
	16	0/52	5%	25%	70%
	17	0/52	5%	25%	70%
	18	0/52	5%	25%	70%
	19	0/52	5%	25%	70%
	20	0/52	5%	25%	70%
	Total	**52/52**			

Source: Authors

The proposed membership functions project a percentage of the actual value of the relative frequency. For the low membership function, 70% of the actual value of the relative frequency corresponds to the inferior risk category (zero to four accidents), 25% of the actual relative frequency value is for the regular risk category (5-11 accidents), and 5% of the actual relative frequency value corresponds to the higher risk category (12-20 accidents).

Example

$$\text{Relative frequency of } 0 \text{ accidents} = Rf(0) = (1/52) = 0.019231 \quad (2)$$

$$\text{Low linguistic projection of } 0 \text{ accidents} = Low(0) = (0.7)*(1/52) = 0.013462 \quad (3)$$

$$\text{Relative frequency of } 5 \text{ accidents} = Rf(5) = 6/52 = 0.11538 \quad (4)$$

$$\text{Low linguistic projection of } 5 \text{ accidents} = Low(5) = (0.25)*(6/52) = 0.02886 \quad (5)$$

$$\text{Relative frequency of } 12 \text{ accidents} = Rf(12) = 0/52 = 0.00000 \quad (6)$$

$$\text{Low linguistic projection of } 12 \text{ accidents} = Low(12) = (0.05)*(0/52) = 0.00000 \quad (7)$$

Fuzzy rules

The proposed system analyzes three types of accidents and is classified into three language labels, for which 21 of the 27 feasible fuzzy rules are determined as the basis of knowledge. This is presented in Table 10.

Implication: total probability theorem

In the research proposal, each linguistic variable has a priority weight and is classified as an independent variable.

$$A_1 = \text{Priority weight of the labor - related variable} \quad (8)$$

$$A_2 = \text{Priority weight of the material - related variable} \quad (9)$$

$$A_3 = \text{Priority weight of the site condition - related variable} \quad (10)$$

where

$$P(A_1) = 0.2; P(A_2) = 0.3; P(A_3) = 0.5$$

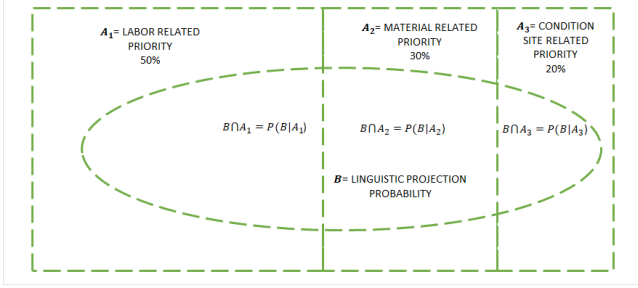
Note that the events A_1, A_2 and A_3 are incompatible and their probabilities add up to 1, so they satisfy the hypotheses of the total probability theorem (Figure 5).

In each proposed fuzzy rule, there is a probabilistic fuzzy value, given that it belongs to a specific linguistic projection represented as $P(B \mid A_i)$.

$$P(B \mid A_1) = \text{Probability of the linguistic projection, given that it belongs to the labor} - \text{related variable} \quad (11)$$

$$P(B \mid A_2) = \text{Probability of the linguistic projection, given that it belongs to the material - related variable} \quad (12)$$

$$P(B \mid A_3) = \text{Probability of the linguistic projection, given that it belongs to the site condition - related variable} \quad (13)$$



As per the total probabilities theorem, the probability of the linguistic projection $P(B)$ is described using Equation (14).

$$P(B) = P(B \mid A_1)P(A_1) + P(B \mid A_2)P(A_2) + P(B \mid A_3)P(A_3) \quad (14)$$

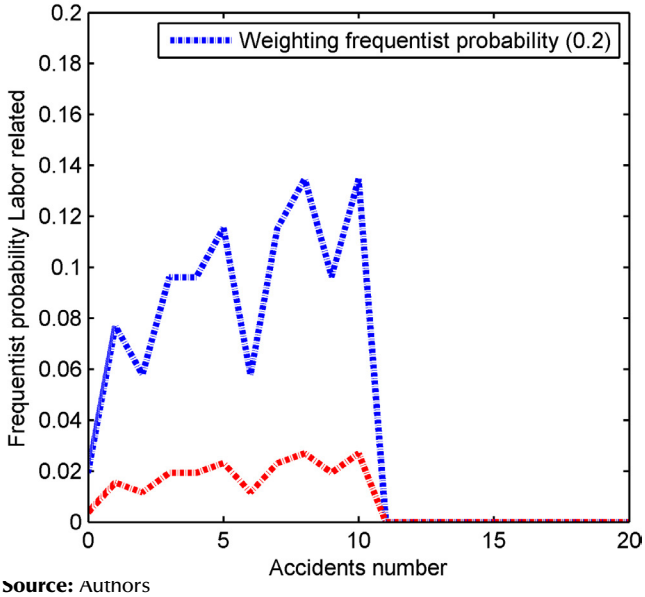
Figure 5 shows the representative Venn diagram of the total probability theorem, where the probability of linguistic

projection $P(B)$ can be determined, since the conditional probabilities of the linguistic labels are known, given the factor to which they belong, $P(B|A_i)$.

The total probability theorem is represented by the product of the weighted probability of the input variable $P(A_i)$ and the known conditional probability $P(B \mid A_i)$. Visually, the product $P(A_i)*P(B \mid A_i)$ is represented by a projection of the known conditional probability. This projection represents the contraction of the conditional probability behavior set at a weighted percentage.

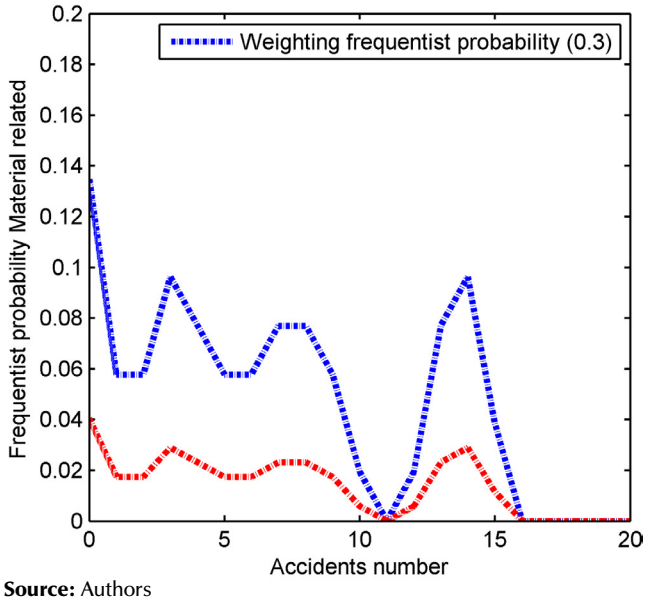
The projections of the weighted frequency probabilities $P(A_i)$for each of the known conditional probabilistic $P(B|A_i)$ are shown below in Figures 6, 7, and 8.

The behavior of the blue trajectory represents the conditional probability $P(B|A_i)$, and the behavior of the red trajectory represents the projection or contraction of the conditional probability, which represents the product $P(A_i)*P(B \mid A_i)$ as a term to determine the total probability proposed in this research. The projection of the weighting behavior of the first type of labor-related accident with 20% of the relative frequency can be seen in Figure 6.



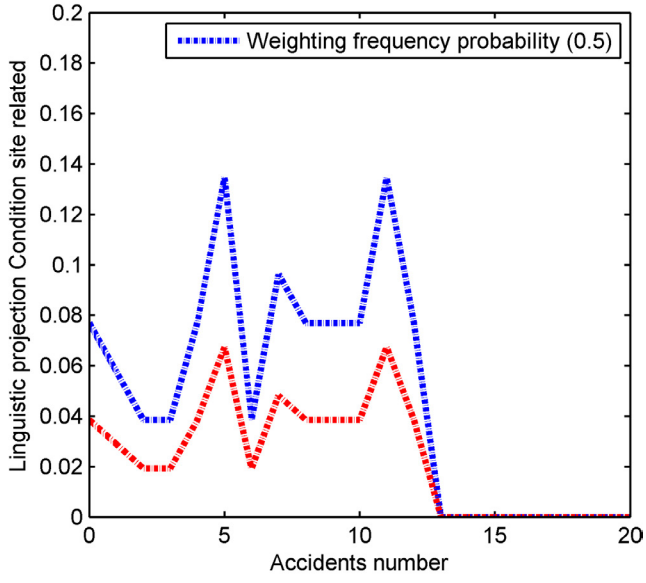
Source: Authors

The blue trajectory represents the conditional probability $P(B \mid A_2)$, and the red one represents the projection or contraction of the conditional probability, which represents the product of the probabilities $P(A_2)*P(B \mid A_2)$ as a term to determine the total probability. The projection of the weighting behavior of the material-related accidents, with 30% of the relative frequency, can be seen in Figure 7.



Source: Authors

The blue trajectory represents the conditional probability $P(B|A_3)$ and the red one represents the projection or contraction of the conditional probability, which represents the product of the probabilities $P(A_3)*P(B \mid A_3)$. The projection of the weighting behavior of site condition-related accidents, with 50% of the relative frequency, can be seen in Figure 8.



Finally, the sum of weighted frequency probabilities is added as the proposed implication mechanism by using the total probability theorem (Figure 9).

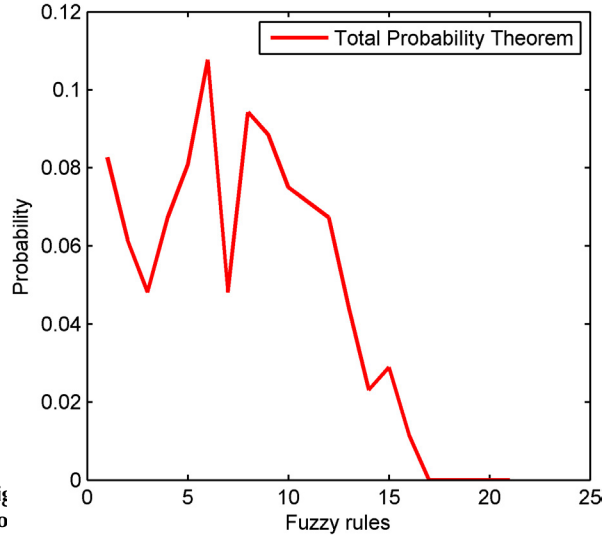


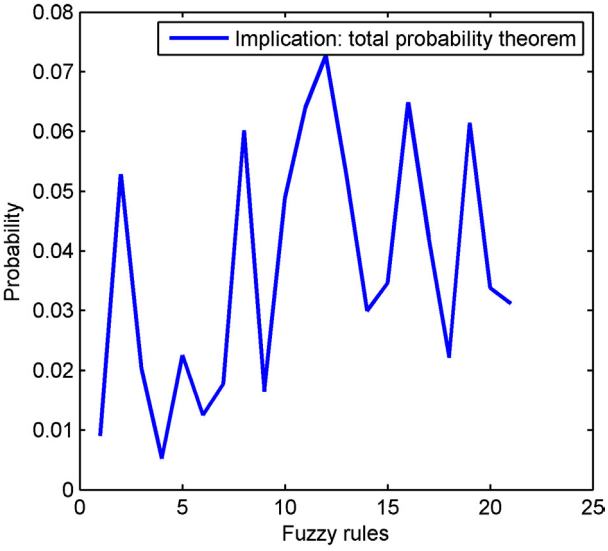
Fig
So

Aggregation

At the aggregation stage, the maximum fuzzy operator is used to find the maximum value of the total probability for each output linguistic label, as well as to identify the linguistic classification of the output variable to which it belongs.

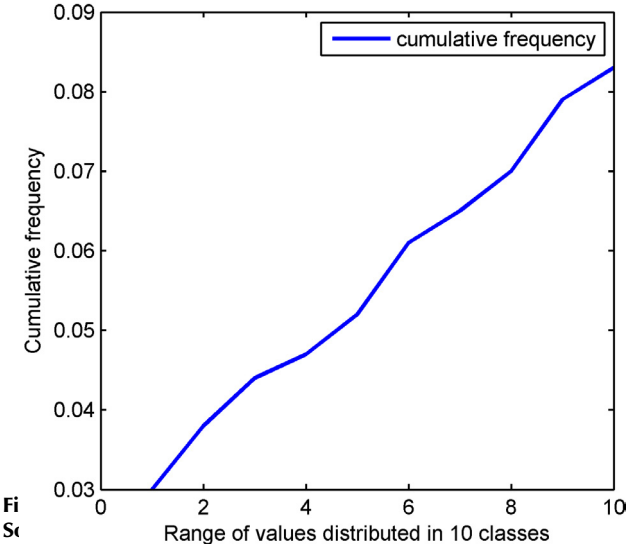
Risk assessment

Below is another scientific contribution of this research, which is related to the assessment phase. The results of 26 training points in the implication phase range from 0 to 0.09**.** Figure 10 shows the results of the implication stage for a training point oscillating between the established range of values.



Fi
Sc

A cumulative probability function is determined in that range of values, distributed in ten classes (Figure 11). Several zones are proposed based on the output risk category.



Fi
Sc

In each zone, the central limit theorem is used for the corresponding evaluations, along with its main mean and standard deviation parameters (Table 6). In summary, when the maximum value is obtained at the aggregation stage, the risk classification is identified, and its normal distribution function in the relevant area is used to assess the company's risk.

Classification and assessment zones

Table 6 shows the classification and assessment zones based on the behavior of the total probability of the system's training points, as well as the parameters of the mean and the standard deviation. The central limit theorem is used as a decoding mechanism, as described using Equation (15).

$$z = \frac{x - \mu}{\sigma} \tag{15}$$

The value of x in the central limit theorem is cleared in Equation (16).

$$x = z\sigma + \mu \tag{16}$$

Table 6. Classification and assessment zones

Output classification	Zones	Probability	Mean	Standard deviation
Very Serious	1	0-0.09	9	1
Moderate	5	0.065-0.09	4	2
		0.061-0.065	7	2
		0.052-0.061	3	2
		0.044-0.052	7	2
		0.000-0.044	9	2
Very Slight	5	0.065-0.09	3	0.75
		0.047-0.065	4	0.75
		0.038-0.047	2.5	0.75
		0.030-0.038	4	0.75
		0.000-0030	1.9	0.75

Source: Authors

Results

The top five causes of accidents are ignorance of safety issues, lack of personal protective equipment, lack of safety measures, unfit equipment, and lack of knowledge and training on equipment. The main types of accidents in the company under study, are related to daily work, material handling, and site conditions.

The company keeps weekly records of the number of accidents. These records were analyzed as empirical collectives to integrate stochastic uncertainty into the proposed system.

The firm specializes in the sale of construction materials, specifically grey, white, and red sand, crushed gravel, *tezontle*, *tepetate*, *confitillo*, and fathom. These materials are delivered to the address indicated by the customer. The main current issues of the company include the lack of control of internal processes and the lack of training for employees regarding material handling and the use of the machinery used. These factors cause various types of accidents and damages to machines and trucks. The owner must incur a maintenance expense, which can result in a large cash outlay, especially with regard to machinery.

Empirical collectives

Each of the accident records related to daily labor, material, and site conditions have been classified as empirical collectives. The value ranges are 0-10 for labor-related accidents, 0-15 for material-related ones, and 0-12 for site condition-related ones (Table 7).

Frequentist probability and risk characterization

The attribute to be analyzed is the weekly relative frequency of the number of accidents of each type over a period of 52 weeks. Table 8 and Figure 12 show the frequentist probability of the number of accidents of each type. A risk characterization based on a defined range of accidents in each category is observed. The range from zero to four represents a low category, the range of five to 11 accidents represents a regular category, and the range of 12 to 20 represents a high-risk category.

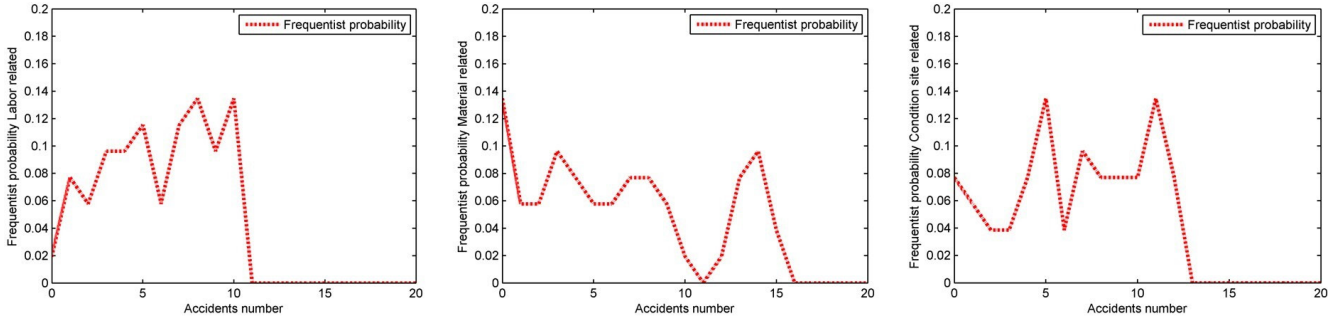


Figure 12. Frequentist probabilities
Source: Authors

Table 7. Accident records

Week	Labor-related	Material-related	Site condition-related	Week	Labor-related	Material-related	Site condition-related	Week	Labor-related	Material-related	Site condition-related
1	1	5	8	**19**	10	9	1	**37**	10	3	9
2	1	9	6	**20**	3	3	12	**38**	10	4	7
3	8	0	8	**21**	7	0	6	**39**	2	6	0
4	9	7	0	**22**	5	13	0	**40**	8	7	3
5	10	14	0	**23**	5	10	11	**41**	3	14	5
6	2	3	2	**24**	8	8	8	**42**	8	0	1
7	6	1	9	**25**	4	7	10	**43**	9	14	11
8	7	2	11	**26**	10	8	12	**44**	9	0	4
9	8	3	5	**27**	9	15	1	**45**	7	8	11
10	5	13	5	**28**	9	12	5	**46**	4	6	11
11	5	2	7	**29**	8	4	7	**47**	4	9	12
12	0	15	7	**30**	5	1	9	**48**	1	5	10
13	10	8	4	**31**	8	4	8	**49**	5	2	11
14	7	4	3	**32**	6	1	5	**50**	10	7	4
15	3	3	2	**33**	4	6	10	**51**	7	13	4
16	7	13	5	**34**	3	5	7	**52**	2	14	10
17	1	0	11	**35**	6	0	9				
18	3	0	5	**36**	4	14					

Source: Authors

Table 8. Frequentist probability

Risk characterization	Accident number	Absolute frequency, labor-related	Relative frequency, labor-related	Absolute frequency, material-related	Relative frequency, material-related	Absolute frequency, site condition-related	Relative frequency, site condition-related
	0	1	0.019230769	7	0.134615385	4	0.07692308
	1	4	0.076923077	3	0.057692308	3	0.05769231
Inferior	2	3	0.057692308	3	0.057692308	2	0.03846154
	3	5	0.096153846	5	0.096153846	2	0.03846154
	4	5	0.096153846	4	0.076923077	4	0.07692308
	5	6	0.115384615	3	0.057692308	7	0.13461538
	6	3	0.057692308	3	0.057692308	2	0.03846154
	7	6	0.115384615	4	0.076923077	5	0.09615385
Regular	8	7	0.134615385	4	0.076923077	4	0.07692308
	9	5	0.096153846	3	0.057692308	4	0.07692308
	10	7	0.134615385	1	0.019230769	4	0.07692308
	11	0	0	0	0	7	0.13461538
	12	0	0	1	0.019230769	4	0.07692308
	13	0	0	4	0.076923077	0	0
	14	0	0	5	0.096153846	0	0
	15	0	0	2	0.038461538	0	0
Superior	16	0	0	0	0	0	0
	17	0	0	0	0	0	0
	18	0	0	0	0	0	0
	19	0	0	0	0	0	0
	20	0	0	0	0	0	0
	Total	**52**		**52**		**52**	

Source: Authors

Linguistic projection

Subsequently, the frequentist probability is projected to three language labels: low, medium, and high. The projection is based on the membership functions proposed in the research. For each risk characterization associated with each language label, a spread is performed based on percentages determined by the experts (Figure 13). Table 9 shows the linguistical projection of labor-related accidents.

$$[x]^{\alpha,\beta}, [Rf]^{\lambda Rf}$$

$$\text{Low} = [I]^{0,4}, [Rf]^{0.7Rf}; [R]^{5,11}, [Rf]^{0.25Rf}; [S]^{12,20}, [Rf]^{0.05Rf}$$

$$\text{Medium} = [I]^{0,4}, [Rf]^{0.05Rf}; [R]^{5,11}, [Rf]^{0.7Rf}; [S]^{12,20}, [Rf]^{0.25Rf}$$

$$\text{High} = [I]^{0,4}, [Rf]^{0.25Rf}; [R]^{5,11}, [Rf]^{0.05Rf}; [S]^{12,20}, [Rf]^{0.70Rf}$$

Fuzzy rules

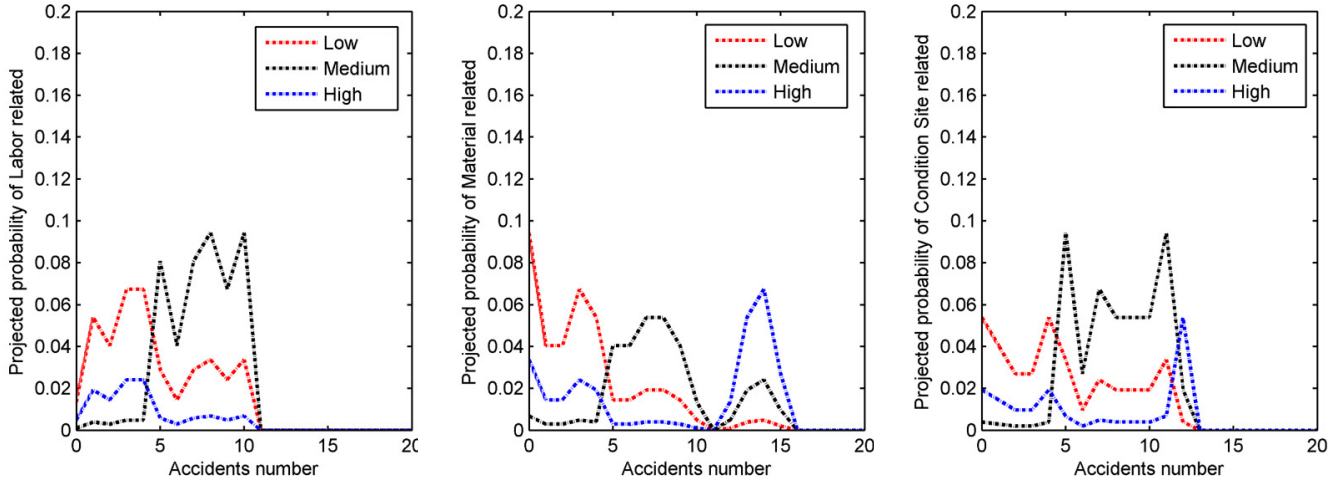
Table 10 shows the 21 fuzzy rules that make up the knowledge base of the proposed system. The output of the system is the classification of a risk as very slight, moderate, and very serious. Seven rules were determined for each label type of the output variable.

Table 10. Knowledge base.

Labor related	Material related	Condition site related	Output classification
HIGH	HIGH	HIGH	VERY SERIOUS
HIGH	HIGH	MEDIUM	VERY SERIOUS
HIGH	MEDIUM	HIGH	VERY SERIOUS
MEDIUM	HIGH	HIGH	VERY SERIOUS
HIGH	HIGH	LOW	VERY SERIOUS
HIGH	LOW	HIGH	VERY SERIOUS
LOW	HIGH	HIGH	VERY SERIOUS
MEDIUM	MEDIUM	MEDIUM	MODERATE
MEDIUM	MEDIUM	HIGH	MODERATE
MEDIUM	HIGH	MEDIUM	MODERATE
HIGH	MEDIUM	MEDIUM	MODERATE
LOW	MEDIUM	MEDIUM	MODERATE
MEDIUM	LOW	MEDIUM	MODERATE
MEDIUM	MEDIUM	LOW	MODERATE
LOW	LOW	LOW	VERY SLIGHT
LOW	LOW	MEDIUM	VERY SLIGHT
LOW	MEDIUM	LOW	VERY SLIGHT
MEDIUM	LOW	LOW	VERY SLIGHT
LOW	LOW	HIGH	VERY SLIGHT
LOW	HIGH	LOW	VERY SLIGHT
HIGH	LOW	LOW	VERY SLIGHT

Source: Authors

Figure 13. Linguistic projection of frequentist probabilities



Source: Authors

Table 9. Linguistic projection of labor related frequentist probability

Test number	Labor-related	Material-related	Site condition-related	Weighted average	System probability	Absolute error	Weighted average	System probability	Efficiency
1	0	1	4	2.3	2.09	0.21	VERY SLIGHT	VERY SLIGHT	1
2	2	4	6	4.6	4.59	0.01	VERY SLIGHT	VERY SLIGHT	1
3	6	5	8	6.7	6.68	0.02	MODERATE	MODERATE	1
4	4	7	12	8.9	8.86	0.04	MODERATE	SERIOUS	0
5	8	6	4	5.4	5.04	0.36	MODERATE	MODERATE	1
6	10	4	2	4.2	4.59	0.39	VERY SLIGHT	VERY SLIGHT	1
7	5	7	1	3.6	3.64	0.04	MODERATE	MODERATE	1
8	5	5	5	5	5.04	0.04	MODERATE	MODERATE	1
9	10	7	5	6.6	6.7	0.1	MODERATE	MODERATE	1
10	12	11	5	8.2	9	0.8	VERY SERIOUS	VERY SERIOUS	1
11	9	10	7	8.3	8.53	0.23	MODERATE	MODERATE	1
12	6	11	10	9.5	9.36	0.14	MODERATE	MODERATE	1
13	5	8	7	6.9	6.56	0.34	MODERATE	MODERATE	1
14	2	4	6	4.6	4.59	0.01	VERY SLIGHT	VERY SLIGHT	1
15	5	7	7	6.6	6.56	0.04	MODERATE	MODERATE	1
16	12	8	9	9.3	9.38	0.08	MODERATE	MODERATE	1
17	8	7	10	8.7	8.07	0.63	VERY SERIOUS	VERY SERIOUS	1
18	10	10	6	8	7.9	0.1	MODERATE	MODERATE	1
19	2	7	4	4.5	4.56	0.06	VERY SLIGHT	VERY SLIGHT	1
20	8	8	8	8	8.07	0.07	MODERATE	MODERATE	1
21	2	3	5	3.8	3.59	0.21	VERY SLIGHT	VERY SLIGHT	1
22	0	2	3	2.1	1.57	0.53	VERY SLIGHT	VERY SLIGHT	1
23	0	1	3	1.8	1.57	0.23	VERY SLIGHT	VERY SLIGHT	1
24	1	3	3	2.6	2.16	0.44	VERY SLIGHT	VERY SLIGHT	1
25	2	6	4	4.2	454	0.34	VERY SLIGHT	VERY SLIGHT	1
26	3	6	5	4.9	5.06	0.16	MODERATE	MODERATE	1

Source: Authors

Implication: Total probability theorem

Test point

Below is a test point of the system's training. The records to be evaluated for each type of accident are as follows: labor-related: 18, material-related: 1, site condition-related: 12. Figure 14 shows the inference values using the total probability theorem for each of the 21 fuzzy rules.

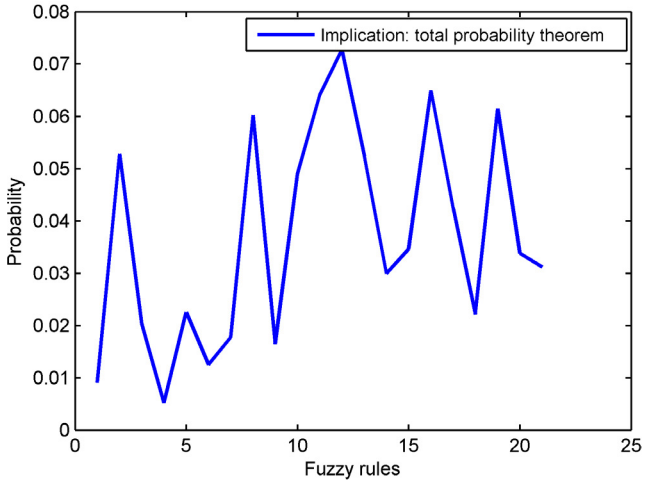


Figure 14. Individual evaluation, inference system
Source: Authors

Aggregation

The maximum value in the implication phase is 0,039, corresponding to fuzzy rule 6, with a *very serious* output label. This value represents the possibility of occurrence for the training point (Figure 14).

Classification: very serious, probability: 3.9%.

Risk assessment

Finally, in the assessment stage, the central limit theorem is used to obtain the corresponding evaluation. The maximum value is scaled with a factor of 10, to be used in Equation (17), since the classification is very serious and there is only one assessment zone with average parameters equal to 9 and a standard deviation 1.

$$x = (z * \sigma) + \mu = (0.39 * 1) + 9 = 9.39 \qquad (17)$$

Figure 15 shows the result of the training point. Evaluation: 9.39.

Comparison of results

Below are 26 system training points. Our comparison with the traditional method is based on a weighted average. The results are reliable, and the efficiency of the system is 96.15%, classifying 25 of 26 training patterns correctly and constitute a robust, reliable, and simple system (Table 11).

The classification made by experts based on the weighted average is flexible with regard to the interpretation of the output classification. Although the weighted average value is 8.9, it was assigned a moderate classification, considering that the first two types of accidents have low or moderate values, preserving subjectivity in their output classification. The proposed model eliminates subjectivity and is based on the integration of stochastic uncertainty by using the relative frequencies of the accidents that occurred, as well as on that of linguistic uncertainty when projecting frequencies based on risk categorization. This was the only pattern that showed variations in the proposed probabilistic risk assessment system, with a reliability of 96,15%.

Figure 16 shows the results of the implication stage for the 26 test points.

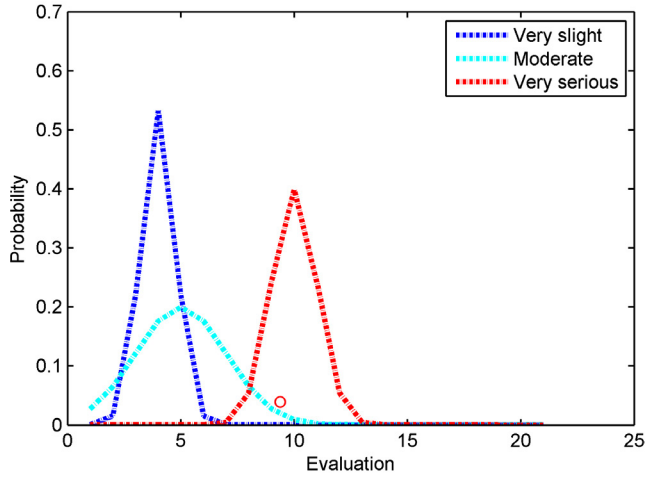


Figure 15. System output
Source: Authors

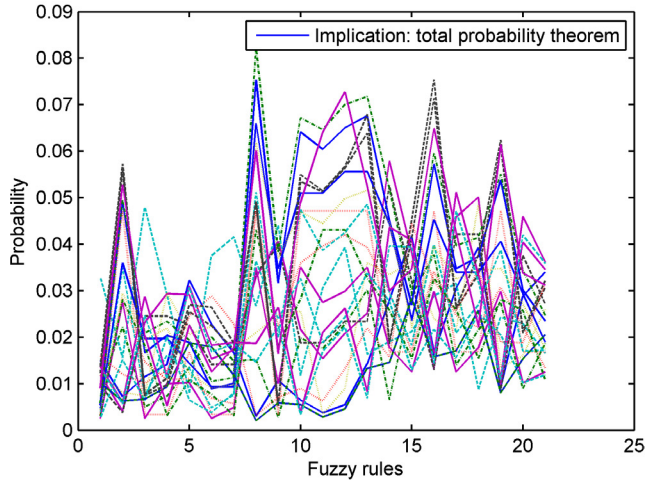


Figure 16. Total evaluation, inference system.
Source: Authors

Figure 17 shows the inference of the proposed system for the 26 test points.

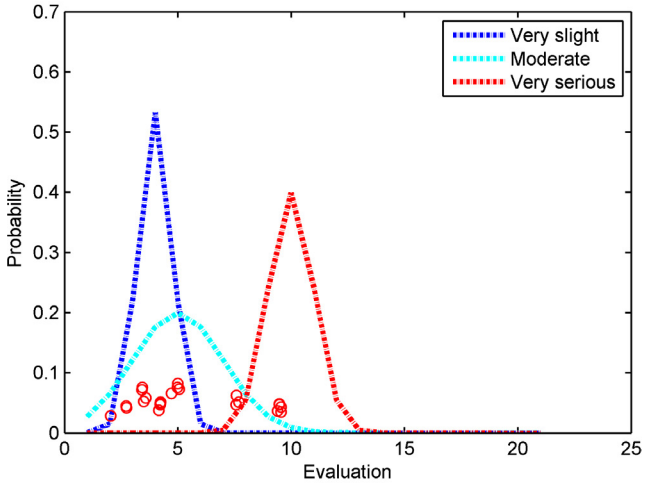


Figure 17. System output
Source: Authors

The results of the proposed system and the weighted average benchmark are compared in Figure 18.

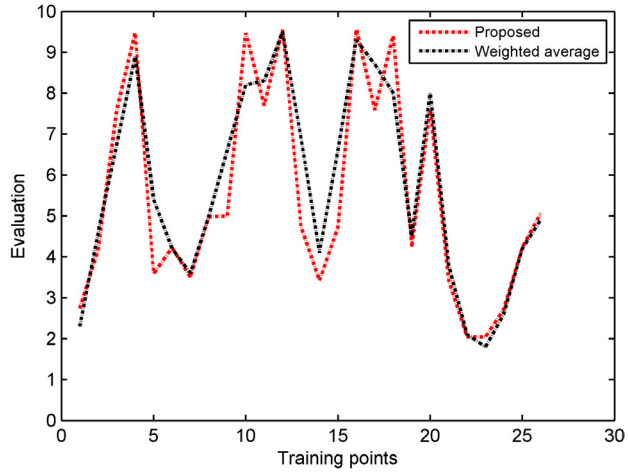


Figure 18. Comparative results
Source: Authors

Table 11. Comparison of results

Test number	Labour related	Material related	Site condition related	Weighted average	Probability system	Absolute error	Weighted average	Probability system	Efficiency
1	0	1	4	2.3	2.09	0.21	VERY SLIGHT	VERY SLIGHT	1
2	2	4	6	4.6	4.59	0.01	VERY SLIGHT	VERY SLIGHT	1
3	6	5	8	6.7	6.68	0.02	MODERATE	MODERATE	1
4	4	7	12	8.9	8.86	0.04	MODERATE	SERIOUS	0
5	8	6	4	5.4	5.04	0.36	MODERATE	MODERATE	1
6	10	4	2	4.2	4.59	0.39	VERY SLIGHT	VERY SLIGHT	1
7	5	7	1	3.6	3.64	0.04	MODERATE	MODERATE	1
8	5	5	5	5	5.04	0.04	MODERATE	MODERATE	1
9	10	7	5	6.6	6.7	0.1	MODERATE	MODERATE	1
10	12	11	5	8.2	9	0.8	VERY SERIOUS	VERY SERIOUS	1
11	9	10	7	8.3	8.53	0.23	MODERATE	MODERATE	1
12	6	11	10	9.5	9.36	0.14	MODERATE	MODERATE	1
13	5	8	7	6.9	6.56	0.34	MODERATE	MODERATE	1
14	2	4	6	4.6	4.59	0.01	VERY SLIGHT	VERY SLIGHT	1
15	5	7	7	6.6	6.56	0.04	MODERATE	MODERATE	1
16	12	8	9	9.3	9.38	0.08	MODERATE	MODERATE	1
17	8	7	10	8.7	8.07	0.63	VERY SERIOUS	VERY SERIOUS	1
18	10	10	6	8	7.9	0.1	MODERATE	MODERATE	1
19	2	7	4	4.5	4.56	0.06	VERY SLIGHT	VERY SLIGHT	1
20	8	8	8	8	8.07	0.07	MODERATE	MODERATE	1
21	2	3	5	3.8	3.59	0.21	VERY SLIGHT	VERY SLIGHT	1
22	0	2	3	2.1	1.57	0.53	VERY SLIGHT	VERY SLIGHT	1
23	0	1	3	1.8	1.57	0.23	VERY SLIGHT	VERY SLIGHT	1
24	1	3	3	2.6	2.16	0.44	VERY SLIGHT	VERY SLIGHT	1
25	2	6	4	4.2	4.54	0.34	VERY SLIGHT	VERY SLIGHT	1
26	3	6	5	4.9	5.06	0.16	MODERATE	MODERATE	1

Source: Authors

Conclusions

Our research proposed a novel global probabilistic fuzzy system for occupational risk assessment. Stochastic uncertainty conditions were integrated, using the frequentist probability of the number of accidents of each type as input variables. Non-stochastic uncertainty was incorporated by classifying the input variable via language labels in a clear and simple language (*i.e.*, high, medium, and low) based on expert interpretation.

The accidents assessed were labor-related, material-related, and site condition-related in a construction company dedicated to the sale of materials.

The conceptualization was based on the analysis of accident records as empirical collectives, where the frequentist probability could be determined for different numbers of identified accidents. A crucial phase in the methodology was to identify the risk category based on parameters corresponding to expert interpretation.

One of the scientific contributions to diffuse probabilistic systems is the projection of frequentist probabilities in three language categories, which are based on the definition of percentages related to the heuristic knowledge of experts and constitute three weighted membership functions. 21 fuzzy rules were defined to infer the result. Another important contribution is the use of the total probability theorem in the implication stage as the inference mechanism of the output variable. In the aggregation stage, the maximum fuzzy operator was used to estimate the probability of occurrence of each type of accident and its classification. The central limit theorem was used in the risk assessment stage. Different zones were proposed for each classification of the output variable. For the *very serious* classification, there is only one zone, and, for the *moderate* and *very slight* classifications, there are five zones. The main parameters of the mean and the standard deviation were determined for each zone. Finally, a benchmarking comparison was performed via a simple weighted average technique used and interpreted by experts. The results show the robustness and reliability of the proposed system.

It should be noted that the heuristic knowledge of experts was used to determine percentages in the frequentist probability breakdown. It was also used in determining the ranges of each risk category.

The proposal for a diffuse probabilistic system can be replicated in any production or service sector, as well as in organizations of any size.

As future work, a novel alternative will be presented for the involvement stage, aiming to group the output categories objectively and not subjectively.

CRedit author statement

All authors have participated in a) conception and design or data analysis and interpretation; (b) drafting the article or revising it critically for important intellectual content; and (c) approving of the final version of the manuscript.

References

Aggarwal, M. (2021). Fuzzy entropy with a general framework. *Experts Systems with Applications, 164*, 113671. https://doi.org/10.1016/j.eswa.2020.113671

Boyaci, A., and Selim A. (2022). Assessment of occupational health and safety risks in a Turkish public hospital using a two-stage hesitant fuzzy linguistic approach. *Environmental Science and Pollution Research, 29*, 36313–36325. https://doi.org/10.1007/s11356-021-18191-x

De, A., Dasb, S., and Karc, S. (2019) Multiple attribute decision making based on probabilistic interval-valued intuitionistic hesitant fuzzy set and extended TOPSIS method, *Journal of Intelligent & Fuzzy Systems, 37*, 5229-5248. https://doi.org/10.3233/JIFS-190205

De Ridder, S. Toscani, N., and Vande, D. (2020). Machine-learning_based hybrid random fuzzy uncertainty quantification for EMC and SI assessment. *IEEE Transactions on Electromagnetic Compability, 62*(6), 2538-2546. https://doi.org/10.1109/TEMC.2020.2980790

Farhadinia, B. Aickelin, U., and Akbarzadeh, H. (2020). Uncertainty measures for probabilistic hesitant fuzzy sets in multiple criteria decision making. *International Journal of Intelligent Systems, 35*(11), 1646-1679. https://doi.org/10.1002/int.22266

Fattahi, R. Tavakkoli-Moghaddam, R., Khalilzaded, M, Shahsavari-Pour, N., and Soltani, R. (2020). A novel FMEA model based on fuzzy multiple-criteria decision makin methods for risk assessment. *Journal of Enterprise Information Management, 33*(5), 881-904. https://doi.org/10.1108/JEIM-09-2019-0282

Gül, M. (2021). A quantitative occupational risk assessment methodology based on Topsis-Sort with its application in aluminum extrusion industry. *International Journal of Pure and Applied Sciences, 7*(1), 163-172. https://doi.org/10.29132/ijpas.943612

Gül, M., and Celik, E. (2018) Fuzzy rule-based Fine–Kinney risk assessment approach for rail transportation systems. *Human and Ecological Risk Assessment: An International Journal, 24*(7), 1786-1812. https://doi.org/10.1080/10807039.2017.1422975

Gupta, M. Srivastava, S., Chaudhary, G., and Parra, J. (2020). Voltage regulation using probabilistic and fuzzy controlled dynamic voltage restorer for oil and gas industry. *Internatinal Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 28*(2), 49-64. https://doi.org/10.1142/S0218488520400139

Huang, W-J. Li, Y-H., and Xu, K-K. (2019). The general probabilistic fuzzy set for modelling and its application in EMG robots. *Journal of Intelligent & Fuzzy Systems, 37*, 2987-2100. https://doi.org/10.3233/JIFS-171121

Jiang, L., and Liao, H. (2021). Network consensus analysis of probabilistic linguistic preference relations for group decision making and its application in urban household waste classification. *Journal of Cleaner Production, 278*, 122766. https://doi.org/10.1016/j.jclepro.2020.122766

Khan, M. W., Ali, Y., Felice, F. D., and Petrillo, A. (2019). Occupational health and safety in construction industry in Pakistan using modified-SIRA method. *Safety Science*, 118, 109-118. https://doi.org/10.1016/j.ssci.2019.05.001

Li, H-X., and Liu, Z. (2009). *A probabilistic fuzzy logic system: Learning in the stochastic environment with incomplete dynamics* [Conference paper]. 2009 IEEE International Conference on Systems, Man and Cybernetics, San Antonio, TX, USA.

Liang, P., Hu, J., Li, B., Liu, Y., and Chen, X. (2020). A group decision making with probability linguistic preference relations based on nonlinear optimization model and fuzzy cooperative games. *Fuzzy Optimization and Decision Making, 19*, 499-528. https://doi.org/10.1007/s10700-020-09329-6

Liang, W., Goh, M., and Wang, Y.-M. (2020). Multi-attribute group decision making method based on prospect theory under hesitant probabilistic fuzzy environment. *Computers & Industrial Engineering, 149*, 106804. https://doi.org/10.1016/j.cie.2020.106804

Modares, M., and Desch, M. (2021). Comparison between probabilistic and possibilistic approaches for structural uncertainty analysis. *Practice Periodical on Structural Design and Construction, 26*(2), 04020070. https://doi.org/10.1061/(ASCE)SC.1943-5576.0000556

Sozhamadevi, N., Sagaya R., and Sathiyamoorthy, S. (2012). *Design and implementation of probabilistic fuzzy logic control system* [Conference paper]. 2012 International Conference on Emerging Trends in Science, Engineering and Technology (INCOSET), Tiruchirappalli, India. https://doi.org/10.1109/INCOSET.2012.6513959

Sozhamadevi, N., and Sathiyamoorthy, S. (2015). Modeling and control of an unstable system using probabilistic fuzzy inference system. *Archives of Control Sciences, 25*(3), 377-396. https://doi.org/10.1515/acsc-2015-0025

Tang, J., Liu, X., and Wang, W (2021) A hybrid risk prioritization method based on generalized TODIM and BWM for Fine-Kinney under interval type-2 fuzzy environment. *Human and Ecological Risk Assessment: An International Journal, 27*(4), 954-979. https://doi.org/10.1080/10807039.2020.1789840

Tian, F., Zhang, M., Zhou, L., Zou, H., Wang, A., and Hao, M. (2018). Qualitative and quantitative differences between common occupational health risk assessment models in typical industries. *Journal of Occupational Health, 60*, 337-347. https://doi.org/10.1539/joh.2018-0039-OA

Wang, Y., Li, L., and Wang, K. (2021). An online operating performance evaluation approach using probabilistic fuzzy theory for chemical processes with uncertainties. *Computers & Chemical Engineering, 144*, 107156. https://doi.org/10.1016/j.compchemeng.2020.107156

Xu, Q., Xu, F., Li, F., Zou, H., Zheng, K., and Zhang, M. (2020). Quantitative differences between common occupational health risk assessment models. *Journal of Occupational Health, 62*, e12164. https://doi.org/10.1002/1348-9585.12164

Xue, Z., Zhao, L-P., Zhang, M., and Sun, B-X. (2020). Three-way decisions based on multi-granulation support intuitionistic fuzzy probabilistic rough sets. *Journal of Intelligent & Fuzzy Systems, 38*, 5013-5031. https://doi.org/10.3233/JIFS-191657

Research Article/Mechanical Engineering, Mechatronics, and Materials Science http://doi.org/10.15446/ing.investig.106383

Determining the Effect of Photovoltaic Module Surface Temperature on Generation Efficiency

Determinación del efecto de la temperatura superficial de los módulos fotovoltaicos en su eficiencia de generación

Cengiz Karaca[1], and Seren Yaver[2]

ABSTRACT

It is imperative to consider the environmental impact of energy production and its cost in deciding how to meet future energy needs. In this regard, it is possible to harness the power of the sun by using photovoltaic (PV) cells. However, when the temperature of a PV cell increases, its generation efficiency is negatively affected. The open-circuit voltage of PV modules is the most sensitive parameter to temperature changes. As the temperature rises, this parameter decreases, and the short-circuit current increases. The circuit's resistance also rises as the electrons' speed is reduced. Temperature also affects the lifespan of PV cells, increasing the rate of thermal decay in their materials. On the other hand, when solar radiation is absorbed at lower temperatures, the system's efficiency, power capacity, and useful life increase. PV module surface temperatures can be reduced in a variety of ways, e.g., the surface can be cooled using water. This work studied hybrid PV-thermal modules under the climate conditions of the Hatay province (Turkey) in order to assess the effect of water cooling on their generation efficiency. The results allow stating that up to 52.6% more electricity can be generated by cooling the module's surface. Additionally, it was found that, in order for PV modules to perform efficiently in Hatay's climate, they must operate at a maximum surface temperature of 55 ˚C.

Keywords: solar PV-T module, PV surface temperature, PV efficiency

RESUMEN

Es imperativo considerar el impacto ambiental de la producción de energía y su costo al decidir cómo satisfacer las necesidades energéticas futuras. A este respecto, es posible aprovechar el poder del sol utilizando células fotovoltaicas (PV). Sin embargo, cuando la temperatura de una célula PV aumenta, su eficiencia de generación se ve negativamente afectada. El voltaje en circuito abierto de los módulos PV es el parámetro más sensible a los cambios de temperatura. A medida que la temperatura aumenta, este parámetro disminuye, y la corriente de cortocircuito aumenta. La resistencia del circuito también se eleva a medida que la velocidad de los electrones se reduce. La temperatura también afecta la vida útil de las células PV, incrementando la tasa de degradación térmica en sus materiales. Por otro lado, cuando la radiación solar se absorbe a temperaturas más bajas, la eficiencia del sistema, la capacidad de potencia y la vida útil aumentan. Las temperaturas superficiales de los módulos PV pueden reducirse de varias maneras, e.g., la superficie puede enfriarse utilizando agua. Este trabajo estudió módulos híbridos PV-térmicos bajo las condiciones climáticas de la provincia de Hatay (Turquía) con el fin de evaluar el efecto del enfriamiento con agua en su eficiencia de generación. Los resultados permiten afirmar que se puede generar hasta un 52.6 % más de electricidad enfriando la superficie del módulo. Además, se encontró que, para que los módulos PV funcionen eficientemente en el clima de Hatay, deben operar a una temperatura superficial máxima de 55 ˚C.

Palabras clave: módulo solar PV-T, temperatura superficial PV, eficiencia PV

Received: December 12th, 2022
Accepted: January 19th, 2024

Nomenclature

P_e: Electrical power (W)
V_i: Instant voltage value (V)
I_i: Instant current value (A)
η_e: Electric conversion efficiency (decimal)
G: Solar radiation on the normal surface (W/m^2)
A_p: Module surface area (m^2)

Introduction

Energy, a key factor for human socio-economic and sustainable development, as well as for improving quality of life, is one of the most important resources in sustaining our lives (Meinshausen *et al*., 2009). Although the world has been seriously harmed as a result of the misuse of natural resources, there is an ongoing search for ways to obtain clean and sustainable energy (Zanlorenzi *et al*., 2018), as the majority of today's energy comes from fossil fuels and nuclear sources.

[1] Professor, Department of Biosystems Engineering, Faculty of Agriculture, Hatay Mustafa Kemal University, Hatay, Türkiye. Email: ckaraca@mku.edu.tr

[2] MSc in Biosystems Engineering, Department of Biosystems Engineering, Faculty of Agriculture, Hatay Mustafa Kemal University, Hatay, Türkiye. Email: serenyaver@hotmail.com

Solar generation is popular, allowing the energy provided by the sun to be used in different ways by converting it into electrical power (Tiwari *et al*., 2006). Solar cells operate on the photovoltaic (PV) principle for power generation, *i.e.*, they generate electricity when they receive light. When solar radiation in the form of photons hits the surface of semiconductor materials such as silicon, electrons are released from atoms, entailing an electric voltage. Solar energy is free and available in abundance. Many applications based on this resource have been developed, including different types of concentrated solar thermal power technologies, solar PV generation, hybrid solar PV and thermal techniques, solar desalination, solar hydrogen production, solar-assisted heat pump technologies, *etc.* (Wang *et al*., 2021).

It is the type of PV employed that has the most bearing on electrical performance. Power generation from solar cells ranges from 6 to 20% depending on weather conditions and the type of cells used in a typical PV module. The remaining incident solar radiation is transformed into heat, which raises the temperature of the PV module and reduces its efficiency (Dubey *et al*., 2013). As PV cells have spectrum-dependent characteristics, ineffective solar radiation increases the cell temperature after it is absorbed, causing a decrease in photoelectric conversion efficiency. Solar PV-thermal (PV-T) technologies employing beam splitting can solve this problem to a certain degree (Wang *et al*., 2023). PV-T collectors can be used to gather this heat from flowing water or air beneath the PV module.

Increases in intrinsic carrier concentrations at higher temperatures lead to increases in the dark saturation current of the p-n junction, which reduces PV cell efficiency. Due to excessive doping, the intrinsic carrier concentration rises as a result of a reduced band gap. The open-circuit voltage drops linearly as the dark saturation current rises, resulting in a temperature change of 2.3 mV/°C for silicon at 300 K, as stated by Peng *et al*. (2017).

At temperatures ranging from 20 to 100 °C, it is estimated that the reduction in the solar cell's band gap increases the short circuit current by 0.1%. Despite this increase in current, reductions in the open-circuit voltage cause a noticeable decrease in the maximum electrical power, as seen in the characteristic curves of PV modules at different operating temperatures (Figure 1) (Andreev *et al*., 1997).

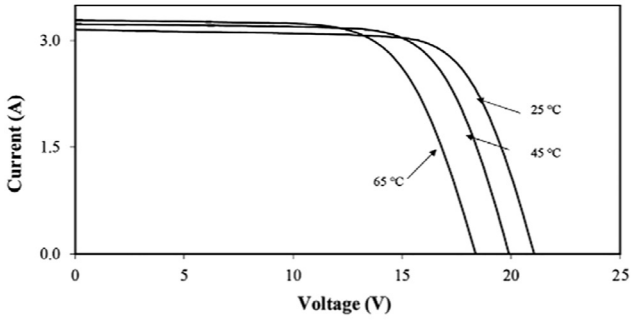


Figure 1. Influence of temperature on a PV module I-V curve
Source: Andreev *et al*. (1997)

A study by Radziemska (2003) determined that the power and voltage produced by monocrystalline silicon solar cells and the maximum power vary at different temperatures (Figure 2).

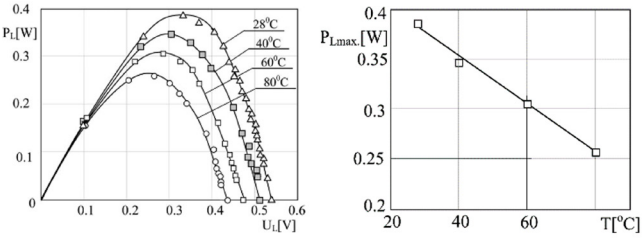


Figure 2. a) Voltage-output power and b) maximum output power-temperature relationships in a monocrystalline silicon cell at different surface temperatures
Source: Radziemska (2003)

One of the advantages of using a solar PV and thermal hybrid system is that it simultaneously generates electricity and thermal energy from the sun. The use of PV-T systems leads to the production of both useful heat energy and electrical power. The solar PV cells currently on the market have a comparatively low efficiency (less than 15%). After PV conversion, more than 80% of the solar energy absorbed is lost as heat into the environment. Cooling PV systems is important, as their efficiency decreases when the temperature goes above a certain point. In order to boost performance, it is critical to keep PV modules cool (Rawat and Dhiran, 2017).

While the short-circuit current of PV cells increases with temperature, the open-circuit voltage decreases. The resistance that reduces the speed of electrons in the circuit also increases with temperature. Therefore, high operating temperatures have a negative impact on power and efficiency in PV systems. With an effect on the thermal degradation of the module's material, temperature also reduces the effective efficiency of the cells. Therefore, lowering the operating surface temperature is an effective way to increase the efficiency of a PV module and reduce its thermal degradation rate. This can be achieved by cooling the module and lowering the heat stored in the PV cells during operation. There are several methods for reducing PV module surface temperatures, including water cooling systems.

The goal of this study is to determine how the efficiency of PV modules changes when cooled via the liquid fluid approach in a PV/T system in the province of Hatay. We expect our results to aid in determining the surface temperature values for the most efficient operation of PV systems, as well as in identifying the efficiency losses regarding electricity production that are caused by temperature in similar hot climate conditions. It is also stated in the studies mentioned above that the panel surface temperature affects the electricity generation efficiency of PV systems. This study aims to determine the extent of this effect and the efficient operating temperature ranges to be used in similar regions. In this vein, a PV-T experimental setup, which can

be operated with or without water cooling, is used to take measurements.

Methodology

Materials

Today, hybrid PV-T solar systems are one of the most popular methods for cooling PV modules. These hybrid systems consist of solar PV modules combined with a cooling system. To cool the surface of the PV modules, water or air is pumped over them, boosting their efficiency.

In the experimental setup, (except PV modules) parts such as a pump circulating the cooling water in the cooling system, a water tank, a solar radiation meter, temperature sensors, data recorders, current and voltage recorders, a charge controller, a light bulb, a mini meteorological station, PE pipe components were used (Figure 3).

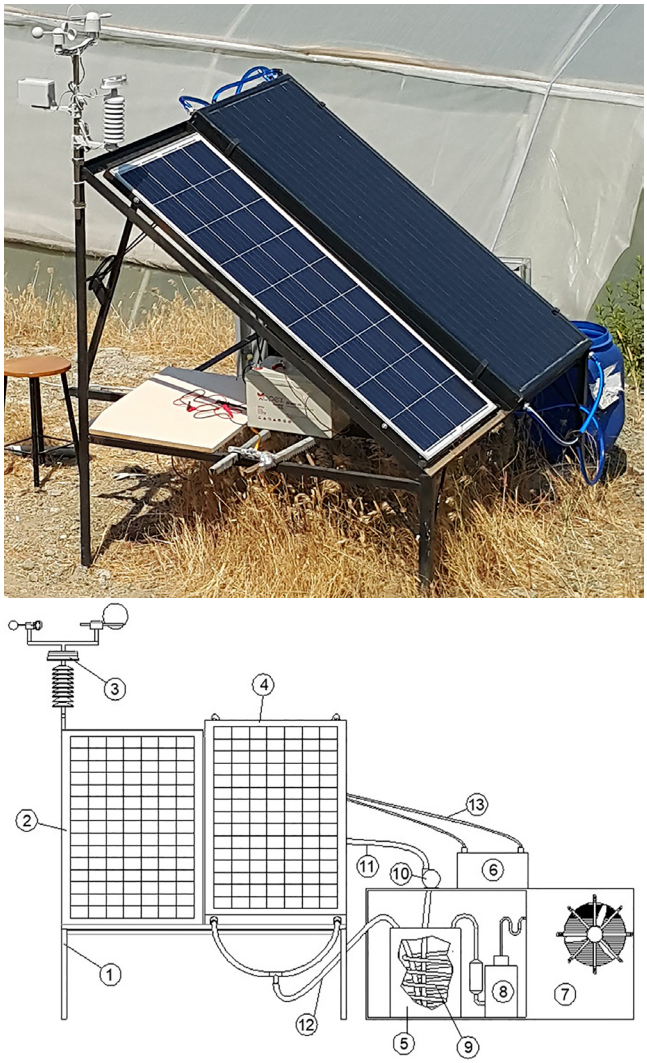


Figure 3. Experimental setup and its parts: 1) scaffold, 2) PV module, 3) weather station, 4) PV-T module, 5) water tank, 6) battery, 7) fan, 8) compressor, 9) serpentine, 10) pump, 11 and 12) PE pipes
Source: Authors

A PV/T module was used, whose features are presented below (Table 1, Figure 4).

Table 1. PV-T module specifications

Properties	**Values**
Dimensions (mm)	1601 x 828 x 90
Mass (kg)	24.4
Nominal power (W)	190
Specific power (W/m^2)	143.4
Nominal current (A)	5.2
Short circuit current (I_{SC})(A)	5.6
Nominal voltage (V)	36.4
Open circuit voltage (V_{OC})(V)	45.2
Module	monocrystalline

Source: Authors

The PV-T module was cooled using water, which was in turn cooled with a compressor cooling system. An evaporator in the form of a spiral copper pipe immersed in the tank was entrusted with water cooling. The cooling system was managed with a digital temperature controller that can be set to a minimum temperature of 18 ˚C (Figure 3, no. 5, 7, 8, 9).

The measurement involved a PWM charge controller; a 12 V, 100 Ah gel-type battery to measure the power produced by the system; and a mini meteorological station to record climate data including wind speed and direction, temperature (ambient, perceived, and dew), and relative and absolute pressure. These data were measured in 5 min intervals (Figure 3, no. 3).

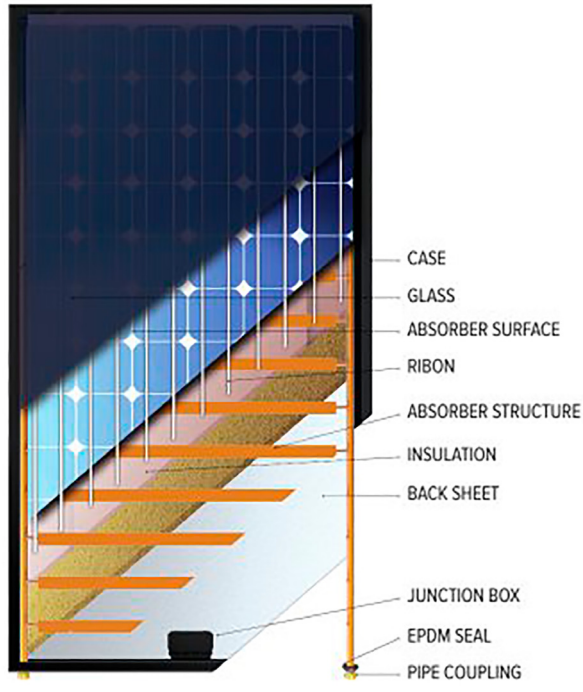


Figure 4. PV/T module
Source: Solimpeks (n.d.)

Method

The electricity generation efficiency of the PV/T module was determined in two ways: with and without cooling. To this effect, as shown in Figure 5, the current and voltage values produced by the module were measured in 5 min intervals by connecting two multimeters. In both experiments, in addition to these measurements, module surface temperature, solar radiation levels, and climate parameters were measured. Measurements were taken between 10:00 a.m. and 3:00 p.m., when the sun was at its most intense.

The water cooled via the compressor system was used to reduce the temperature of the PV/T module. With the help of a 12/24 V DC pump, water at a temperature of around 20 °C was pushed through the PV/T module and discharged back into the water tank, providing perpetual circulation.

Moreover, for the sake of comparison, the current, voltage, and surface temperature values of the PV module were measured in the experimental setup.

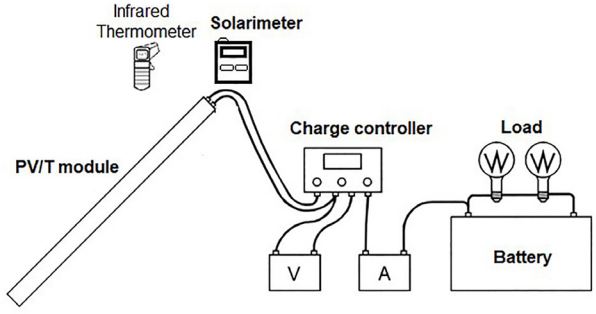


Figure 5. Connection diagram of the experimental setup's measurement system
Source: Authors

The power produced by the solar modules was calculated using Equations (1) and (2), aiming to determine the efficiency of the modules. The power consumed by the cooling water system was not considered.

$$P_e = I_i V_i \quad (1)$$

$$\eta_e = \frac{P_e}{G\,A_p} \quad (2)$$

The results of these measurements were graphed. This includes the evolution of the surface temperature of the PV-T module with and without cooling as well as the module's efficiency, aiming to determine their relationship. The temperature changes were calculated by taking the differences between the two measurement points, and the efficiency change was determined by calculating the percentage of the difference between the two measurement points.

Results

Current-voltage, solar radiation-module power, and temperature-efficiency graphs for a day of operating the PV-T cooling module. A total of six days was plotted. The data for June 24th, 2019, are presented in Figures 6, 7, and 8.

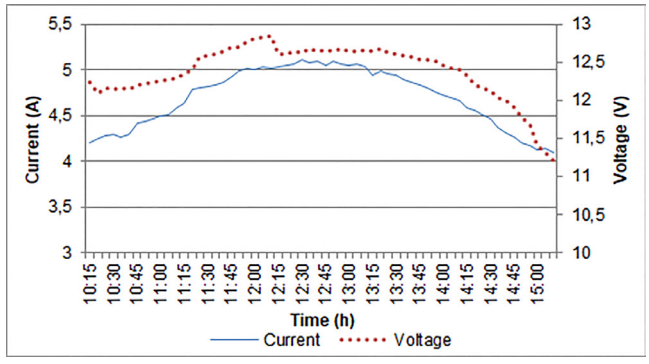


Figure 6. PV-T module with cooling, current-voltage graph (24 June 2019)
Source: Authors

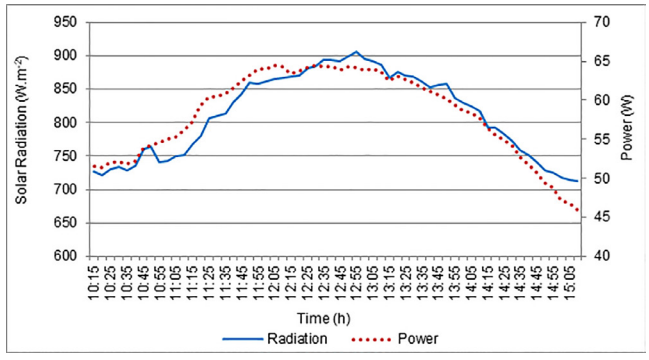


Figure 7. PV-T module with cooling, solar radiation-module power generation graph (24 June 2019)
Source: Authors

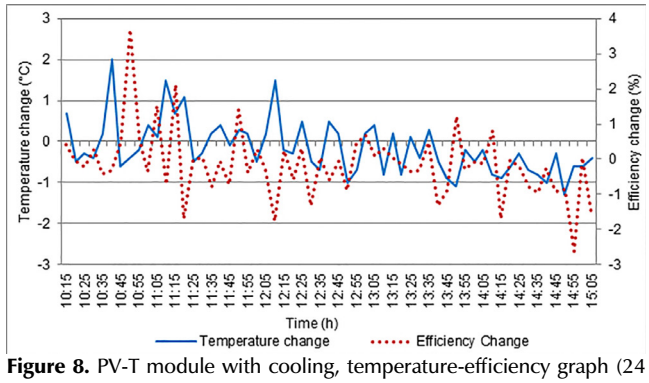


Figure 8. PV-T module with cooling, temperature-efficiency graph (24 June 2019)
Source: Authors

The results showed that the current ranges from 4.0 to 5.11 A, while the voltage ranges from 11.0 to 12.90 V. It was determined that the highest current and voltage values of the PV-T module are reported between 11:45 a.m. and 1:30 p.m. (Figure 6).

Regarding the solar radiation-module power graph, it was observed that the radiation ranges from 711.6 to 905.2 W/

m², and the electrical power generated ranges from 45.92 to 64.53 W. The highest values occurred between 11:45 a.m. and 1:30 p.m. (Figure 7).

The module's surface temperature value ranged from 46.0 to 60.7 °C, and its electricity generation efficiency ranged from 4.98 to 5.84%. At many measurement points, temperature increases caused a decrease in efficiency (Figure 8). The temperature between the measurement points went from -1.3 to 2.0 °C, and the electricity generation efficiency of the module increased when surface temperature decreased. According to the calculations, the module's efficiency varied between -2.63 and +3.69% with regard to variations in surface temperature.

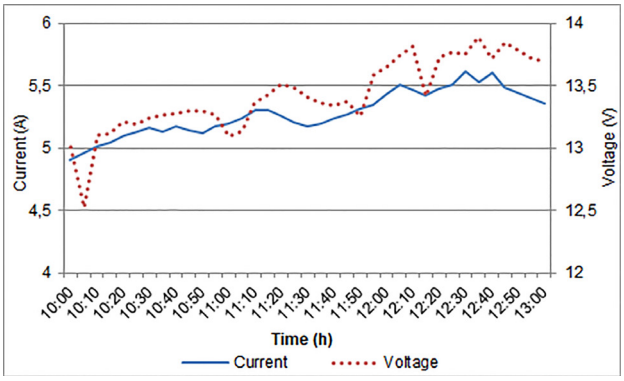


Figure 9. PV-T module without cooling, current-voltage graph (30 June 2019)
Source: Authors

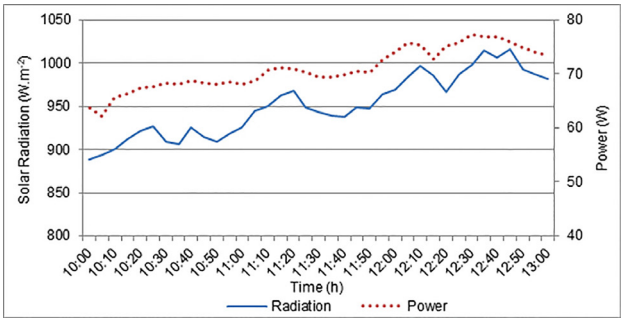


Figure 10. PV-T module without cooling, solar radiation-module power generation graph (30 June 2019)
Source: Authors

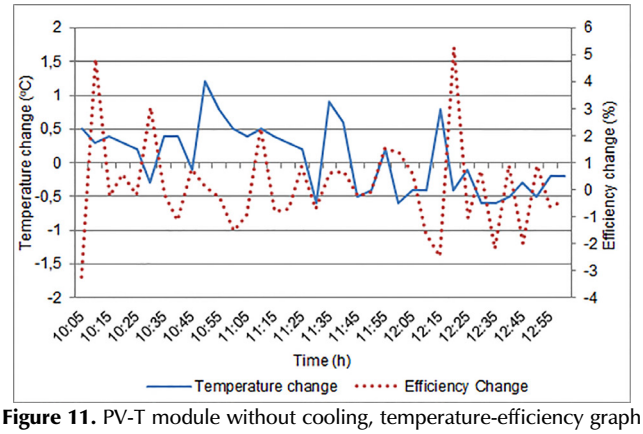


Figure 11. PV-T module without cooling, temperature-efficiency graph (30 June 2019)
Source: Authors

On June 30th, 2019, the current ranged from 4.9 to 5.6 A, while the voltage went from 12.52 to 13.89 V. The highest current and voltage values of the PV-T module occurred between 11:45 a.m. and 1:30 p.m. (Figure 9).

Solar radiation ranged from 888.3 to 1016 W/m², and power generation ranged from 62.1 to 77.3 W. The highest values were reported between 11:45 a.m. and1:30 p.m. (Figure 10).

The module surface temperature value ranged from 46.0 to 60.7 °C, and the electricity generation efficiency went from 4.3 to 6.0%. At many measurements points, temperature increases also caused a decrease in efficiency (Figure 11). Temperature changes between measurement points ranged from -0.6 to 1.2 °C. The efficiency of the module increased when the surface temperature decreased. According to the calculations, the changes in generation efficiency are between -3.2 and +5.29% with regard to surface temperature.

The overall variation ranges for all days of measurement are presented in Table 2.

Table 2. Variations in temperature and efficiency in the system with and without cooling

	With cooling	Without cooling
Highest temperature change (°C)	4,20	2,60
Lowest temperature change (°C)	-5,20	-2,10
Highest efficiency change (%)	3,54	0,55
Lowest efficiency change (%)	-3,45	-0,57

Source: Authors

For all measurements, the temperature change interval of the system without cooling was 7.7 °C, and that of the system with cooling was 9.2 °C. In light of this, the efficiency change intervals were 1.12 and 6.99%, respectively. Based on these results, it can be stated that an efficiency increase and decrease of 3.5% is generated by changes in the module surface temperature.

Table 3 presents the descriptive statistics of the module surface temperature and efficiency values for all measurement days in both scenarios (with cooling and without cooling).

Table 3. Temperature and efficiency values with and without cooling

Descriptive statistics	With cooling		Without cooling	
	Temperature (°C)	Efficiency (%)	Temperature (°C)	Efficiency (%)
Minimum	33.30	4.42	52.80	4.52
Maximum	61.50	9.31	75.20	6.10
Mean	52.02	5.95	64.28	5.57
Standard error	0.381	0.056	0.334	0.022

Source: Authors

For the PV-T module without cooling, the module surface temperature reported values of up to 75.20 °C. With cooling, the lowest value was 33.30 °C. The efficiency of the system without cooling reached a maximum of 6.10% depending on the surface temperature. Moreover, the highest efficiency value was 9.31% with cooling. By comparing both scenarios, an efficiency increase of 52.62% was observed with the use of the cooling system.

Several studies have been carried out in various applications to reduce the surface temperature of PV panels. Some of these are reviewed below. These works have aimed to increase panel efficiency and generation by reducing the surface temperature.

The efficiency increase achieved with our cooling system is similar to that of Akbarzadeh and Wadowski (1996), who measured water cooling in a hybrid PV-T solar system, with the purpose of increasing the output power by almost 50%. They also determined that the module's cooling process prevents the surface temperature of the solar cells from going above 46 °C for a period of 4 h.

As reported by Teo *et al*. (2012), when a module with no active cooling only exhibits an 8-9% efficiency. However, when it is cooled, the temperature drops significantly, and its efficiency increases to 12-14%.

Abu-Rahmeh (2017) compared three different PV panel cooling methods: water, nanofluid TiO_2 (0.04% by weight), and rectangular aluminum fins. With regard to a non-cooled system, this study reported efficiency increases of 5.37, 2.62, and 1.34% for the TiO_2, water, and fin cooling methods, respectively. In comparison with our study, this work showed that TiO_2 cooling is more effective than using water, while rectangular aluminum fins reported the lowest efficiency improvements.

In the study by Peng *et al*. (2017), which compared the ice cooling method applied behind the PV panel against a non-cooled system, the highest value reported by the latter was 4.98%. The cooled system, however, reached a 7.32% efficiency, which represents a 47% improvement.

Haidara *et al*. (2018) set out to determine the decrease in panel temperature and the changes in panel efficiency while using an evaporative cooling system that involved wetting a fabric and placing it on the back surface of the PV panel. They stated that a 14% increase in electricity generation efficiency was achieved, in response to a temperature drop of more than 20 °C in the PV panel.

In a study conducted by Gül and Akyüz (2019), the electrical and thermal performances of a hybrid PV-T system were evaluated. These authors observed a thermal efficiency range of 49.9-52.11% in measurements taken at fluid velocities of 0.015, 0.044, and 0.069 kg/s. A comparison of the module's electrical performance with and without cooling showed a 12.9% increase in the electrical output's maximum power point, and the electrical efficiency was calculated as 12%.

In the study carried out by Luboń *et al*. (2020), which aimed to reduce the module temperature by pouring cold tap water in the form of a water film on the surface of the PV module, a power increase of up to 20.2 W/m^2 was observed. These authors determined that a 20% increase in the total power generation can be obtained by continuously cooling the module's surface.

The effect of the water flow rate on PV panel performance was determined in the study by Govardhanan *et al*. (2020), used a uniform water flow on the PV panel surface for cooling. It was determined that cooling the PV module increases the output power by 15% when compared to the conventional approach. The cooling water flow rate that yielded the highest PV module power was 5.3 kg/min.

In a study that experimentally evaluated the effect of simultaneous dual-surface cooling on the output performance of a PV module, where both surfaces were cooled using a water flow on the panel surface and an absorbing cotton wick mesh on the rear surface, an improvement of approximately 30.3% in the output power of the panel was obtained. In addition, it was stated that the average efficiency of the non-cooled panel was 12.83%, against an average efficiency of 14.36% in the cooled system (Agyekum *et al*., 2021).

Zubeer and Ali (2022) set out to determine the changes in panel performance as a result of water cooling in concentrator photovoltaic (CPV) panels. They stated that the surface temperature decreased from 64.1 to 365 °C with this approach. This temperature drop increased the panel's electricity generation efficiency from 14.2 to 17%. In addition, in the water-cooled CPV system, the open circuit voltage and the short circuit current increased by 9 and 5.2%, respectively.

Shalaby *et al*. (2022) employed water passing through PVC pipes placed behind a PV panel for cooling. Their results showed an improvement of 14.1% in the power generation of the cooled PV panel. In addition, they stated that the electrical efficiency of the PV module with cooling reached 19.8% *vs*. the 17.4% value achieved without cooling.

The aforementioned studies show changes in panel surface temperature and electricity generation efficiency depending on the method applied. While the panel surface temperature decreases by over 20 °C with cooling applications, increments of up to 50% in power generation have been observed. According to the cooling method, increments in panel electricity generation efficiency ranging from 12 to 47% have been reported. In our study, it was determined that a maximum temperature difference of 13.7 °C is obtained by cooling the panel surface. Moreover, an efficiency increase of 52.6% was achieved in scenarios involving cooling. These data are similar to those obtained in previous studies.

When analyzing the surface temperature and efficiency graphs for the PV-T module in both scenarios, it was observed that decreases in surface temperature entail increased efficiency. By examining the measurements separately, we noticed a partial decrease in efficiency with increasing surface temperature. This parallel increase in efficiency and temperature could be a result of increased solar radiation. Above a certain temperature, the electricity production of PV modules becomes heat loss, reducing their efficiency. The main reason for this is that, as the temperature of the cell rises, the voltage produced by it decreases, resulting in thermal resistance.

Figure 12 provides a clearer graph of the relationship between the module's surface temperature and electricity generation efficiency throughout the analysis period.

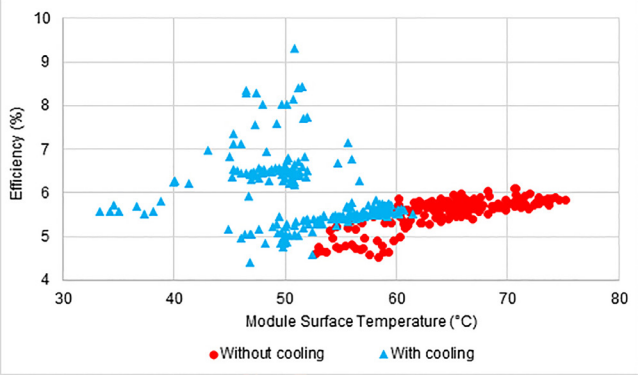


Figure 12. PV-T module with and without cooling, overall surface temperature-efficiency graph
Source: Authors

By evaluating all measurements together, a relationship between surface temperature and generation was observed (Figure 12). In all measurements taken without the cooling system, we noticed an efficiency mostly below 6%. Furthermore, the panel surface temperature did not fall below 50 °C. With cooling however, the surface temperature remained below 60 °C. For this scenario, the calculated efficiency values increased by up to 9%. It was determined that a panel surface temperature of 45-55 °C is necessary to achieve the highest PV module efficiency in summer, as losses were reported at higher temperatures.

The climate data (ambient temperature, humidity, and wind speed) on the days of the experiment were also recorded. The climate parameters for all days were approximately the same. Only on June 30th, 2019, was the air humidity very low in comparison. This day showed the highest values, especially regarding solar radiation (888-1 016 W/m^2).

Conclusions

This section summarizes the results obtained during our, which aimed to determine the effect of PV panel surface temperature on electricity production efficiency while using a water-cooled PV-T system.

- Solar radiation was observed to be in the range of 600-1 000 W/m^2 throughout the analyzed period.
- The highest power generation value was 77.3 W without cooling and 95.7 W with cooling.
- The module surface temperature was between 51.6 and 75.2 °C in the system without cooling and between 33.3 and 61.5 °C in the one with cooling.
- The module generation efficiency was 6.10% without cooling and 9.30% with cooling. This represents an increase of 52.62%.
- According to the scatterplot showing the relationship between module surface temperature and electricity generation efficiency, the most suitable module surface temperature is 45-55 °C, which yields a high efficiency in PV panels operating during the summer months in the Hatay province. This is also valid for regions with similar climate conditions.
- Furthermore, air humidity was shown to have a negative impact on direct solar radiation levels.

CRedit author statement

This article was elaborated within the framework of MSc studies and supported by Hatay Mustafa Kemal University, Coordinatorship of Scientific Research Projects (project no. 18.YL.053).

Conflicts of interest

The authors declare no conflict of interest.

Author contributions

All authors have participated in a) conception and design or data analysis and interpretation; (b) drafting the article or revising it critically for important intellectual content; and (c) approving of the final version of the manuscript.

References

Abu-Rahmeh, T. M. (2017). Efficiency of photovoltaic modules using different cooling methods: A comparative study. *Journal of Power and Energy Engineering, 5*, 32-45. https://doi.org/10.4236/jpee.2017.59003

Agyekum, E. B, PraveenKumar, S., Alwan, N. T., Velkin, V. I., and Shcheklein, S. E. (2021). Effect of dual surface cooling of solar photovoltaic panel on the efficiency of the module: experimental investigation. *Heliyon, 7*(9), e07920, https://doi.org/10.1016/j.heliyon.2021.e07920

Akbarzadeh, A., and Wadowski, T. (1996). Heat-pipe-based cooling systems for photovoltaic cells under concentrated solar radiation. *Applied Thermal Engineering, 16*(1), 81-87. https://doi.org/10.1016/1359-4311(95)00012-3

Andreev, V. M., Grilikhes, V. A., and Rumyantev V. D. (1997). *Photovoltaic conversion of concentrated sunlight*. Wiley.

Dubey, S., Sarvaiya, J. N., and Seshadri, B. (2013). Temperature dependent photovoltaic (PV) efficiency and its effect on PV production in the world: A review. *Energy Procedia, 33*, 311-321. http://doi.org/10.1016/j.egypro.2013.05.072

Govardhanan, M. S., Kumaraguruparan, G., Kameswari, M., Saravanan, R., Vivar, M., and Srithar, K. (2020). Photovoltaic module with uniform water flow on top surface. *International Journal of Photoenergy, 2020*, 9. https://doi.org/10.1155/2020/8473253

Gül, M., and Akyüz, E. (2019). Performance investigation of a hybrid photovoltaic-thermal (PV/T) system. *Journal of Balıkesir University Institute of Science and Technology, 21*(1), 444-458. http://doi.org/10.25092/baunfbed.548728

Haidara, Z. A., Orfib, J., and Kaneesamkandi, Z. (2018). Experimental investigation of evaporative cooling for enhancing photovoltaic panels efficiency. *Results in Physics, 11*, 690-697 https://doi.org/10.1016/j.rinp.2018.10.016

Luboń, W., Pełka, G., Janowski, M., Pająk, L., Stefaniuk, M., Kotyza, J., and Reczek, P. (2020). Assessing the impact of water cooling on PV modules efficiency. *Energies, 13*(10), 2414. https://doi.org/10.3390/en13102414

Meinshausen, M., Meinshausen, N., Hare, W., Raper, S. C. B., Frieler, K., Knutti, R., Frame, D. J., and Allen, M. R. (2009). Greenhouse-gas emission targets for limiting global warming to 2 °C. *Nature, 458*, 1158-1163. http://doi.org/10.1038/nature08017

Peng, Z., Herfatmanesh, M. R., and Liu, Y. (2017). Cooled solar PV panels for output energy efficiency optimisation. *Energy Conversion and Management, 150*, 949-955. https://doi.org/10.1016/j.enconman.2017.07.007

Radziemska, E. (2003). The effect of temperature on the power drop in crystalline silicon solar cells. *Renewable Energy, 28*, 1-12. https://doi.org/10.1016/S0960-1481(02)00015-0

Rawat, P., and Dhiran, T. S. (2017). Comparative analysis of solar photovoltaic thermal (PVT) water and solar photovoltaic thermal (PVT) air systems. *International Journal of Civil, Mechanical and Energy Science, 3*(1), 8-12. https://www.researchgate.net/publication/315657756_Comparative_Analysis_of_Solar_Photovoltaic_Thermal_PVT_Water_and_Solar_Photovoltaic_Thermal_PVT_Air_Systems

Shalaby, S. M., Elfakharany, M. K., Moharram, B. M., and Abosheiasha, H. F. (2022). Experimental study on the performance of PV with water cooling. *Energy Reports, 8*(Supp. 1), 957-961. https://doi.org/10.1016/j.egyr.2021.11.155

Shukla, A., Kanta, K., Sharmaa, S., and Biwoleb, P. H. (2017). Cooling methodologies of photovoltaic module for enhancing electrical efficiency: A review. *Solar Energy Materials & Solar Cells, 160*, 275-286. https://doi.org/10.1016/j.solmat.2016.10.047.

Solimpeks (n.d.) *Product Catalogue*. https://solimpeks.com.tr/wp-content/uploads/14-powerwolt-hibrit.pdf

Teo, H. G., Lee, P. S., and Hawlader, M. N. A. (2012). An active cooling system for photovoltaic modules. *Applied Energy, 90*, 309-315. https://doi.org/10.1016/j.apenergy.2011.01.017

Tiwari, A., Sodha, M. S., Chandra, A., and Joshi, J. C. (2006). Performance evaluation of photovoltaic thermal solar air collector for composite climate of India. *Solar Energy Materials & Solar Cells, 90*(2), 175-89. https://doi.org/10.1016/j.solmat.2005.03.002

Wang, G., Chao, Y., and Chen, Z. (2021). Promoting developments of hydrogen powered vehicle and solar PV hydrogen production in China: A study based on evolutionary game theory method. *Energy, 237*, 121649. https://doi.org/10.1016/j.energy.2021.121649

Wang, G., Zhang, Z., and Chen, Z. (2023). Design and performance evaluation of a novel CPV-T system using nano-fluid spectrum filter and with high solar concentrating uniformity. *Energy, 267*, 126616. https://doi.org/10.1016/j.energy.2023.126616

Zanlorenzi, G., Szejka, A. L., and Canciglieri, O. (2018). Hybrid photovoltaic module for efficiency improvement through an automatic water cooling system: A prototype case study. *Journal of Cleaner Production, 196*, 535-546. https://doi.org/10.1016/j.jclepro.2018.06.065

Zubeer, S. A., and Ali, O. M. (2022). Experimental and numerical study of low concentration and water-cooling effect on PV module performance. *Case Studies in Thermal Engineering, 34*, 102007. https://doi.org/10.1016/j.csite.2022.102007

Research Article/Mechanical, Mechatronics, and Materials Engineering

http://doi.org/10.15446/ing.investig.111284

Analysis, Modeling, and Simulation Solution of Induced-Draft Fan Rotor with Excessive Vibration: A Case Study

Solución de análisis, modelado y simulación de rotor de ventilador de tiro inducido con vibración excesiva: un caso de estudio

Erick Alejandro González-Barbosa[1], José Juan Vázquez-Martínez[2], Fernando Jurado-Pérez[3], Héctor Castro-Mosqueda[4], Francisco Javier Rodríguez-Ornelas[5], and José-Joel González-Barbosa*[6]

ABSTRACT

In the modern industry, computer modeling and simulation tools have become fundamental to estimating the behavior of rotodynamic systems. These computational tools allow analyzing possible modifications as well as alternative solutions to changes in design, with the aim of improving performance. Nowadays, rotodynamic systems, present in various industrial applications, require greater efficiency and reliability. Although there are deep learning methodologies for monitoring and diagnosing failures which improve these standards, the main challenge is the lack of databases for learning, a problem that can be addressed through experimental monitoring and computer analysis. This work analyzes the vibrations of two induced-draft fans with excess vibration in a thermoelectric plant in Mexico. A vibration analysis was carried out through the instrumentation and monitoring of accelerometers located at crucial points in the fans. The results of this experimental analysis were validated by computer simulation based on FEM. The results show that the operating speed of the induced-draft fans is very close to their natural frequency, causing considerable stress and potential failures due to excessive vibration. Finally, this work presents a practical solution to modify the natural frequency of induced-draft fans, so that they can function correctly at the required operating speed, thus mitigating excessive vibration issues.

Keywords: Monitoring and data analysis, fault diagnosis, computer simulation, mitigate mechanical vibrations.

RESUMEN

En la industria moderna, las herramientas de modelado y simulación computacional se han vuelto fundamentales para estimar el comportamiento de los sistemas rotodinámicos. Estas herramientas computacionales permiten analizar posibles modificaciones y soluciones alternativas a cambios en el diseño, con el objetivo de mejorar el rendimiento. Hoy en día, los sistemas rotodinámicos, presentes en diversas aplicaciones industriales, requieren mayor eficiencia y fiabilidad. Aunque existen metodologías de aprendizaje profundo para el monitoreo y diagnóstico de fallas que mejoran estos estándares, el principal desafío es la falta de bases de datos para el aprendizaje. Este problema puede ser abordado a través del monitoreo experimental y el análisis computacional. Este trabajo analiza las vibraciones de dos ventiladores de tiro inducido con exceso de vibración en una planta termoeléctrica en México. Se realizó un análisis de vibración a través de la instrumentación y el monitoreo de acelerómetros ubicados en puntos cruciales de los ventiladores. Los resultados de este análisis experimental fueron validados por simulación computacional basada en el método de elementos finitos. Los resultados muestran que la velocidad de operación de los ventiladores de tiro inducido está muy cerca de su frecuencia natural, causando un estrés considerable y posibles fallas debido a la vibración excesiva. Finalmente, este trabajo presenta una solución práctica para modificar la frecuencia natural de los ventiladores de tiro inducido, de modo que puedan funcionar correctamente a la velocidad de operación requerida, mitigando así los problemas de vibración excesiva.

Palabras clave: Monitoreo y análisis de datos, diagnóstico de fallas, simulación por computadora, mitigar vibraciones mecánicas.

Received: September 25th 2023
Accepted: February 27th 2024

Introduction

Vibration issues in rotating machinery depend on multiple factors, including misalignment, critical speeds or resonances, system deterioration due to continuous use, bearings defects, and imbalance. It is well known in the field of vibrations that critical velocities or resonances are physical phenomena that arise when the natural frequency of the system matches the operating frequency. These occurrences can be mitigated by altering the system design and adjusting its natural or operating frequency (Xiangyang *et al.*, 2023). Simultaneously, imbalances can manifest as either mechanical or electrical in nature. Mechanical imbalance arises when the principal axis of inertia does not align with the geometric axis of the system. According to (Blanco-Ortega *et al.*, 2010), several active and passive methods or devices have been developed trough vibration analysis to mitigate this mechanical phenomenon. Electrical imbalance is generated by voltage variations or harmonic distortion in the voltage, which, in the typical case of induction motors, affects their dynamic behavior and vibrations (Donolo *et al.*, 2016; Ren *et al.*, 2023). For steam turbines, the vibration in the rotor is divided into two categories (Kaneko *et al.*, 2022): forced vibration and self-excited vibration. Forced vibration is caused by an external

[1]Tecnológico Nacional de México/ITS de Irapuato, México
[2] Instituto Politécnico Nacional, Queréretaro, México. *E-mail: jgonzalezba@ipn.mx

force and is divided as follows: a) by rotor imbalance, b) by mechanical imbalance in gears and connections or couplings, c) by electric excitation from the motor or generator, d) by fluid excitation, and e) by uncommon factors. Self-excited vibration is mainly caused by oil whip and steam whirl.

Long-term vibration defects in the machinery significantly impact both the equipment's service life and the regular, stable operation of the unit (Li *et al.*, 2020). According to studies on different types of turbomachines, the leading causes of vibrations are cavitation and resonance. In addition, excess speed and long periods of operation contribute to this effect (Doshi *et al.*, 2021). Although vibrations are unwanted in systems, they also serve as support to identify defective designs, poor installation, and wear or deterioration. Over time, improper installation, manufacturing defects, waste accumulation, and the gradual degradation of machines can lead to rotor eccentricity. This eccentricity adversely impacts their performance, causing high noise and vibration levels, elevated losses, and the risk of overheating (Chapagain and Silwal, 2023; Yu-Ling *et al.*, 2023).

Induced-draft fans are used to evacuate air from a space or to create a negative air pressure in a system. They are one of the crucial elements of a thermal power plant's auxiliary equipment. Vibration studies on industrial fans focus on the main static and dynamic parts, such as the rotor and bearings (Zenglin and Gordon, 2003; Shabaneh and Zu, 2003; Trebuna *et al.*, 2014; Jagtap *et al.*, 2020), the blades (Pingchao *et al.*, 2018; Wang *et al.*, 2020), the casing and structure (Niko *et al.*, 2011), and the ducts. At times, vibration issues arise from the intricate interaction between various components of the fan-duct-foundation system.

Rotating machinery can be analyzed using physical models (Shabaneh and Zu, 2003; Zenglin and Gordon, 2003), experimental tests through instrumentation (Jerzy *et al.*, 2014), simulation via mechanical/acoustic and computational methods (Manish *et al.*, 2015; Kalmár-Nagy *et al.*, 2015), or a combination of approaches (Pingchao *et al.*, 2018; Novotný *et al.*, 2019; Xie *et al.*, 2023). In the case of rotating machinery (*e.g.*, fans), vibration measurement has become a fundamental tool for monitoring operating conditions (Guo *et al.*, 2021; Dhamande *et al.*, 2023). In some studies, comprehensive analyses are conducted to identify fan failure. The methods employed encompass vibration analysis, noise measurement, wear debris analysis, and ultrasonic monitoring (Jagtap *et al.*, 2020).

In recent years, various modern techniques have emerged from analyzing, determining, and predicting failures in rotating machinery based on a) artificial intelligence (Liu *et al.*, 2018) through data acquisition, feature extraction, and fault recognition; b) neural networks (Benrahmoune *et al.*, 2018), by analyzing and comparing monitoring data; and c) machine learning models (Benchekroun *et al.*, 2023) to accurately predict the vibration problem in fans. Considering the above, and bolstered by preventive maintenance to reduce downtime and enhance efficiency, operators can proactively identify potential issues before they reach critical levels. (Di *et al.*, 2022) propose a novel method for detecting anomalies in rotating machinery, leveraging vibration vectors inspired by the Polar plot. These vectors contain amplitude and phase values from various characteristic frequencies. By converting the original Polar vectors to Cartesian plots using fast Fourier transform-based order analysis (FFT-OA), calculations are simplified and visualization is improved. Wei *et al.* (2022) introduce an innovative condition monitoring approach for induced-draft systems. They combine a genetic algorithm with a long-short-term memory network (GA-LSTM) to establish dynamic and static thresholds for anomaly detection. Applying this method to a coal-fired power plant enables early fault diagnosis and detection. Experimental results demonstrate the effectiveness of this approach, successfully detecting minor anomalies in advance and contributing to improved system reliability and proactive maintenance.

Physical models allow estimating a system's real behavior to better perform and interpret computational and simulation analyses (Xie *et al.*, 2023; Novotný *et al.*, 2019). The study by Zenglin and Gordon (2003) is based on the Jeffcott rotor model with external damping for vertical and horizontal directions. The Lund Stability method is applied for the horizontal direction, whereas an analytical and numerical analysis is conducted for the vertical direction to study the threshold speeds. The authors also consider improving the stability characteristics of the system by adjusting external damping values. Shabaneh and Zu (2003) perform a dynamic analysis of a single-rotor shaft system with nonlinear elastic bearings at the ends mounted on viscoelastic suspension. The Timoshenko shaft model represents the flexibility of the shaft, and the viscoelastic supports are modeled using the Kelvin-Voigt model. The authors also find that the primary resonance peak is modified for higher frequencies based on increased nonlinear elastic bearing characteristics.

Bearings play a critical role in rotating machines. The inevitability of degradation and bearing failures arises from prolonged and continuous operation, encompassing factors such as poor maintenance practices, inadequate care, and dynamic stress induced by the dynamic loads of rotating parts. Recent research endeavors employ innovative analysis techniques for the vibration analysis of bearings, recognizing their inherent complexity (Mohamad *et al.*, 2023; Lu *et al.*, 2023). Furthermore, probabilistic techniques such as the remaining useful life (RUL) as well as those based on deep learning have become very important. To predict the failure of bearings through vibration analysis or estimate their remaining service life, modern techniques are complemented with the Vold-Kalman parametric filter method (Cui *et al.*, 2019). This method, based on the state-space framework, has recently found applications in the navigation and control of vehicles and spacecraft. In the context of techniques rooted in deep learning, such as the innovative PIResNet (Ni *et al.*, 2023), their necessity arises from the inherent variations in speed and load experienced by bearings during the operation of machinery. PIResNet stands out by offering a consistent physical solution even in the presence of imperfect data. Given the critical significance of bearings, their behavior has undergone evaluation in both previous unsuccessful actions and our ongoing research project. In both instances, they have been ruled out as the causes of excessive vibrations.

The main contribution of this work lies in the comprehensive study and identification of the cause of excessive vibration in two induced-draft fans of a thermoelectric plant in Mexico.

In addition, through computer simulation, a solution is proposed. Previous attempts to resolve the vibration issue had been unsuccessful, prompting the need for a more rigorous and systematic approach. The hypothesis of this work assumes that a practical and theoretical analysis through simulation will shed light on the leading causes of excessive vibration, thus allowing to propose solutions. This hypothesis employed a methodology involving the instrumentation of rotor bearings with accelerometers, fan modeling, simulation, and data analysis.

The research objectives of this work are as follows: (i) to study and identify the cause of excessive vibration in an induced-draft fan at a Mexican thermoelectric power plant; (ii) to develop a methodology that includes instrumentation, fan modeling, simulation, and data analysis to evaluate vibrations and determine the natural frequencies and vibration modes of the fan's operation; (iii) to assess the hydraulic bearings' stiffness and damping constants for a more accurate computer simulation; (iv) to emphasize the importance of vibration measurement and monitoring for the proper operation and maintenance of fans; (v) to provide valuable insights and recommendations for mitigating vibration issues and improving the performance and reliability of induced-draft fans in thermoelectric power plants.

This document is organized into sections focusing on analyzing, modeling, and simulating an induced-draft fan model with excessive vibration. It starts with an *Introduction* that provides an overview of vibration issues in rotating machines and a case study conducted in a thermoelectric plant in Mexico, in addition to stressing the need for analyzing fan vibration issues. Figure 1 depicts the proposed approach while considering previous actions and presents an overview of the methodology. The *Problem definition* section provides information on the system's design, vibration limits, and the need for mitigating excessive vibrations. It also describes previous actions to address the vibration issues. The *Metodology* section outlines the analysis, covering both experimental instrumentation and computer simulation. It explains the approach to vibration analysis, which involved the placement of accelerometers in the fan bearings to validate the natural frequencies. The document further details the process of modeling and simulating the fan using the finite element method (FEM) in the ANSYS software. The *Results* section presents the data obtained from static and dynamic tests, encompassing the modeling and meshing of individual fan components as well as the complete simulation. Subsequently, the *Discussion* and *Conclusion* sections draw upon relevant studies, offering valuable insights and recommendations to enhance the performance and reliability of a particular induced-draft fan in a thermoelectric power plant.

Problem definition

Unpredictable rotor vibrations have been observed in two fans at a thermoelectric plant in Mexico, with each occurrence displaying distinctive characteristics. In order to mitigate random vibrations, it red was necessary to balance the shaft-compressor system of the rotor. This process has been frequently performed over the past decade whenever the system experiences vibration. However, this issue has resulted in significant economic losses due to the inadequate operation of fan motors or insufficient maintenance times to achieve balance.

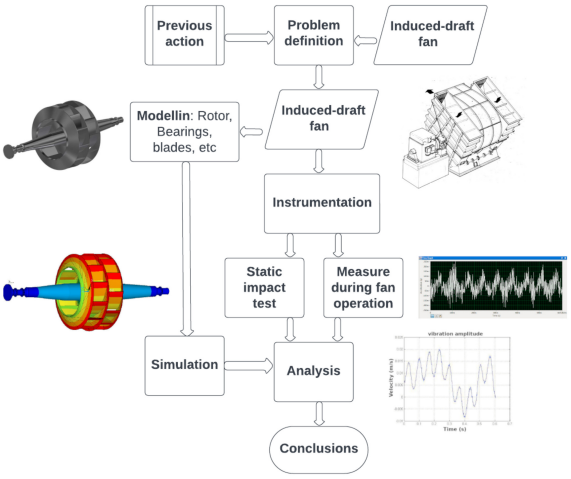


Figure 1. Proposed approach
Source: Authors

Table 1. Fan information

Description	Quantity
Volumetric flow	2000 $\frac{m^3}{hr}$
Static pressure	0.602 m water column
Temperature	417.5 K
Type	Double suction blade or vane control, constant speed

Source: Authors

The system comprises two 2500 kW motors, each integrated into a rotor, and each duct is supplied by one of these motors (Figure 2). Table 1 presents information on the system. Airflow is supplied to the ducts of these motors. The air inflow is in the vertical direction from the outside in, and the air outflow is in the horizontal direction towards the entrance of the ducts (Figures 3a and 3b).

The mechanical vibration issues of the system have persisted since it was put into operation two decades ago, with the bearings reaching unbalanced speeds between 10 and

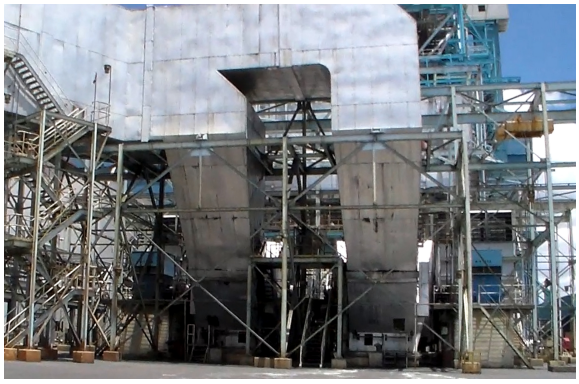
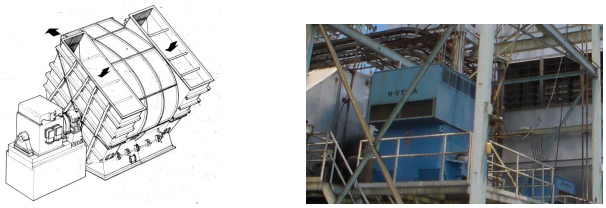


Figure 2. Lateral section of the ducts and actuation engines
Source: Authors



(a) Manufacturer design **(b)** Induced-draft motor

Figure 3. Fan motor
Source: Authors

Table 2. Typical band boundaries regarding the effective vibration velocity of machines in class III according to STN ISO 10816-1 (Trebuna *et al.*, 2014)

Band	v_{eff} ($\frac{mm}{s}$)	Characteristics
A	0.28-1.80	Typical vibration of new machines
B	2.28-4.50	Vibration of machines assigned for non-stop long-term operation
C	7.10-11.2	Vibration of machines that are unsuitable for long-term non-stop operation
D	18.0-45.0	No vibration allowed in the machines

Source: Authors

$11mm/s$, exceeding the limit established by the regulations (Table 2). While induced-draft fan 1A initially experienced minimal issues, it has required frequent balancing since 2002. Achieving motor stability as it passes through the system's natural frequencies is crucial, and the balancing process varies for each of the two 2500 kW motors. Sometimes, up to 20 motor starts are necessary for correct balancing without complications.

Previous actions

Earlier, the internal staff of the thermoelectric, as well as private companies, used to analyze vibration issues in the system. In this regard, previous actions included the following:

- In 1996, a new rotor and shaft design was implemented for induced draft-fan 1A, significantly reducing the need for more continuous balancing. The revised design specifically involved the switch to a more rigid steel material.
- Ash residues, with quantities of approximately 0.300 kg, were detected on the fan blades. In response, a maintenance and cleaning program with increased frequency was proposed. While this initiative has contributed to a partial mitigation of excessive vibration issues, it remains a necessary measure.
- In some instances, repairing the edges of the fan blades has been necessary.
- To achieve system stability, the structure of the ducts was modified, and gates were added to the fan suction. These actions were also unsuccessful.
- The bearings were also tested to determine whether they were the cause of excessive vibration, and the foundation of the fans was modified. Both modifications were unsuccessful.

In light of the above, the following hypotheses were proposed:

1. The operating frequency of the engine is very close to one of the system's natural frequencies (simulation, hammer test and instrumentation sections will be used to test these hypotheses).
2. Improper balancing, whereby an additional mass is added to a subsequent imbalance, may be the cause of excessive vibration.
3. The accumulation of external elements (*i.e.*, ash, additives, moisture) or blade wear causes the rotor to be unbalanced.

In light of the above, a methodology was structured, with the aim of examining the system's excessive vibration issues. This methodology comprises two main parts: experimental instrumentation and computer simulation.

Methodology

Our approach began with the system's instrumentation for static and dynamic testing. We modeled and meshed the fan parts. In this model, master nodes connect the meshed parts. A vibration analysis of the complete system was carried out through a computer simulation of the fan's operation.

Instrumentation of the induced-draft fan

Vibration curves and natural frequencies were obtained by conducting dynamic and static tests with accelerometers placed on two fans. One fan was operational for the dynamic case, while the other was turned off for the static case.

Dynamic testing

Through the instrumentation process, the operation of the induced draft-fan was evaluated in order to estimate the system's natural frequencies. To properly take measurements of these machines' bearings, the first step involved adjusting the position of the accelerometers. One accelerometer was used to measure vibrations in the vertical direction of the bearing, and the other one was employed to measure vibrations in the longitudinal or horizontal direction of the bearing (Figure 4).

Static impact test

An impact test was performed on induced-draft fan 1B, taking advantage of the fact that the unit was out of service due to major maintenance (*i.e.*, joint expansion replacement). The impact test is a vibration measurement using accelerometers, wherein the system is excited using a high-intensity mechanical impulse, forcing it to vibrate in its natural modes. If the system is operating, the damping factor of the system can also be obtained. Nevertheless, the only information that could be obtained from this test were the natural frequencies and the first vibration modes.

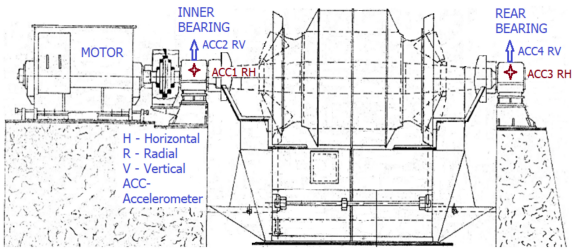
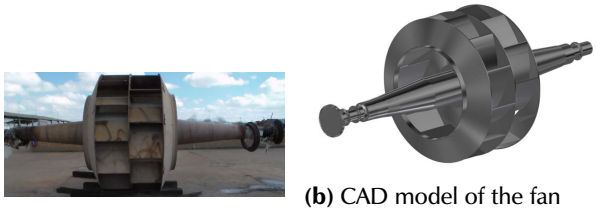


Figure 4. Placement of the accelerometers and directions of response measurements in the induced-draft fan
Source: Authors



(a) Real model of the fan rotor rotor

(b) CAD model of the fan

Figure 5. Rotor and CAD model
Source: Authors

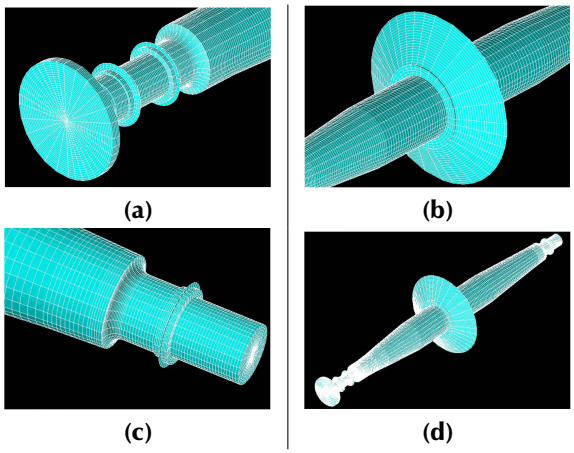


Figure 6. Sectional meshing: a) connection to the motor of the rotor shaft component, b) middle section of the rotor shaft component, c) final section of the rotor shaft component, d) complete rotor shaft mesh
Source: Authors

Modeling the induced-draft fans

The goal of this vibration analysis was to provide a solution to the vibration issues, which cause mechanical damage and result in downtimes, leading to economic losses for the company. The analytical solution aimed to extend the service life of the fans and ensure optimal performance without frequent maintenance.

This work focuses on three main aspects:

1. the material's mechanical properties, using specialized techniques to understand its behavior under mechanical stress;
2. the root causes of vibration failure in the system;
3. proposing a series of solutions to mitigate vibration issues in the system.

Rotor

In rotating machinery, vibration can stem from various sources, with imbalance being a major cause. While rotors ideally consist of homogeneous material with a uniform mass distribution along their axis and perfect symmetry, the presence of small randomly distributed mass concentrations unfortunately leads to imbalance. This imbalance can be caused by factors such as corrosion and dirt. Balancing a rotor involves ensuring that its principal axis of inertia aligns with the center of gravity, coinciding with the axis of rotation, nearly eliminating vibration.

Rotating machinery may experience instability due to various factors. Factors that can be attributed to incorrect rotor assemblies include inadequate shaft alignment and incorrect bearing placement. In many cases, rotors operate above their natural frequency, causing resonance to occur briefly when the operating speed (or impeller speed) passes through this frequency. These occurrences can cause damage to machinery components, such as seals and bearings, leading to instability. In addition, high-speed operation and friction in the bearings can lead to wear and temperature increases, modifying the dynamic characteristics of lubricants and resulting in poor lubrication and further contributing to instability in the bearings.

There are various models available for dynamic rotor analysis, ranging from very simple to extremely complex. The complexity of these models is based on the number of degrees of freedom (Zenglin and Gordon, 2003; Shabaneh and Zu, 2003; Yang *et al.*, 2024). While some models may be unrealistic and inadequate, others may need to be more complex to be practically useful, making their mathematical representation and solution nearly impossible. Figure 5a shows the fan rotor. The model was compared against design models based on manufacturer and field measurements to ensure accuracy. The 3D CAD model of the fan rotor is shown in Figure 5b.

After creating the geometric model, the rotor was meshed to generate a sectional model. This involved selecting certain parts or components of the rotor and assembling them into a single model, as illustrated in Figure 6.

The mesh corresponding to the complete rotor shaft shell has a complex geometry due to the shape of the blades, as shown in Figure 7, which depicts the appropriate assembly of elements.

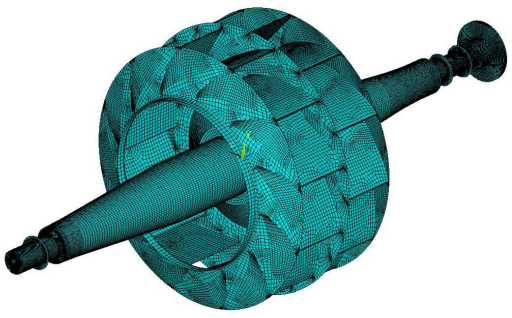


Figure 7. Total mesh of the rotor shaft
Source: Authors

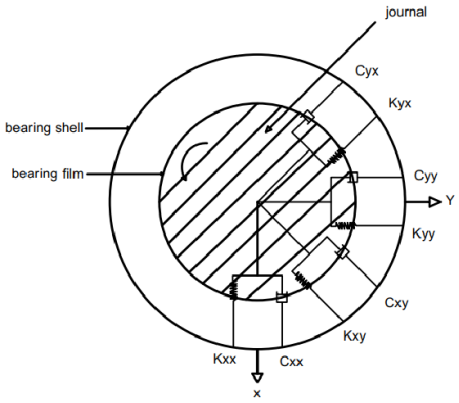


Figure 8. Schematic representation of a hydraulic bearing
Source: Authors

Table 3. Stiffness and damping coefficients for the simulated hydraulic bearing

Constant	Value
K_{xx}	$190.673496665 \times 10^6 [N/m]$
K_{yy}	$52.599585288 \times 10^6 [N/m]$
C_{xx}	$4.389542179 \times 10^6 [(N-s)/m]$
C_{yy}	$0.6270774542 \times 10^6 [(N-s)/m]$
K_{xy}	$203.823392957 \times 10^6 [N/m]$
K_{yx}	$85.474326105 \times 10^6 [N/m]$
C_{xy}	$0.2429925136 \times 10^6 [(N-s)/m]$
C_{yx}	$0.219477109 \times 10^6 [(N-s)/m]$

Bearings

The bearings used in these fans are hydrodynamic and can be represented by four stiffness and four damping coefficients. These coefficients are usually modeled as four spring-damper systems placed at a separation of 45^o (Jorgen *et al.*, 1965; Čorović and Miljavec, 2020; Xie *et al.*, 2023), as shown in Figure 8.

Based on Sommerfeld numbers, and assuming that the system is in resonance when the natural and the operating frequencies through the angular velocity are similar, the eight stiffness and damping constants to be introduced into the software for simulation are defined in Table 3. The following is also considered:

- Each bearing carries 9394.849 kg due to the weight of the rotor.
- The radial clearance of the bearing is c = 0.0014 m.
- The critical angular velocity of the system is also considered.

Based on a Sommerfeld number of 0.9, and assuming that the first resonance is the most pronounced, the critical speed is 0.9 (890 rpm) = 801 rpm (83.88 rad/s). The value of 890 rpm is the system's operating speed and corresponds to a frequency of 14.8333 Hz. The value of 0,9 was chosen instead of 1 to prevent the system's natural and operating frequencies from being equal, thus only approaching resonance.

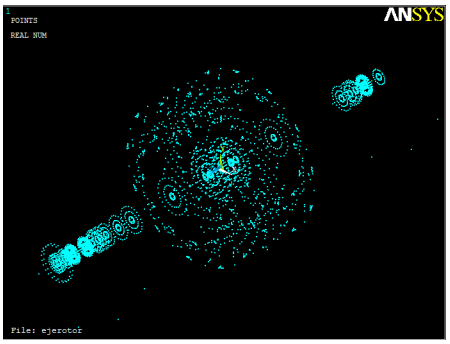


Figure 9. Key points to be connected with the master node
Source: Authors

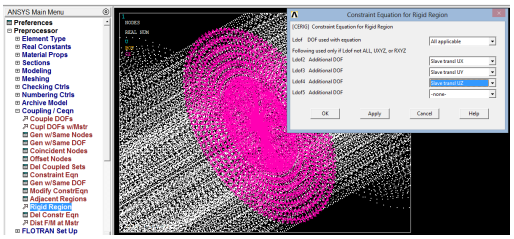


Figure 10. Association of slave nodes with the master node
Source: Authors

Preparing the performance simulation

The simulation was carried out in the ANSYS software. The analysis was performed based on several factors, *i.e.*, for two materials, considering increases in diameter and the accuracy of bearings. The materials defined for the rotor shaft were 1026 steel (E=205GPa) and 4340 steel (E=215GPa), both with a Young modulus of 0.29. The MASS21 element assigned mass values to key points and consequently provided the values of the loads applied to each bearing. These key points were used as master nodes. For the real constant of the MASS21 element, the approximate value of the bearing weight was introduced.

To simulate the spring-damper pairs, the COMBIN14 element must be assigned to assume the values in Table 3 for the eight different values of the stiffness and damping constants.

Key points were created in the part of the fan where the bearings are mounted. These key points were then meshed as a mass element and used as a master node with which all slave nodes were associated. Attributes were assigned to the previously created key points, the element type (MASS21), and its real constant. This yielded a mesh, which implies a mass element. The key points were created to be connected with lines to the mass element (master node) that simulates the springs and dampers through the COMBIN14 element. This process can be observed in Figure 9. Here, the key points are outside the rotor.

Afterwards, the master node was associated with the slave nodes. The master node was the MASS21 element, and the slave nodes were all those around it, as shown in Figure 10.

As depicted in Figure 11, lines were created to represent the spring-damper system. Each of the lines starts from the master node (MASS21 element) and goes towards the different key points that were created earlier.

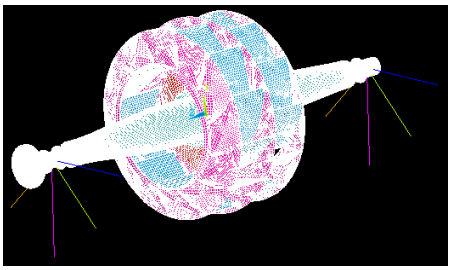


Figure 11. Lines to simulate each spring-damper assembly
Source: Authors

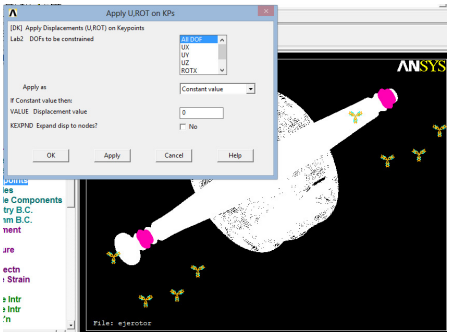


Figure 12. Restriction of displacements in the spring-damper system
Source: Authors

To assign attributes to the previously created lines created, these steps were followed:

1. The element type (COMBIN14) was given.
2. The corresponding real constant for each line was assigned.
3. Meshing was performed on the lines.

These steps aimed to simulate the operation of the spring-damper system. To avoid unwanted displacement, the following actions were performed (Figure 12):

1. The key points connected to the elements of COMBIN14 were restricted in all directions.
2. The master node was restricted in the direction of z.

The finished model can be observed in Figure 13, which is sufficient for the modal analysis. This analysis is very important in the design and development of a fan to prevent premature failure (Manish *et al.*, 2015; Čorović and Miljavec, 2020), as it helps to obtain the vibration modes of the system.

Results

The results presented herein start with the data obtained from static and dynamic experimental tests using instrumentation. Following this, the modeling and meshing of the fan parts are detailed, leading to the ultimate stage of computer simulation.

Dynamic test

Measurements were obtained for the induced-draft fan 1A. It was possible to measure the vibration in both bearings for

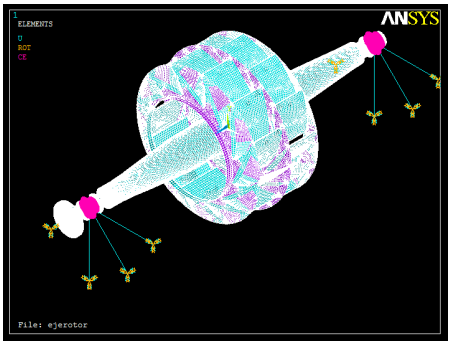


Figure 13. Completed model prior to modal analysis
Source: Authors

the connection closest to the motor (inner bearing) and the opposite or furthest from the motor (rear bearing) (Figure 4). Figures 14 and 15 show the vibration corresponding to the horizontal and vertical directions. Despite the presence of some noise, both figures exhibit a harmonic behavior. This consistent behavior allowed eliminating any possibility of damage to the bearings.

Figures 16 and 17 show the integral of the acceleration curves, which clearly shows the harmonic behavior of the cycles, eliminating the noise as well.

Figures 18 and 19 show the Fourier spectrum and the integration of the spectrum of the rotating behavior of the bearing in the horizontal and vertical directions. The maximum peak occurs at approximately 15 Hz for both cases.

The vibration values in the horizontal direction exceed those obtained in the vertical direction. This discrepancy arises from the lower rigidity in the horizontal direction, in contrast to the vertical direction, due to the presence of vertical supports. This behavior has been experimentally confirmed.

Static impact test

Figure 20 shows the results obtained from the impact test conducted on the out-of-operation fan 1B. The top part shows the vibration amplitude, and the bottom part shows the Fourier spectrum indicating the natural frequencies. As seen in this test, values of 15 and 20 Hz were obtained for this particular configuration. It is important to emphasize that the unit was out of service: the bearings were not operating, and there was no mechanical connection between the electric motor and the fan rotor. The results show that the natural frequencies of the rotor are very close to its operating frequency, as already validated with the accelerometers mounted during the operation of the other fan.

Simulation without bearings

Through simulation, it is possible to determine the different vibration modes of a fan. These modes are fundamental to determining primary conditions such as angular velocity or system design, which are essential for a correct operation. The simulation of the system is based on the real operating conditions of the fan, meaning that the model is close to reality. Figure 21 shows the frequencies of the first vibration mode of the rotor for two different materials.

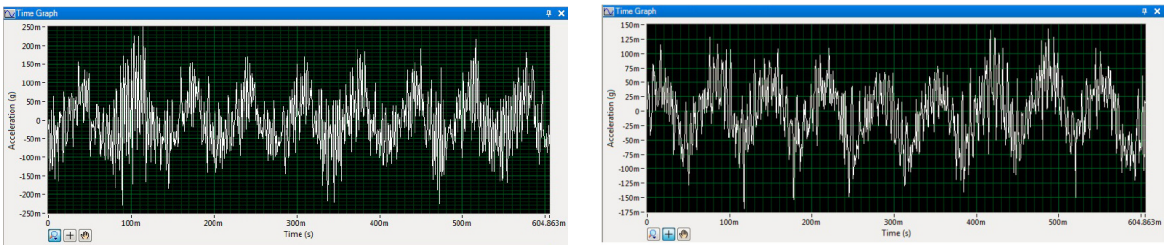


Figure 14. Vibration obtained in the horizontal direction of the inner bearing, left; and rear bearing, right. Induced-draft fan 1A.
Source: Authors

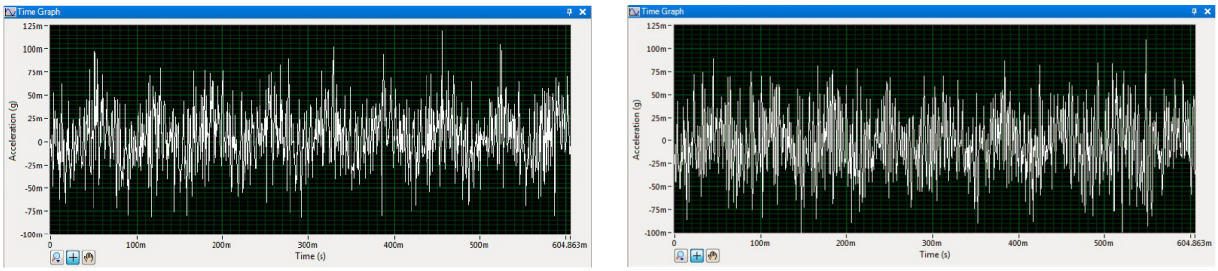


Figure 15. Vibration obtained in the vertical direction of the inner bearing, left; and rear bearing, right. Induced-draft fan 1A.
Source: Authors

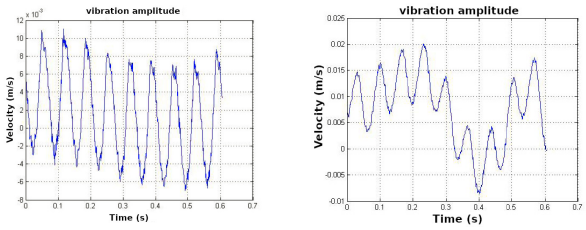


Figure 16. Integral of the vibration obtained in the horizontal direction of the inner bearing, left; and rear bearing, right. Induced-draft fan 1A.
Source: Authors

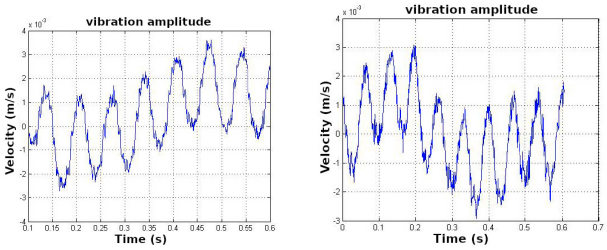


Figure 17. Integral of the vibration obtained in the vertical direction of the inner bearing, left; and rear bearing, right. Induced-draft fan 1A.
Source: Authors

According to the results shown in Figure 21, the system's natural frequency is sensitive to the material. Another factor that we consider fundamental is the rotor diameter. In this vein, to increase the cross-section, analyses were conducted in order to evaluate the impact on the natural frequency. The results are shown in Table 4.

According to the results regarding the change in the radius of the inner shaft, by increasing the stiffness of the shaft, it is possible to increase the natural frequencies of the system's vibration modes. It should be noted that the longitudinal dimension cannot be modified. It was also necessary to analyze the system's vibration modes. Figures 21 and 22 show the first to the fourth mode. These modes were included to rule out their similarity to the functional frequency of the system.

Table 4. Natural frequency of the undamped system according to the radial increment of the shaft and the two materials used

Radial increment [m]	Material	Natural frequency [Hz]
0.01	1026 steel	19,302
0.02		20.009
0.03		20.834
0.04		21.343
0.05		25.254
0.01	4340 steel	19.671
0.02		20.392
0.03		21.111
0.04		21.923
0.05		25.945

Source: Authors

Table 5. Results of the two main simulations

First mode of vibration without bearings	First mode of vibration with bearings
18.559 Hz	15.3622 Hz

Simulation with bearings

Figure 23 shows the vibration mode obtained with the original characteristics of the system, including the bearings. It can be observed that the natural frequency (15.3622 Hz) of the system is very close to its excitation or operating frequency (890 rpm, 14.8333 Hz), which likely indicates the reason for excessive vibration in the system.

Table 5 compares the results obtained from the two initial simulations of the system in its first mode of vibration, represented in Figures 21 and 23.

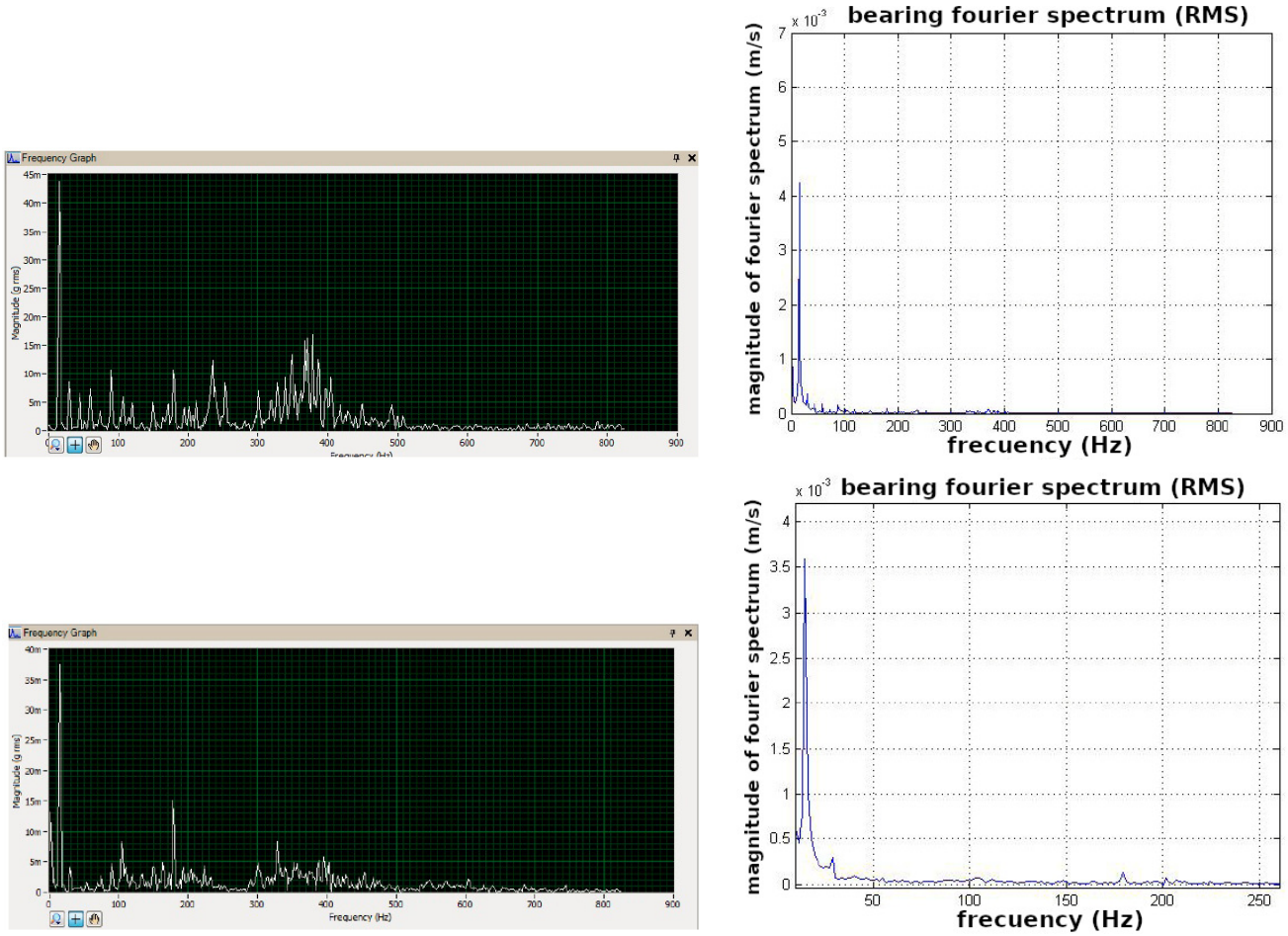


Figure 18. Fourier spectrum and integral of the Fourier effect in the horizontal direction of the inner bearing, top; rear bearing, below. Induced-draft fan 1A.
Source: Authors

Discussion

The first natural frequency is 15.3622 Hz. This frequency was obtained from experimental tests, and it was validated using computer simulation (Table 5). The operating frequency (14.833 Hz), obtained based on the impeller speed (890 rpm), is close to the system's natural frequency. Excessive vibration occurs due to the convergence of these values toward mechanical resonance.

Modifying the rotation speed of the fan rotor in order to move it away from its natural frequency values, with the purpose of preventing resonance, was not possible due to two main factors: a) the airflow and the efficiency of the system are altered and no longer meet the functionality requirements; and b), according to Qingjie *et al.* (2020), it is not possible to eliminate the torsional vibration generated in this type of systems when adding a variable frequency drive. Inertial factors generate this torsional vibration due to speed changes in induced draft fans, and the torsional vibration is in turn caused by frequency conversion driving technology due to dynamic electromechanical coupling. Several alternatives to eliminate the torsional vibration of the shaft were considered in (Qingjie *et al.*, 2020), finding that it is only possible to displace its resonance region and reduce its amplitude but not eliminate it.

M and K (2023) propose a MATLAB-based methodology to prevent machine downtime. This methodology consists of continuously monitoring the rotor-bearing system with regard to vibration responses obtained through data collection and the fast Fourier transform (FFT) analyzer. This work, however, only proposes the detection of the fault and makes some suggestions. In our work, conducting vibration monitoring in the rotor-bearing system was imperative to identifying the primary vibration modes and establishing their similarity to the operating frequency. This understanding was crucial to define an effective solution. To facilitate our vibration analysis, we utilized the MATLAB software and integrated the Fourier spectrum of the bearing's rotating behavior in both the horizontal and vertical directions.

Numerous research projects such as that of Manish *et al.* (2015) have employed the FEM for the modal analysis of rotating industrial equipment. Some works, beyond relying solely on computational approaches, also incorporate physical models to validate their findings, as demonstrated by Čorović and Miljavec (2020) and Noureddine and Noureddine (2022). In Manish *et al.* (2015), a comprehensive analysis of the vibration modes of a centrifugal fan was undertaken, focusing on the acquisition of the first ten vibration modes. This extensive examination aimed to validate that these modes did not closely align with the fan's operating speed or frequency, thus avoiding vibration issues. The authors of this study emphasized the crucial nature of such analyses for rotating machinery,

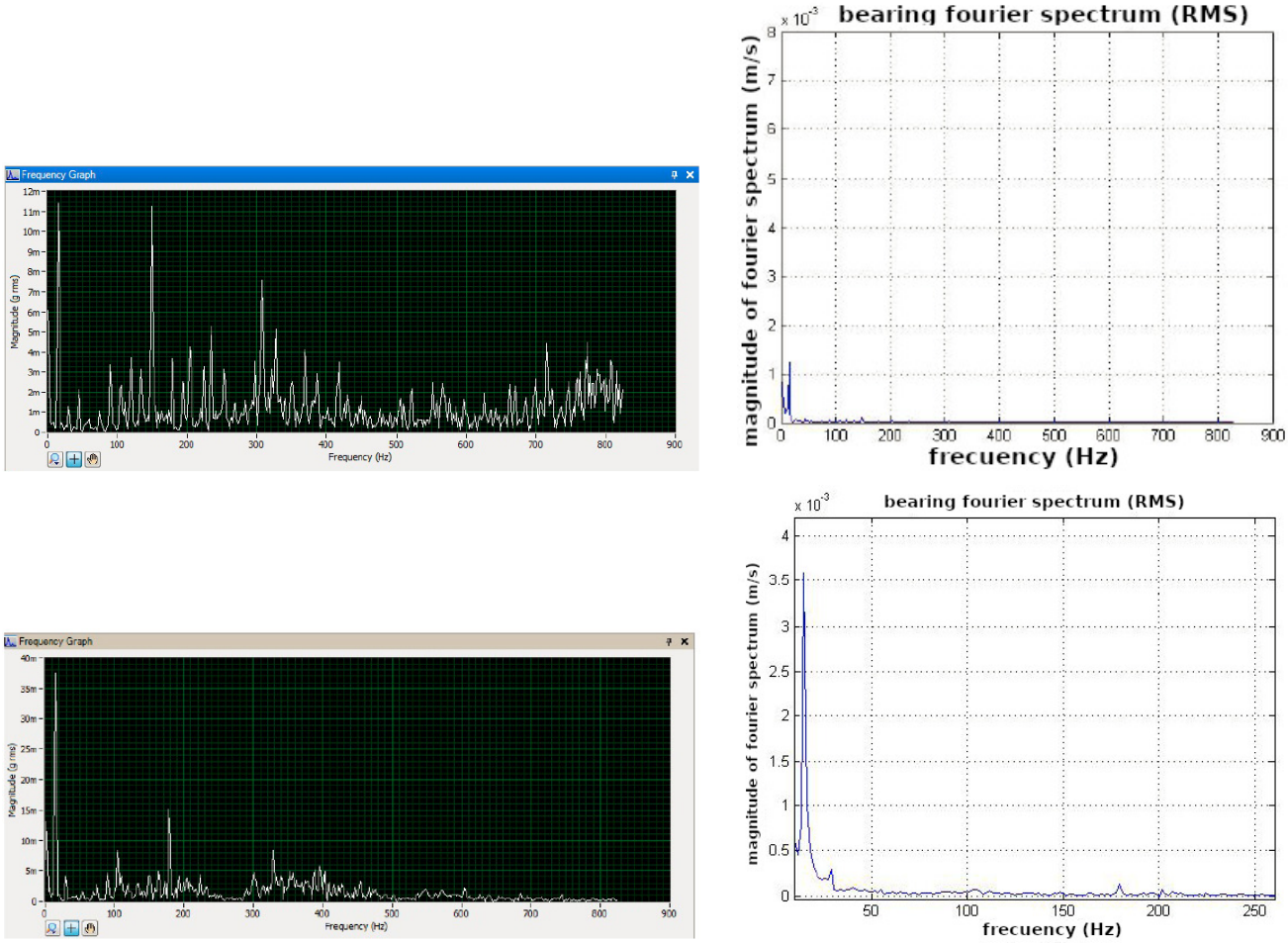


Figure 19. Fourier spectrum and integral of the Fourier effect in the vertical direction of the inner bearing, top; rear bearing, below. Induced-draft fan 1A.
Source: Authors

stressing its indispensability in predicting design outcomes and enhancing performance. In the context of our project, we conducted a similar analysis, acquiring data on the first six vibration modes. This document presents findings up to the fourth mode (Figure 22). Our study revealed a proximity between the first natural vibration frequency and the operating frequency of the induced-draft fan. This correlation provides enough evidence to attribute the excessive vibration issues to the alignment of the first natural vibration frequency with the fan's operating frequency. Based on our results, in the subsequent vibration modes, the values move away from their operating frequency, as in Manish *et al.* (2015), Čorović and Miljavec (2020), and Noureddine and Noureddine (2022). These modes increase their numerical value as shown in Figures 21 and 22.

In the study conducted by Trebuna *et al.* (2014), a methodology similar to ours was developed. Their approach involved finite element analysis, strategically placing accelerometers in both rotor bearings, and conducting static impact tests. These measures were implemented in order to scrutinize excessive vibration in two air extraction fans. As in our case, both fans were of the same model, resulting in the same behavior. One distinctive aspect of the work by Trebuna *et al.* (2014) was its focus on air extraction fans with mixed-flow characteristics, combining both flow directions (axial and radial). In this vein, three accelerometers were required for instrumentation: one of the accelerometers was placed in the axial direction of the rotor, while, for the horizontal and vertical directions, two accelerometers were used to assess the radial flow effects. Given that the air extraction fans operated within a speed range of 400-1 000 rpm, modal and experimental analyses were conducted at distinct frequency values in orded to capture the nuances of their behavior.

Similarly, Tarek *et al.* (2018) employed accelerometers positioned in all three directions. In an effort to rectify fan behavior issues, their maintenance department opted for a comprehensive replacement of the rotor, encompassing the shaft, fan, and roller bearings. In contrast, in our project, we placed the accelerometers in the two necessary directions: horizontal and vertical. This approach was prompted by the radial flow characteristics of the induced-draft fan under study. Through our experimental and modal analyses, we identified a distinctive pattern: the fan exhibited an operating frequency closely aligned with the first natural frequency of the system. Consequently, we recommend a design modification strategy centered around altering the rotor's diameter, including the replacement of the fan rotor itself.

Our methodology was shaped by a combination of factors, including an assessment of ineffective previous approaches, our team's extensive experience, and a thorough analysis

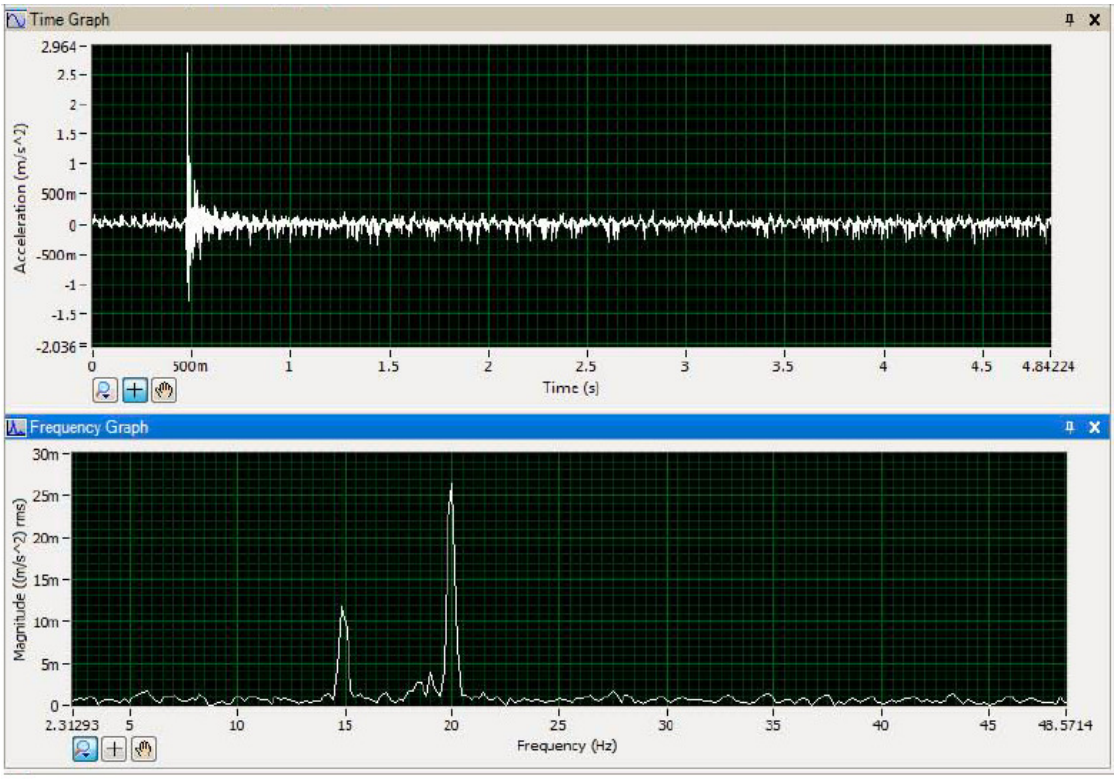
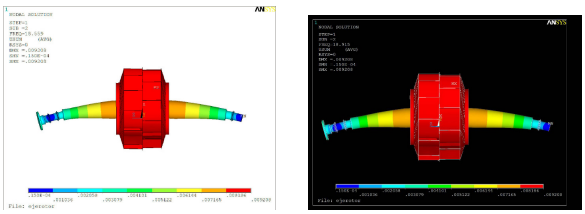


Figure 20. Results of the impact test. 1B induced-draft fan.
Source: Authors



(a) (b)

Figure 21. First mode of vibration of the original model without bearings: (a) natural frequency of 18.559 Hz, 1026 steel; (b) natural frequency of 18.915 Hz, 4340 steel
Source: Authors

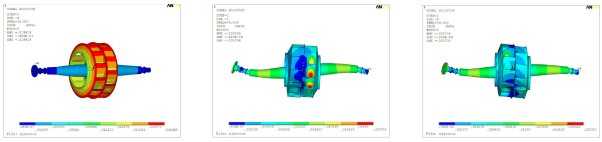


Figure 22. Main vibration modes of the fan rotor with a 1026 steel central shaft and original diameter dimensions. Left: second mode of vibration, with a frequency of 20.091 Hz. Center: third mode of vibration, with a frequency of 79.628 Hz. Right: fourth mode of vibration, with a frequency of 79.632 Hz.
Source: Authors

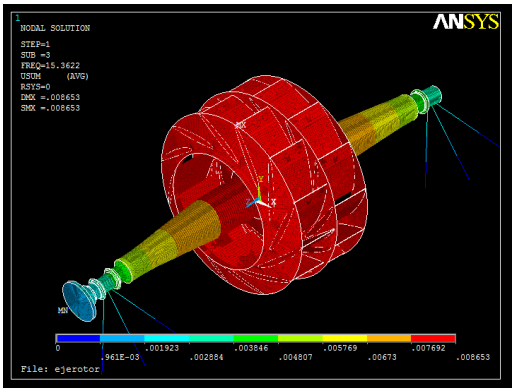


Figure 23. First mode of vibration of the original model with bearings, with a natural frequency obtained of 15.3622 Hz.
Source: Authors

of the induced-draft fan's performance. We also considered and explored alternative solutions to tackle the studied issue.

We proposed a solution based on instrumentation and computer simulation to address the excessive vibration issue, which had not been considered by companies in the public and private sectors and company maintenance personnel. The root cause of the problem was the need for more analysis and communication between the fan manufacturer and the company's engineering and maintenance area, as they were operating in areas close to the fan's resonance, *i.e.*, they failed to consider that the first natural frequency of the system would be close to its operating frequency in the plant.

To support our proposal, Table 4 should be considered, as it shows that an increase in the rotor's diameter raises the system's natural frequency, causing it to move away from

the operating frequency or speed. This finding aligns with turbomachinery principles, as it is advisable for the natural frequencies to exceed the operating one.

Conclusions

In this work, mechanical vibratory analysis was conducted to address the issue of excessive vibrations in two induced-draft fans at a thermal power plant in Mexico. Excessive vibrations occurred suddenly and frequently, causing continuous shutdowns to balance the rotors, thus generating economic losses.

The research methodology comprised four key processes: a) analyzing previous solutions to the problem, b) conducting a literature review encompassing background and standards, c) performing experiments by instrumenting accelerometers at crucial points in the fans with the purpose of determining their natural frequencies (a crucial component for subsequent vibration analysis), and d) validating the natural frequencies through modeling and computer simulation. This validation allowed defining a solution to mitigate the problem. It is worth adding that the computer simulation was based on finite element analysis.

Based on the data obtained from both experimentation and simulation, the following was observed:

- The first natural frequency of the system is closely aligned with the operating frequency of the induced-draft fans. Consequently, the system is close to mechanical resonance, resulting in excessive vibration.
- After reviewing the available literature and considering the essential operation of the fans under analysis, implementing a variable frequency drive to adjust the operating speed was not considered a viable option.

In light of the results, it is imperative to modify the fans' geometry. This is necessary because the first natural frequency is close to the operating frequency or impeller speed. In this context, the most practical approach would involve modifying the shaft's cross-section to one with a larger diameter. Although the expenses associated with this solution may be substantial, it remains a necessary measure to rectify the initial design flaws and address the ongoing issues.

Future work could leverage real-time monitoring, computer modeling, and simulation tools, along with the integration of deep learning methodologies. These approaches are poised to offer crucial insights into the behavior of systems, aiding in identifying potential faults and developing corresponding solutions. In this case study, establishing a database lays a foundation for applying deep neural network learning in forthcoming projects. Another improvement would be enhancing the accuracy of sensitivity measurements, similarly to the methods outlined in this study, which could be further advanced through techniques such as speckle interferometry (Dhiya *et al.*, 2023; Jesús *et al.*, 2024).

CRedit author statement

Erick-Alejandro González-Barbosa conceived the idea and conducted preliminary research. José Juan Vázquez Martínez conducted further research. Fernando Jurado-Pérez oversaw the evaluation process and provided research supervision. Héctor Castro Mosqueda contributed to the methodology, software development, and original draft writing. Francisco Javier Rodríguez Ornelas conducted the literature review, formal analysis, and data collection processes. José-Joel González-Barbosa assisted with data collection and supervision, and he offered valuable feedback. All authors contributed to the manuscript and approved its final version for publication.

Conflicts of interest

The authors declare that they have no conflict of interest

References

Benchekroun, M. T., Zaki, S., Hezzem, B., and Laacha, H. (2023). Kiln process fan vibrations prediction based on machine learning models: Application to the raw mill fan. In *Computer sciences and mathematics forum* (Vol. 6, p. 6 https://doi.org/10.3390/cmsf2023006006.).

Benrahmoune, M., Ahmed, H., Mouloud, G., and XiaoQi, C. (2018). Detection and modeling vibrational behavior of a gas turbine based on dynamic neural networks approach. *Strojnícky casopis-Journal of Mechanical Engineering*, *68*(3), 143–166 https://doi.org/10.2478/scjme-2018-0032.

Blanco-Ortega, A., Beltrán-Carbajal, F., Silva-Navarro, G., and Méndez-Azúa, H. (2010). Control de vibraciones en maquinaria rotatoria. *Revista Iberoamericana de Automática e Informática Industrial RIAI*, *7*(4), 36–43 https://doi.org/10.1016/S1697-7912(10)70058-3.

Chapagain, A., and Silwal, B. (2023). Influence of rotor eccentricity on large synchronous generator. In *Ressd 2023 international conference on role of energy for sustainable social development* (p. 1-4).

Čorović, S., and Miljavec, D. (2020). Modal analysis and rotor-dynamics of an interior permanent magnet synchronous motor: An experimental and theoretical study. *Applied Sciences*, *10*(17), 5881 https://doi.org/10.3390/app10175881.

Cui, L., Wang, X., Xu, Y., Jiang, H., and Zhou, J. (2019). A novel switching unscented kalman filter method for remaining useful life prediction of rolling bearing. *Measurement*, *135*, 678-684 https://doi.org/10.1016/j.measurement.2018.12.028.

Dhamande, L. S., Bhaurkar, V. P., and Patil, P. N. (2023). Vibration analysis of induced draught fan: A case study. *Materials Today: Proceedings*, *72*, 657–663 https://doi.org/10.1016/j.matpr.2022.08.329.

Dhiya, F., R., M. B., and A., Y. H. (2023, dic.). Interferómetro mach-zehnder recubierto de óxido de grafeno basado en sensor de gas de amoníaco. *Nexo Revista Científica*, *36*(06), 1132-1140 https://doi.org/10.5377/nexo.v36i06.17469.

Di, H., Chen, Z., Tao, Y., and Gang, C. (2022). An intelligent anomaly detection method for rotating machinery based on vibration vectors. *IEEE Sensors Journal*, *22*(14), 14294-14305 https://doi.org/10.1109/JSEN.2022.3179740.

Donolo, P., Bossio, G., De Angelo, C., García, G., and Donolo, M. (2016). Voltage unbalance and harmonic distortion effects on induction motor power, torque and vibrations. *Electric power systems research*, *140*, 866–873 https://doi.org/10.1016/j.epsr.2016.04.018.

Doshi, S., Katoch, A., Suresh, A., Razak, F. A., Datta, S., Madhavan, S., ... Gundabattini, E. (2021). A review on vibrations in various turbomachines such as fans, compressors, turbines and pumps. *Journal of Vibration Engineering and Technologies*, *9*(7), 1557–1575 https://doi.org/10.1007/s42417-021-00313-x.

Guo, R., Zhang, G., Zhang, Q., Zhou, L., Yu, H., Lei, M., and Lv, Y. (2021). An adaptive early fault detection model of induced draft fans based on multivariate state estimation technique. *Energies*, *14*(16), 4787 https://doi.org/10.3390/en14164787.

Jagtap, H. P., Bewoor, A. K., and Kumar, R. (2020). Failure analysis of induced draft fan used in a thermal power plant using coordinated condition monitoring approach: A case study. *Engineering Failure Analysis*, *111*, 104442 https://doi.org/10.1016/j.engfailanal.2020.104442.

Jerzy, C., Przemyslaw, M., and Piotr, O. (2014). Tests of rotary machines vibrations in steady and unsteady states on the basis of large diameter centrifugal fans. *Eksploatacja i Niezawodność*, *16*(2), 211–216.

Jesús, T., Liline-Daniel, C.-H., Alejandra, A. P., José-Joel, G.-B., and Juan-B, H.-R. (2024). Vibrations of a violin while the f-holes were sequentially cut. *Acoustics Australia*, *52*(1). https://doi.org/10.1007/s40857-024-00313-3).

Jorgen, W. L., Elie, B. A., Cheng, H. S., W., P. C., and Coda, H. (1965). *Rotor bearings dynamic design technology, part iii:: desing handbook for fluid film type bearings* (1st ed.). Mechanical Technology Inc.

Kalmár-Nagy, T., Bak, B. D., Benedek, T., and Vad, J. (2015). Vibration and noise of an axial flow fan. *Periodica Polytechnica Mechanical Engineering*, *59*(3), 109–113 https://doi.org/10.3311/PPme.7948.

Kaneko, Y., Kawashita, R., and Kanki, H. (2022). 8 - steam turbine rotor design and rotor dynamics analysis. In T. Tanuma (Ed.), *Advances in steam turbines for modern power plants (second edition)* (Second Edition ed., p. 163-194). Woodhead Publishing. Retrieved from https://doi.org/10.1016/B978-0-12-824359-6.00011-1

Li, P., Pang, L., and Lin, Z. (2020, mar). Vibration fault diagnosis and dynamic balance processing analysis of blower in thermal power plant. *IOP Conference Series: Earth and Environmental Science*, *467*(1), 012110 https://doi.org/10.1088/1755-1315/467/1/012110.

Liu, R., Yang, B., Zio, E., and Chen, X. (2018). Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mechanical Systems and Signal Processing*, *108*, 33–47 https://doi.org/10.1016/j.ymssp.2018.02.016.

Lu, H., Pavan Nemani, V., Barzegar, V., Allen, C., Hu, C., Laflamme, S., ... Zimmerman, A. T. (2023). A physics-informed feature weighting method for bearing fault diagnostics. *Mechanical Systems and Signal Processing*, *191*, 110171 https://doi.org/10.1016/j.ymssp.2023.110171.

M, J., and K, P. (2023). Matlab-based fault diagnosis of industrial rotor-bearing systems. *Insight-Non-Destructive Testing and Condition Monitoring*, *65*(5), 278–283 https://doi.org/10.1784/insi.2023.65.5.278.

Manish, D., Kumar, J. S., Vikas, S., Kumar, S. S., and Dhirendra, A. (2015). Fatigue (FEA) and modal analysis of a centrifugal fan. *International Journal of Recent advances in Mechanical Engineering (IJMECH)*, *4*(2), 77–91 https://doi.org/10.14810/ijmech.2015.4209.

Mohamad, T. H., Abbasi, A., Kappaganthu, K., and Nataraj, C. (2023). On extraction, ranking and selection of data-driven and physics-informed features for bearing fault diagnostics. *Knowledge-Based Systems*, *276*, 110744 https://doi.org/10.1016/j.knosys.2023.110744.

Ni, Q., Ji, J., Halkon, B., Feng, K., and Nandi, A. K. (2023). Physics-informed residual network (piresnet) for rolling element bearing fault diagnostics. *Mechanical Systems and Signal Processing*, *200*, 110544 https://doi.org/10.1016/j.ymssp.2023.110544.

Niko, L., Jari, P., and Esa, P. (2011). The effect of foundation on fan vibration response. *Journal of Structural Mechanics*, *44*(1), 1–20.

Noureddine, A., and Noureddine, M. (2022). Computational investigation of vibration characteristics analysis for industrial rotor. *Acta Mechanica et Automatica*, *16*(4 https://doi.org/10.2478/ama-2022-0044.).

Novotný, P., Hrabovský, J., Juračka, J., Klíma, J., and Hort, V. (2019). Effective thrust bearing model for simulations of transient rotor dynamics. *International Journal of Mechanical Sciences*, *157*, 374–383 https://doi.org/10.1016/j.ijmecsci.2019.04.057.

Pingchao, Y., Dayi, Z., Yanhong, M., and Jie, H. (2018). Dynamic modeling and vibration characteristics analysis of the aero-engine dual-rotor system with fan blade out. *Mechanical Systems and Signal Processing*, *106*, 158–175 https://doi.org/10.1016/j.ymssp.2017.12.012.

Qingjie, Z., Guangxiang, L., You, X., and Chengyu, Z. (2020). Torsional vibration analysis of shaft in an induced draft fan due to variable frequency drive. *IEEE Access*, *8*, 174723–174735 https://doi.org/10.1109/ACCESS.2020.3026325.

Ren, G.-P., Zhang, H.-T., Wu, Y., and Ding, H. (2023). A general double-input synchronous signal processor for imbalanced vibration mitigation in amb-rotor systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *53*, 3823–3832 https://doi.org/10.1109/TSMC.2022.3232517.

Shabaneh, N., and Zu, J. W. (2003). Nonlinear dynamic analysis of a rotor shaft system with viscoelastically supported bearings. *Journal of Vibration and Acoustics*, *125*(3), 290–298 https://doi.org/10.1115/1.1547684.

Tarek, K., Nouredine, O., and Abderrazek, D. (2018). Experimental vibratory analysis of a fan motor in industrial environment. *The International Journal of Advanced Manufacturing Technology*, *98*(9-12), 2439–2447 https://doi.org/10.1007/s00170-018-2391-1.

Trebuna, F., Simcák, F., Bocko, J., Hunady, R., and Pástor, M. (2014). Complex approach to the vibrodiagnostic analysis of excessive vibration of the exhaust fan. *Engineering Failure Analysis*, *37*, 86-95.

Wang, Z., Chen, Y., Ouyang, H., and Wang, A. (2020). Investigation of vibration characteristics of titanium wide-chord fan blade. *Journal of Vibration Engineering and Technologies*, *8*, 529–540 https://doi.org/10.1007/s42417-019-00108-1.

Wei, W., You, L., Yizhuo, L., Qing, L., Shisen, X., and Jun, L. (2022). An early fault detection method of the induced draft fan based on long-short term memory network and double warning thresholds. In *2022 34th chinese control and decision conference* (*ccdc*) (p. 5262-5267 https://doi.org/10.1109/CCDC55256.2022.10033743.).

Xiangyang, D., Shiqiang, C., Zhenlin, L., Zhulong, Z., and Yihan, C. (2023). Experimental study on fan aerodynamic noise variation characteristics under non-proportional variation law. *Sustainability*, *15*(3 https://doi.org/10.3390/su15032025.).

Xie, Z., Yang, K., He, T., and Jiao, J. (2023). Experimental and theoretical analysis on the nonlinear rotor-dynamic performances and vibration characteristics of a novel bearing-rotor system. *Mechanical Systems and Signal Processing*, *199*, 110416 https://doi.org/10.1016/j.ymssp.2023.110416.

Yang, Y., Zhang, Y., Zeng, J., Ma, H., Yang, Y., and Cao, D. (2024). Insight on uncertainty of geometrically nonlinear rotor with rub-impact under maneuvering motion. *Journal of Sound and Vibration*, *570*, 118018 https://doi.org/10.1016/j.jsv.2023.118018.

Yu-Ling, H., Ling, T., Kai, S., Wen-Hao, Z., Xue-Wei, W., and Hai-Peng, W. (2023). Impact of static air-gap eccentricity fault on synchronous generator efficiency. *Energies*, *16*(7 https://doi.org/10.3390/en16073294.).

Zenglin, G., and Gordon, K. R. (2003). Instability boundary for rotor-hydrodynamic bearing systems, part 1: Jeffcott rotor with external damping. *Journal of Vibration and Acoustics*, *125*(4), 417–422 https://doi.org/10.1115/1.1547466.

Research Article/Mechanical Engineering, Mechatronics, and Materials Science http://doi.org/10.15446/ing.investig.108697

Perspectives in the Study of Renewable Gaseous Fuels Autoignition at Low and Moderate Temperatures

Perspectivas en el estudio de la autoignición de combustibles renovables gaseosos en temperaturas moderadas y bajas: una revisión

Hernando A. Yepes[1], Adalberto Salazar[2], and José D. Yepes[3]

ABSTRACT

The increased energy demand and restrictions regarding pollutant emissions have sparked the search for different renewable energy sources, as well as for the improvement of thermal processes, with a focus on reducing the consumption of fossil fuels. Renewable gaseous fuels seem to be a promising alternative for solving this issue, along with the different high-efficiency, low-emissions technologies that operate at low and moderate temperatures (600-1000 K). However, the implementation of these approaches is limited by the autoignition phenomenon and the different difficulties in predicting its occurrence in the aforementioned operation range. To identify the reasons for this fact, we carried out a review of the different research works conducted in the field. It was evidenced that most studies focus on performing adjustment processes that require prior experimentation. This allowed identifying the need to conduct a research work focused on the autoignition phenomenon in the low and moderate temperature range while using renewable gaseous fuels, as well as on improving the predictive models for calculating ignition delay times.

Keywords: autoignition, gaseous renewable fuels, detailed reaction mechanism, ignition delay time

RESUMEN

El aumento en la demanda energética y las restricciones respecto a la emisión de contaminantes han suscitado la búsqueda de diferentes fuentes de energía renovables, así como del mejoramiento de los procesos térmicos enfocados en reducir el consumo de combustibles fósiles. Los combustibles renovables gaseosos se muestran como una alternativa útil para resolver este problema, al igual que diferentes tecnologías de alta eficiencia y bajas emisiones que operan a temperaturas bajas y medias (600-1000 K). Sin embargo, la implementación de estos enfoques se ve limitada por el fenómeno de autoignición y las diferentes dificultades para predecir su aparición en el mencionado rango de operación. Para identificar las razones de este último hecho, se llevó a cabo una revisión de los diferentes trabajos de investigación realizados en el área. Se evidenció que la mayoría de los estudios se centran en realizar procesos de ajuste que requieren una experimentación previa. Lo anterior permitió identificar la necesidad de llevar a cabo una investigación enfocada en el fenómeno de autoignición en el rango de bajas y medias temperaturas usando combustibles renovables gaseosos y en el mejoramiento de los modelos predictivos para el cálculo del tiempo de retraso de la ignición.

Palabras clave: autoignición, combustibles renovables gaseosos, mecanismo detallado de reacción, tiempo de retraso de la ignición

Received: May 2th, 2023
Accepted: April 16th, 2024

Introduction

The different climate challenges and increased pollutant emissions have caused the use of renewable energies in the global energy market to become more relevant. Prior to the COVID-19 pandemic, an increase in the utilization of this kind of energy was observed over a period of five years, reaching 12.4 and 29% of the global energy demand in 2023 and 2040, respectively (Chen *et al*., 2020; US Energy Information Administration, 2016).

During the first phase of the pandemic, a significant decrease in greenhouse gas emissions was recorded. CO_2 levels decreased by 17% in April 2020 compared to the same month in 2019 (Kuzemko *et al*., 2020), and a general 12.2% reduction in CO_2 emissions of was recorded in 2020 (Q. Wang and Wang, 2020). These reductions are related to the decrease in energy consumption by the industry and business sectors due to the restrictions imposed by the sanitary authorities. In contrast, an increase in domestic energy consumption was recorded (B. Wang *et al*., 2020). However, after the end of the lockdown, and due to the economic reactivation, a significant increase in energy consumption and pollutant

[1] Mechanical engineer, Universidad Nacional de Colombia –Medellín Campus, Colombia. MSc in Engineering, with an emphasis on Energy, Universidad de Antioquia, Colombia. Affiliation: Assistant professor, Department of Mechanical Engineering, Applications in Thermofluids, Energy, and Advanced Nanomaterials – ATENA, Universidad Francisco de Paula Santander, Ocaña, Colombia. PhD candidate in Mechanical Engineering, Universidad Técnica Federico Santa María, Chile. Email: hayepest@ufpso.edu.co, hernando.yepes@sansano.usm.cl

[2] Mechanical engineer, Universidad del Norte, Colombia. MSc in Mechanical Engineering, Universidad Técnica Federico Santa María, Chile. Affiliation: Department of Mechanical Engineering, Universidad Técnica Federico Santa María, Chile. Email: adalberto.salazar@sansano.usm.cl

[3] Engineering physicist, Universidad Nacional de Colombia – Medellín Campus, Colombia. Email: jdyepest@unal.edu.co

Nomenclature

β	Weight factor in the Sankaran expression	NTC	Negative temperature coefficient
ϕ'	Equivalence ratio fluctuations	RCM	Rapid compression machine
τ	Ignition delay time	R_{ef}	Reynolds flame number
τ_f	Characteristic flame time	RGF	Renewable gaseous fuels
T'	Temperature fluctuations	S	Factor considering the correlation of T and ϕ
D_{al}	Damköhler number	S_{ap}	Sankaran number
DRM	Detailed reaction mechanisms	SG	Syngas
ICE	Internal combustion engines	ST	Shock tube
MN	Methane number		

emissions was to be expected. This represented a challenge, especially for developing countries (Q. Wang and Wang, 2020). Other factors contributed to this issue, as is the case of the increase in private transport, since many people chose this alternative during the pandemic to avoid agglomeration in public transport (Cheshmehzangi, 2020). Considering that industry and transport are the highest energy consumers in many Latin American countries, whose requirements are provided for by oil-derived fuels sources, *e.g.*, natural gas for electricity production, these sectors are potential targets for the implementation of renewable energies and the reduction of greenhouse gas emissions (Burdack *et al*., 2023; Grangeia *et al*., 2023; Simsek *et al*., 2019; Wolde-Rufael and Mulat-Weldemeskel, 2022).

Renewable gaseous fuels (RGFs) are promising alternatives for achieving the aforementioned objective, especially those containing hydrogen (H_2). This component has many benefits during the combustion process in terms of emissions and efficiency, given its high reactivity and burning velocity (Elsemary *et al*., 2016). However, the addition of RGF in thermal devices cannot be arbitrary, as it may entail undesired phenomena due to changes in combustion properties. Mixture reactivity increases when RGF is added, especially in the presence of H_2, and early autoignition may occur before reaching the combustion chamber, implying significant damage to the system. In internal combustion engines (ICEs), the addition of great large amounts of RGF can produce knocking phenomena. If the compression ratio is increased to achieve higher efficiency, this issue could take place (Arunachalam and Olsen, 2012; Szwaja *et al*., 2013; Z. Wang *et al*., 2017).

In recent years, the relevance of RGFs as an energy source has grown, with an annual rate of 10.8% until 2016, but technical issues have limited their contributions to the energy supply (International Energy Agency, 2016, 2018). In this sense, understanding the phenomenon of RGF autoignition becomes relevant with regard to broadening their use in thermal devices. The most common parameter used to study autoignition in fuels is the *ignition delay time* (τ), defined as the time required by a reactive mixture under specific thermodynamic conditions to begin the reaction by itself without any external energy source. Since the 90s, interest in measuring τ for RGFs has increased, and several studies have been conducted. The experimental values of τ have been used to develop detailed reaction mechanisms (DRM) in order to predict and analyze autoignition. τ prediction is a crucial aspect in improving the design of thermal devices, as it allows controlling this phenomenon during operation and preventing its occurrence in zones that can produce damage.

For several years, most of the τ data used to adjust and validate DRMs corresponded to high temperature conditions (>1000 K), since many combustion devices operated in this range. However, due to environmental regulations, international agreements, improvements to industrial processes, and the development of new technologies and techniques operating at moderate (800-1000 K) and low (600-900 K) temperatures, *e.g.*, lean and diluted mixtures, recirculation gases, flameless combustion, HCCI engines, among others, this trend has changed in recent years. Thus, there was an increase in the number of studies on autoignition and experimental measurements of τ in these ranges. Nevertheless, τ prediction under these conditions via the zero-dimensional approach still exhibits significant disagreements, mainly when testing RGFs. The main issue is that the zero-dimensional model, which uses the DRM, overpredicts the value of τ by more than one order of magnitude, which can clearly generate oversizing or failures in a combustion device. In order to solve this issue, different adjustment approaches have been developed, which are presented in the following sections, although many of them have several limitations, since they focus on corrective actions and do not address the phenomenon in order to fix prediction issues. Most of these approaches require prior experimental measurements, which limits their application in the field of RGFs, whose chemical composition is highly variable. This hinders progress in the use of this kind of fuel.

Although some reviews on autoignition can be found in the literature, most of them focus on the behavior of engines (Chintala and Subramanian, 2017; Dimitriou and Javaid, 2020; Rönn *et al*., 2023). Some of them also encompass advances in kinetics and experimental setups to perform measurements (Goldsborough *et al*., 2017; Sung and Curran, 2014; Zádor *et al*., 2011). However, the literature review

carried out in this study could not find a work focused on the issues of autoignition prediction at low temperatures while using RGF, on the connection with the autoignition regime, or on modeling perspectives. In this vein, it is necessary to adequately understand the autoignition phenomenon under moderate and low temperature conditions, as well as its effect on RGFs compared to conventional fuels. The first step to achieving this objective is to understand the actual state of the studies related to this phenomenon and their considerations for further research. The main objectives of this work are to determine the current landscape of autoignition studies in relation to RGFs and to identify the remaining gaps. Furthermore, this review can be used as a starting point to propose and conduct additional studies focused on improving models and DRMs related to RGFs. The *General features* section presents some fundamental concepts related to τ and a brief description of the experimental setups used to measure it. In *Disagreement at low temperatures*, we outline the existing disagreement between experimental and numerical results for τ at moderate and low temperatures. Different approaches for τ prediction are discussed in the *Methodologies to improve τ prediction*. Finally, the section titled *Inhomogeneity effects* addresses the relationship with the ignition regime.

Methodology

The scope of this work encompasses the autoignition phenomenon. Therefore, our literature review focused on scientific articles in specialized combustion journals from major publishers such as ScienceDirect, Taylor and Francis, and ACS, among others. All selected works belong to journals indexed in the Web of Science database. Experimental and numerical τ results, along with the disagreement reported between them when using RGFs, were the main topics of the selected studies. The reported results were classified into three global aspects, as presented in the next sections. A critical analysis was performed to identify the current gaps in autoignition modeling at low and moderate temperatures while using RGFs. Additionally, the effect of H_2 on the disagreement between zero-dimensional model predictions was analyzed, as well as the relationship between the autoignition regime and abnormal combustion in ICEs. All calculations were carried out in the Cantera software, modeling a closed, constant-volume, adiabatic, and completely mixed reactor while only considering time as an independent variable. Surface reactions were not considered. Five DRMs were used in the calculations in order to test the effects of different kinetic models on τ prediction capabilities at low and moderate temperatures.

General features

Autoignition

Under specific pressure and temperature conditions, *autoignition* can be described as the phenomenon resulting from the start of the reaction process in a fuel-oxidant mixture without the assistance of an external energy source. Thereupon, the autoignition process is not immediate and is considered a transient phenomenon. The best way to describe this phenomenon is to take time into account. As previously mentioned, ignition delay is the most frequently used parameter with regard to the autoignition process. *Ignition delay time* is defined as the time needed by the reactive mixture to start and maintain the reaction process under fixed thermodynamic conditions. Temperature increments are regarded as minimal during this time. Only when the radical concentration in the mixture reaches a certain level does ignition take place; these radicals release the necessary energy through fuel consumption, leading to the combustion of the mixture (Warnatz *et al.*, 1996).

Furthermore, several research groups have developed different numerical and experimental approaches to determine the value of τ. Generally, the most common method consists of determining the highest concentration point of the intermediate species OH, or determining the interval until the inflection point is reached in the pressure profile (Man *et al.*, 2013; Tang *et al.*, 2013; B. L. Wang *et al.*, 2003). Since τ is a function of the elementary reactions associated with the type of fuel, and the reaction rates are highly dependent on temperature, it is accurate to deduce that τ is intrinsically temperature-dependent. The relationship between these parameters can be described using the Arrhenius form as follows:

$$\tau = A exp\left(\frac{B}{T}\right) \quad (1)$$

where A is a factor related to the reactive concentration and the preexponential coefficient, and B is linked to the activation energy, assuming a one-step global reaction. Nevertheless, this document demonstrates that this approach is more accurate under high temperature conditions.

Experimental measurement of τ

Calculating τ is crucial for developing models that address complex combustion processes in different thermal devices, as well as for building DRMs that contribute to the aforementioned objective. Therefore, various thermal reactors have been used to control pressure and temperature conditions, thereby facilitating the calculation of this parameter. However, not all reactors are suitable for measuring τ across varying pressure ranges and temperatures. Figure 1 shows the operating range of devices used to determine τ experimentally. According to the literature, shock tubes (STs) and the rapid compression machines (RCMs) are the two most popular thermal reactors, as a large portion of thermal equipment operates under similar conditions.

An ST consists of a long-closed tube, divided into the *driver section* (high pressure) and the *driven section* (low pressure). These sections are separated by a diaphragm. The experiment fills the driven section with the test

mixture while an inert gas is introduced into the second section to increase the pressure until the diaphragm ruptures. A shock wave travels rapidly to the driven section, where the pressure and temperature of the test mixture increase. Subsequently, the wave is reflected at the end of this section, generating additional compression and allowing the necessary reactions for ignition to begin. The value of τ is determined by measuring the time elapsed from the initial heating until the rapid release of energy or the formation of radicals. Therefore, in many cases, chemiluminescence measurements via some radical such as OH and induced excitation (S. Wang *et al.*, 2019) are employed.

RCMs can have different versions, but, in general, all have a test or driven section inside a cylinder where the test mixture is located, which is rapidly compressed by the movement of a piston toward the test section. The test section consists of a small chamber, like a combustion chamber in an engine. A sensor is used throughout the procedure to continuously record the pressure-time profile. In general, this refers to the time between the end of compression and the location where the temporal pressure reaches its maximum variation (Liu *et al.*, 2018).

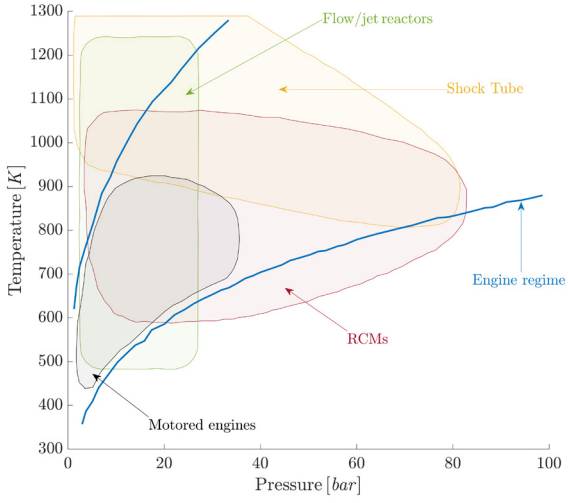


Figure 1. Operating ranges of the devices used for τ measurements
Source: Adapted from Goldsborough *et al.* (2017)

Disagreement at low temperatures

Several studies have focused on autoignition measurement and modeling by means of τ calculation. The most common model to determine τ is the adiabatic, homogeneous, close, constant-volume reactor, where the reactive mixture is considered to be an ideal gas. In general, for gaseous fuels, and especially for RGFs, the calculated values of τ at high pressures and low temperatures (600-950 K) shows significant disagreement with experimental values. Figure 2 presents a comparison between the experimental and numerical ignition delay values of $CH_4/C_2H_6/C_3H_8/H_2$, as obtained by Kuppa *et al.* (2018) using two different detailed reaction mechanisms. A better agreement can be seen in the low ignition delay values in comparison with the high ones (>200 μs). The former data group is associated with high temperatures, where the DRM and autoignition models have been more extensively studied, leading to improved performance and precision. On the other hand, for the high ignition delay values related to low and moderate temperatures, the scatter increases and becomes more extensive as ignition delay increases.

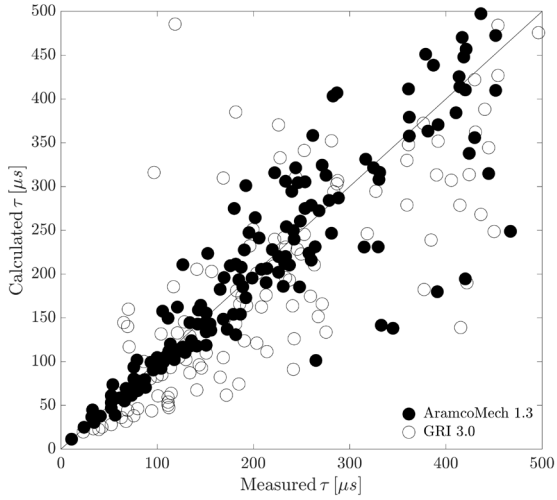


Figure 2. Ignition delay comparison for $CH_4/C_2H_6/C_3H_8/H_2$ mixtures
Source: Adapted from Kuppa *et al.* (2018)

Figure 3 shows the experimental data collected by Zhang *et al.* (2012) for CH_4 ignition delay times, along with the predictions obtained using different DRMs. The thermodynamic conditions correspond to high temperature and pressure. The excellent agreement between experimental and numerical data is evident, regardless of which DRM was used. This behavior suggests an adequate chemical kinetic description by both the DRM-evaluated and the model approximations. However, the trend changes as the temperature decreases. Figure 4 compares the predicted and experimentally measured ignition delay times for a natural gas mixture composed of 78.8% CH_4, 14% C_2H_6, 3.4% C_3H_8, and higher hydrocarbons (% in vol.) (Vallabhuni *et al.*, 2018). It is evident that a good agreement between experimental and numerical results can only be achieved for temperatures higher than 1000 K. The opposite behavior is observed at lower temperatures.

The model cannot capture the observed behavior at these temperatures. In this case, the predicted values are lower than the experimental ones. This can be considered safe from a practical standpoint, as autoignition will occur later experimentally than is expected numerically., but it is evident that the model cannot capture the actual process. According Vallabhuni *et al.* (2018), the deviations are due to the high concentration of hydrocarbons heavier than C_2. However, a detailed explanation is missing in the manuscript.

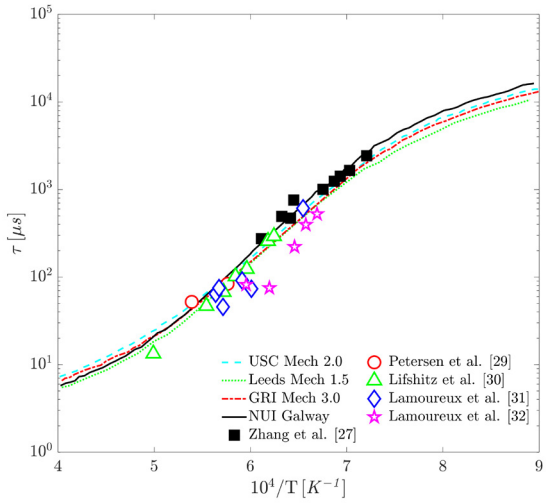


Figure 3. Ignition delay time comparison for CH_4
Source: Data from Lamoureux *et al*. (2002), Lamoureux and Paillard (2003), Lifshitz *et al*. (1971), Petersen, Hall, *et al*. (2007), and Zhang *et al*. (2012)

In the case of RGFs, the behavior is similar, but the disagreement increases at low temperatures. A comparison between the experimental and numerical values of τ, carried out by McDonell and Dunn-Rankin (2008), is shown in Figure 5. It is evident that the models cannot reproduce the observed behavior of τ at lower temperatures with any of the DRMs used. According to the authors, the low accuracy of the models may be related to the fact that H_2 autoignition does not occur in a homogenous manner. However, this is also observed in fuel mixtures without H_2, so the disagreement cannot be attributed to this factor. Moreover, the work by Davidson *et al*. (2019), for τ values of gasoline surrogates, reported disagreements in the 700-900 K temperature range, thereby supporting that the presence of H_2 in the fuel mixture is not the only reason for this behavior.

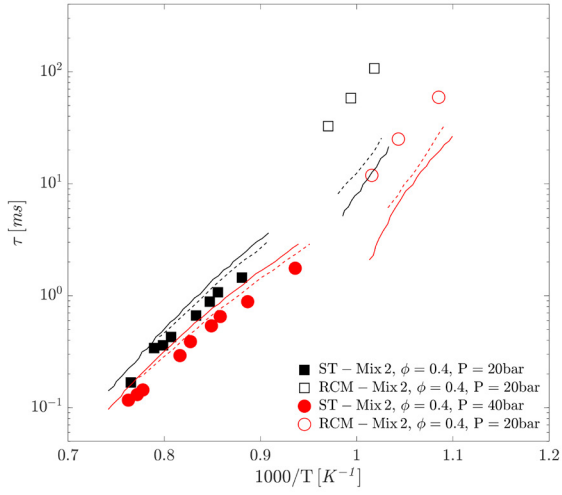


Figure 4. Ignition delay for natural gas mixture. Experimental data (symbols). Numerical data (lines) using the LLNL (straight) and NUIG mechanisms (dashed).
Source: Adapted from Vallabhuni *et al*. (2018)

Syngas (SG) also exhibits a similar behavior. Figure 6 shows ignition delay values for an SG/air mixture (7.33% H_2, 9.71% CO, 1.98% CO_2, 17.01% O_2, and 63.97% N_2 in vol.) (Yu *et al*., 2018). Significant disagreements between the experimental and numerical results appear at low temperatures. The same trend was reported by Barak et al. (2017) for an equimolar H_2/CO fuel mixture. In addition, Lee et al. (2014) collected ignition delay data for different SG compositions and highlighted the need to obtain more experimental data on these fuels at low temperatures, along with their pressure curves, in order to improve the model predictions.

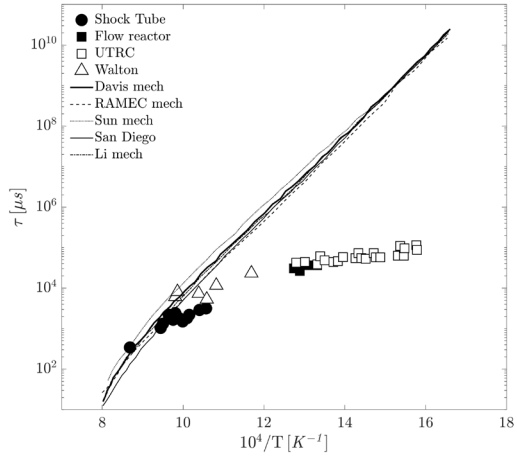


Figure 5. Experimental and numerical ignition delay values for H_2
Source: Adapted from McDonell and Dunn-Rankin (2008)McDonell and Dunn-Rankin (2008)

Hu *et al*. (2016) evaluated ten different DRMs in τ calculations for H_2/O_2 mixtures for the 850-1500 K temperature range. Only four of them allowed obtaining adequate τ predictions, exhibiting increased disagreement at low temperatures. Mansfield and Wooldridge (2014) reported the poor performance of the DRM in J. Li *et al*. (2007) regarding τ calculations for an SG, especially at temperatures lower than 1000 K. A similar behavior was observed by Donohoe *et al*. (2014) in CH_4/H_2 mixtures with the GRI-Mech 3.0 and the Aramco 1.3 reaction mechanisms.

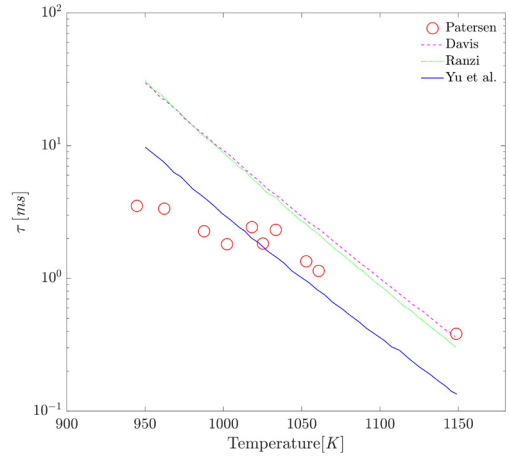


Figure 6. Ignition delay for a syngas mixture
Source: Adapted from Yu *et al*. (2018)

In order to improve the prediction capability of the models and DRMs regarding τ under low temperature conditions, different approaches have been implemented, which can be summarized in three principles: preexponential factor modifications, inclusion of the pressure rate in the model, and the consideration of catalytic effects. In the next section, some works about these principles are outlined while analyzing advantages and limitations.

Methodologies to improve τ prediction

Different methodologies have been proposed to improve the prediction capability of τ at moderate and low temperatures using zero-dimensional reactors. The main characteristic of these strategies is their corrective approach, as they use experimental results in a direct or partial manner to correct the numerical values, bringing them closer to the experimental ones. This is due to the fact that, in many cases, the objective is to use numerical results to perform a kinetic analysis, and the objective, in principle, is not to obtain an adequate prediction. In order to analyze the scope of these strategies, as well as their limitations, this work categorized them into two global groups according to the principle implemented: thermodynamics and chemical kinetics.

Thermodynamic approaches

One of the most common thermodynamic approaches is including the experimental pressure rate in the model, generally expressed as a percent per time units. This methodology is focused on considering non-ideal effects by means of the pressure increase before ignition occurs inside the reactor. Thus, the temperature rises, and the ignition acceleration is captured. As a recommendation, this approach must be applied when the experimental value of τ is greater than 1.5 ms, which implies the first and most important limitation: the need for a previously known τ before performing any calculations to improve the numerical result. On the other hand, the value of the pressure rate is not constant, and some studies report a range from 2 to 50%/ms (Lieuwen *et al*., 2009; Zhang *et al*., 2012). These values have been obtained from experimental measurements according to the pressure curve (Figure 7). It is evident that this approach can be applied when the objective is to determine ignition delay times without performing experimental tests.

Figure 8 shows some experimental τ data for a 70% H_2/30% CO (in vol.) mixture obtained in an ST, along with the numerical data generated using the SENKIN subroutine of the CHEMKIN-PRO software. The red line corresponds to the results generated with the model without any modification. As can be seen, when the temperature decreases, the disagreement between experimental and numerical data increases. The black line represents the numerical prediction using a pressure rate of 4.22%/ms in zero-dimensional model. The use of the pressure rate considerably improves the τ predictions at moderate and low temperatures. In an analogous study, albeit using only H_2 as fuel, Pang *et al*. (2009) used a value of 2%/ms. Although the prediction capability also improved, it is noticeable that the dp/dt value depends on the fuel mixture and requires previous experimental measurements.

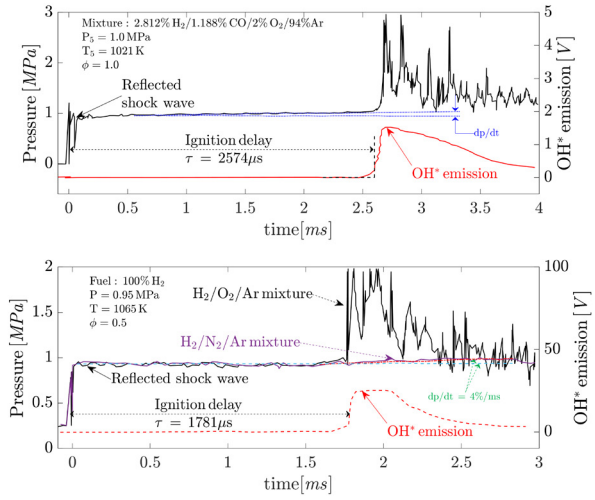


Figure 7. Pressure trace rate determination
Source: Adapted from Thi *et al*. (2014) and Zhang *et al*. (2012)

For the RCM results, the pressure and temperature at the end of compression are used to obtain τ predictions. However, in order to improve the numerical results, the pressure history curves are employed to determine the thermodynamic conditions, which must be used in the zero-dimensional model to calculate τ. The pressure equivalent is calculated as the integral average from the end of the compression to the maximum dp/dt before ignition (Walton *et al*., 2007).

The thermodynamic approaches also include dynamic predictor-corrector methodologies to improve τ predictions at moderate and low temperatures. These approaches focus on temperature correction at each time step while solving the energy equation. An example of this is the CHEMSHOCK model, which begins the solution by calculating the temperature as a conventional zero-dimensional reactor model at an instant t. Afterwards, the model assumes that the reactive mixture is isentropically compressed and recalculates the temperature for the time $t+\Delta t$. The main limitation of this approach is related to the calculation of $T_{t+\Delta t}$, wherein a measured pressure is needed, implying the need for a prior experimental test (H. Li *et al*., 2008). Figure 9 shows a comparison between experimental and numerical τ results for H_2, as obtained using the conventional constant volume reactor and CHEMSHOCK. As can be seen, the agreement improves when this approach is applied at low temperature conditions.

The approaches outlined above allow for remarkable improvements to the τ predictions, in comparison with the conventional zero-dimensional model. However, all of them require experimental measurements under the desired thermodynamic conditions. Their application is not viable for fuel mixtures with a variable chemical composition, as

is the case of RGFs and their mixtures with oil-derived fuels, where it is highly likely that experimental measurements are unavailable.

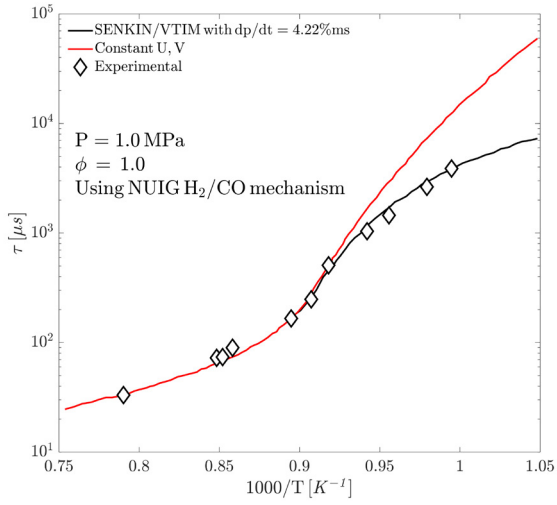


Figure 8. Ignition delay predictions, as improved by including dp/dt inclusion
Source: Adapted from Thi *et al*. (2014)

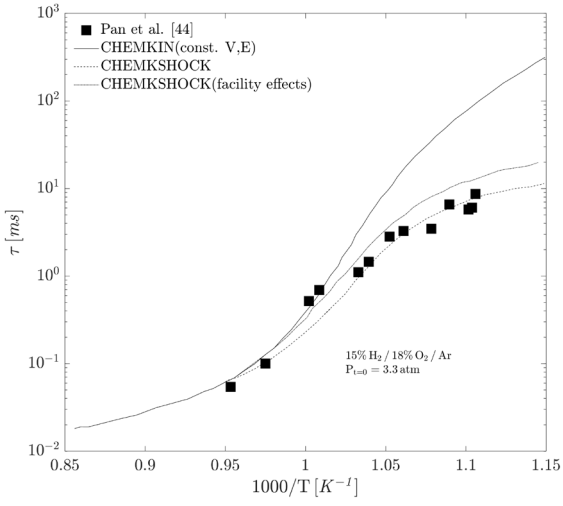


Figure 9. Ignition delay predictions for H_2, as improved using CHEMSHOCK
Source: Adapted from Pang *et al*. (2009)

Chemical-kinetic approaches

Modifying preexponential factor and kinetic parameters is another common approach to improving τ predictions (Cavaliere *et al*., 2010; Vallabhuni *et al*., 2018; Walton *et al*., 2007; Zhang *et al*., 2012). However, in this case, the main issue is that the changes are unrelated to specific reactions. Although they are commonly focused on chemical reactions in the presence of the HO_2 and H_2O_2 radicals (which is relevant in kinetics at moderate and low temperatures), there are no general rules to determine which reactions must be modified. This is dependent on the chemical composition of the fuel mixture.

Figure 10 shows a comparison between the experimental and predicted τ values reported by Amador Díaz (2017) using a DRM without any modifications and with changes in the preexponential factor of the following reactions:

$$2(OH)+M \leftrightarrow H_2O_2+M \quad (2)$$

$$H+H_2O_2 \leftrightarrow HO_2+H_2 \quad (3)$$

According to the results, only modification III significantly improves the τ predictions. Although the other alternatives do not work for this case, the studies where they were proposed exhibited good results. This suggests that the modifications cannot be universally applied, and that an adequate performance necessitates adjustments for each case – which, similarly to the thermodynamic approaches, requires experimental results.

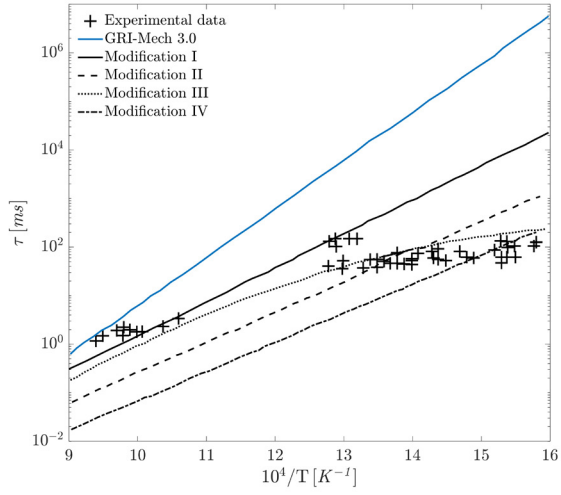


Figure 10. Ignition delay predictions improved using different modified reactions in the DRM
Source: Amador Díaz (2017)

Lee *et al*. (2017) used a genetic algorithm to obtain a reduced mechanism based on Aramco 1.3, wherein the kinetic parameters of the following reactions were modified:

$$H+O_2+CO_2=HO_2+CO_2 \quad (4)$$

$$CH_4+OH=CH_3+H_2O \quad (5)$$

The methodology was applied for H_2/CO/CH_4/CO_2 fuel mixtures. According to the authors, the reduced mechanics, along with the modified values, are suitable for calculating the ignition delays of syngas, biogas, and natural gas with the addition of H_2O, N_2, C_2H_6, and C_3H_8. However, for mixtures with C_4H_{10} or C_5H_{12}, the disagreement is significant and therefore not recommended. Another limitation of the approach becomes evident when the chemical composition is variable.

Another, different, and less common approach consists of adding new elementary reactions to the DRM, but, of course, this alternative requires important kinetic knowledge. Prince and Williams (2012) studied τ in the 500-1000 K temperature range for C_3H_8 and C_2H_6 using the San Diego DRM. When the unmodified version was used, the disagreement between experimental and numerical τ data was much greater. Although the prediction capability improved significantly with the modifications, differences remained at low temperatures. Besides, this approach also needs a fitting process that requires experimental results. In a study with a more phenomenological focus, Medvedev *et al*. (2016) used H_2 as fuel. They performed a quantum correction in the reaction rates of some initiation reactions, thereby improving the τ predictions at low temperatures and high pressures.

Zhang *et al*. (2012) conducted a numerical and experimental study using H_2, CH_4, O_2, and AR mixtures using a temperature range of 1000-2000 K and pressures from 0.5 to 2.0 MPa. In order to improve the τ predictions, the authors changed the preexponential factor of two reactions containing HO_2. Additionally, when the τ value was higher than 1.5 ms, a pressure increase rate of 4%/ms was included in the model. The differences between the experimental and numerical τ results were attributed to the uncertainties of the $H+O_2+M=HO_2+M$ reaction. However, as will be explained later, the parameters of this reaction are still being studied in light of relevant variations in τ predictions.

A review of DRMs and the subsequent proposal for a new one was performed by Jiang *et al*. (2019). The authors gathered experimental τ data for H_2/CH_4 mixtures at temperatures of 900-2000 K and pressures from 1.4 to 70 atm. The DRM obtained has 18 species and 39 reversible reactions. However, there is still disagreement for temperatures under 1000 K.

Another kinetic approach consists of compiling the seemingly most elementary reactions under moderate and low temperature conditions. Pan and Wallace (2019) applied this approach to develop the low-temperature natural gas (LTNG) reaction mechanism, focused on the combustion of natural gas mixtures. Their work included reactions considering CH_3O_2, CH_3O_2H, C_2H_5O, $C_2H_5O_2H$, and $C_2H_5O_2$ radicals in a previous DRM. Their results show an improvement in τ predictions for temperatures from 1100 and 1200 K. However, performance decreases as the temperature approaches 1000 K, with disagreements of up to four times the experimental value.

Recently, W. Li *et al*. (2023) evaluated the collision efficiency of CO_2 in $H+O_2+M=HO_2+M$, using a syngas mixture with a high CO content (>50% in vol.) under stoichiometric conditions in a ST setup. As a result, they recommended new CO_2 collision efficiency values and presented an uncertainty analysis. The new model led to a substantially improved prediction of τ at 5 atm and was validated with other syngas data reported in the literature. However, the issue under study still remains, as this approach involves fitting a specific type of fuel, and performance with other fuel mixtures needs to be tested.

The main issue with these approaches is that kinetic parameter modifications cannot be generalized for different fuel mixture compositions, thereby limiting their application (Amador Díaz, 2017). Regarding the development of DRMs focused on moderate and low temperatures, two of the most recent (Baigmohammadi *et al*., 2020; Pan and Wallace, 2019) still exhibit significant disagreements when the temperature is 1100 K or lower and the mixture contains H_2. This suggests that the differences between experimental and numerical results are not only due to kinetic effects, but also to physical effects that zero-dimensional models are not able to capture.

Inhomogeneity effects

A wide range of numerical and experimental studies have been conducted to analyze the autoignition phenomenon at moderate and low temperatures in comparison with high temperature conditions. Some characteristics and behaviors related to autoignition inhomogeneity have been identified and categorized as weak or mixed autoignition regimes. In the mixed case, a stratified ignition first occurs, and then a homogeneous behavior appears. In this section, some of the most relevant works are presented along with their findings and methodologies. Additionally, some approaches to the prediction of this phenomenon are analyzed.

In a numerical study, Basal and Im (2011) evaluated the effects of temperature and composition inhomogeneity using three possible scenarios: homogeneous concentration with inhomogeneous temperature, uncorrelated inhomogeneities in both fields, and correlated inhomogeneities. They performed DNS simulations in one and three dimensions, using heat release rate as an indicator for the autoignition regime. They found that, when inhomogeneities are present, the τ value is lower than in the homogeneous case, indicating an acceleration effect in the autoignition process. This reveals the relationship between inhomogeneities and the disagreements obtained with the zero-dimensional models.

In further work, Im *et al*. (2015) established criteria to predict the autoignition regime using the Sankaran number while only considering temperature inhomogeneities. Additionally, they defined the mixing Damköhler number from τ and the mixing time scale associated with the Taylor microscale for the temperature field as a complementary criterion, with the purpose of identifying weak and mixed regimes. The criteria were evaluated using the experimental values reported by Mansfield and Wooldridge (2014). When performing the measurements, image capture was used to identify the autoignition regimes. H_2/CO under lean conditions was used as the reactive mixture. For all analyzed cases, according to the experimental values, the criterion to predict homogeneous and inhomogeneous autoignition

exhibited an adequate performance. However, the criterion to categorize inhomogeneous autoignition as weak or mixed showed some disagreements. These may be associated with the turbulence quantities, as they were arbitrarily assumed. Although the criteria are important and valuable in predicting the autoignition regime, they do not allow establishing and direct relationship between τ predictions and the way to use ignition regime data to improve the process, especially when using zero-dimensional reactors.

Due to the evident differences and the gap in the τ prediction capabilities of the models under moderate and low temperature conditions, some researchers have recently focused on the use of molecular dynamics to more profoundly understand the autoignition phenomenon, aiming to clarify the origin of the studied disagreements.

Sirmas and Radulescu (2017) carried out a 2D simulation with the objective of analyzing the thermal fluctuations during ignition through the molecular dynamics approach. For the simulation, only one generic reaction was considered, with two reactants A and B and one product C, all of them with the same weight and size. The triggering of the reaction was determined using a limit velocity value during the collisions of A and B. This study defined τ as the time during which the limiting reactant was consumed. The effects of activation energy and heat release were analyzed. According to the results, differences of up to 40% may appear in τ estimations when the activation energy and the heat release are high. The τ obtained via molecular dynamics is lower than that calculated through the equilibrium approach. Subsequently, the authors performed a new study considering 3D simulations (Murugesan *et al*., 2019). As in the 2D case, they found that it is possible to obtain different temperature profiles for the same simulation conditions when considering random particle paths. This behavior allows stating that the ignition phenomenon has a stochastic effect. Similarly to the 2D results, when the activation energy was high, the τ values obtained via molecular dynamics were lower than those calculated using conventional kinetic gases theory. Finally, the authors made an analogy between some H_2 oxidation reactions and their molecular dynamic results, suggesting that the observed behaviors correspond to this type of reaction. However, the comparison was only qualitatively made. These works' results can help explain the disagreement observed in other approaches, as the calculations were performed with a generic reactant and therefore are not focused on a specific fuel. Nevertheless, as stated by the authors, additional studies are required to confirm and broaden the observed trends.

On the other hand, in a recent study, Luong *et al*. (2020) extended the criterion proposed by Im *et al*. (2015) for defining the ignition regime, considering fuels with a negative temperature coefficient (NTC). Additionally, for the criterion, they considered concentration inhomogeneity and its negative correlation with temperature due to fuel vaporization. Thus, a new way to calculate the Sankaran number was defined via the following equations:

$$S_{ap} = KD_{al}^{-\frac{1}{2}} \tag{6}$$

$$K = \beta \frac{1}{\left(\tau_f \tau\right)^{0.5}} \left(\left| \frac{\partial \tau}{\partial T} T' \right| + s \left| \frac{\partial \tau_{ig}}{\partial \phi} \phi' \right| \right) \tag{7}$$

For a S_{ap} greater than one, the autoignition is classified as mixed/weak, and the regime is strong for lower values. The results reveal that the modified criterion performs adequately for fuels with a NTC and allows establishing the ignition regime. However, the behavior in the NTC region is different because the temperature sensibility of τ changes considerably, *i.e.*, in the NTC region, an inhomogeneous ignition can be observed at 900 K, but, at 770 and 1045 K (outside said region), a homogeneous ignition takes place. In the NTC region, τ increases when the temperature grows, which reduces the ignition wave in the mixture, whereas the flame velocity increases, entailing a faster growth of the deflagrative flame fronts, which have more time to develop due to the higher τ.

Finally, the compilation of 1398 τ experimental data reported in the literature (Beerer and McDonell, 2008; de Vries and Petersen, 2007; Huang *et al*., 2006; Kalitan *et al*., 2007; Kéromnès *et al*., 2013; Petersen, Hall, *et al*., 2007; Petersen, Kalitan, *et al*., 2007; Shao *et al*., 2019; Vallabhuni *et al*., 2018; Walton *et al*., 2007; Zhang *et al*., 2012) and comparisons with zero-dimensional prediction models using five DRMs, *i.e.*, NUI Galway–C5_2008 (Bourque *et al*., 2009), GRI-Mech 3.0 (Smith *et al*., 2000), Aramco 1.3 (Metcalfe *et al*., 2013), 56.54 Mech (Burke *et al*., 2015), and NUI Galway–C5_2010 (Donato *et al*., 2010; Healy, Donato, *et al*., 2010a, 2010b; Healy, Kalitan, *et al*., 2010; Healy, M. Kopp, *et al*., 2010) were performed in this work in order to evaluate the effect of H_2 on prediction capabilities under moderate and low temperature conditions, wherein the effect of inhomogeneity is relevant. The data were classified into two groups: fuels with H_2 (WH) and without it (WOH). We found relative differences greater than 50% between the experimental and the numerical results. The actual overestimation actual was around 52% of the WH data group. In contrast, this disagreement level only appeared in 15% of the WOH data group. The deviations for both groups are shown in Figure 11. Disagreements higher than one order of magnitude were found for both groups. This behavior suggests that the physical and kinetic-chemical properties of H_2 have a remarkable effect on τ prediction capabilities at moderate and low temperatures (<1000 K). The greater disagreements are probably related to the high burning velocity of H_2, which promotes the reaction fronts that appear in weak/mixed ignition regimes, accelerating the phenomenon and decreasing the experimental value of τ, given that zero-dimensional models cannot capture this phenomenon.

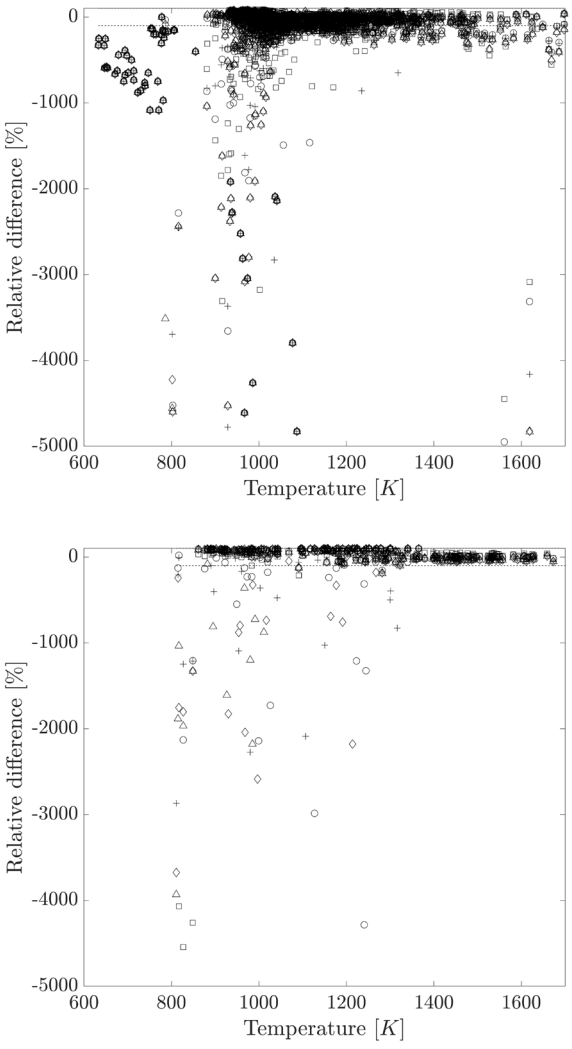


Figure 11. Relative differences between experimental and numerical τ values. WH group (up). WHO group (down). NUI Galway–C5_2008 (circles). GRI-Mech 3.0 (squares). Aramco 1.3 (diamonds). 56.54 Mech (triangles). NUI Galway–C5_2010. (cross).
Source: Authors

Autoignition regime and knocking

Abnormal combustion or knocking is one of the most relevant issues that can take place in ICEs, and it is a limitation in the use of some RGFs, especially those with high H_2 concentrations. The fundamental phenomenon related to knocking is the autoignition of the unburned mixture before it is consumed by the flame front coming from the spark plug. Therefore, a strong correlation between τ and the occurrence of knocking is expected. As presented in the previous sections, the value if τ can be affected by the autoignition regime, and, therefore, it can also influence knocking behavior. This section presents a brief analysis of the relationship between knocking and autoignition regimes, using characteristic numbers for the phenomena.

The methane number (MN) is a characteristic parameter to estimate the knocking tendency of gaseous fuel; a lower MN means a higher tendency towards detonation (Malenshek and Olsen, 2009). The MNs of different fuels, as measured by Gómez *et al*. (2016), were used for the analysis. The fuel compositions cover mixtures of CH_4, C_3H_8, H_2, and CO_2. The specific values and the designation for each mixture are shown in Table 1.

Table 1. Chemical composition of the analyzed gaseous fuel mixtures

Mixture designation	MN	Chemical composition			
		%CH4	%CO2	%C3H8	%H2
100P	36.5	0	0	100	0
50B50M	120	80	20	0	0
54B36M10H	96.5	68	22	0	10
100B	140	60	40	0	0
100M	100	100	0	0	0
83B17P	65.8	50	33	17	0
79B16P5H	65.2	47	32	16	5
75B15P10H	63.8	45	30	15	10
57B38M5H	105.3	72	23	0	5
54.5B36.5M9H	100	69.2	21.8	0	9
90B10P	100	54	36	10	0
88B12H	100	53	35	0	12

Source: Gómez Montoya *et al*. (2016)

The corresponding S_{ap} and Reynolds flame numbers (R_{ef}) were calculated for each mixture. Typical ICE thermodynamic conditions were selected, namely a pressure of 3000 kPa and temperatures of 800, 900, and 1000 K. Figure 12 shows the S_{ap} for the different MN values at the different temperatures. An increase in the S_{ap} as the MN grows is the general trend. However, this behavior depends on the temperature. For the highest temperature, all S_{ap} values are lower than one, which means that, if autoignition occurs, it will be strong (homogenous). In addition, the data are less dispersed in comparison with the other temperature values.

For 800 and 900 K, almost all the S_{ap} are higher than one, suggesting mixed or weak autoignition. As previously mentioned, τ prediction is less accurate in this autoignition regime. Therefore, the time available for the unburned mixture to be consumed by the flame front coming from the spark plug can be overestimated if it is calculated with conventional zero-dimensional models. This means that the occurrence of knocking can probably be omitted in ICE modeling.

On the other hand, it is evident that, for the same MN value, the S_{ap} can vary under the same thermodynamic conditions. For example, four different S_{ap} values were obtained with an MN of 100 at 900 K (0.83, 1.5, 1.7, and 2.1). According to that, it is possible for the same knocking tendency to entail different autoignition behaviors, which can be associated with differences in the combustion properties of the fuel

mixtures. An example of this corresponds to the variations in the R_{ef}, which are related to the laminar burning velocity of the fuel mixture (Figure 13). Nevertheless, a deeper dive is necessary to relate autoignition regimes and the knocking tendency of RGFs.

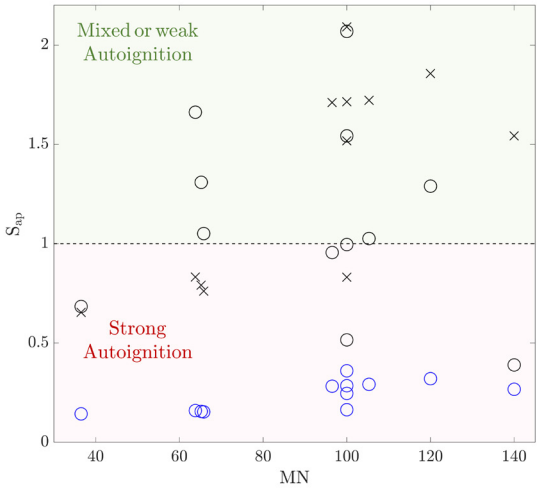


Figure 12. S_{ap} for different MN at 3000 kPa. Black circles: 800 K. Cross: 900 K. Blue circles: 1000 K.
Source: Authors

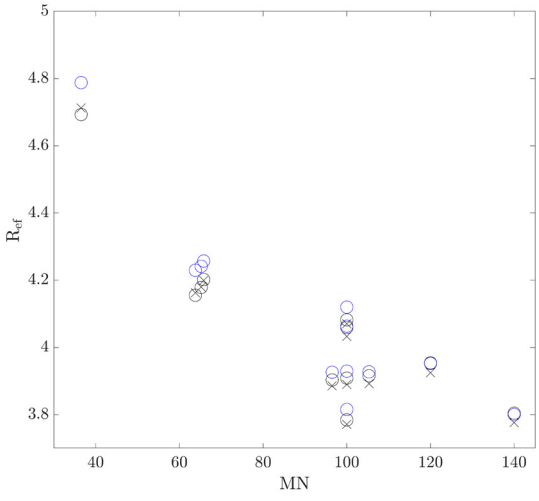


Figure 13. R_{ef} for different MN at 3000 kPa. Black circles: 800 K. Cross: 900 K. Blue circles: 1000 K.
Source: Authors

Conclusions

RGFs are a relevant alternative energy source that contributes to decreasing pollutant emissions. However, due to the high variability of their chemical composition, their fundamental combustion properties (*e.g.*, τ) have not been entirely characterized. Thus, preventing undesired autoignition phenomena becomes a complex issue, limiting the use of this kind of fuel in practical systems (engines, furnaces, and turbines). In order to broaden the use of RGFs in current thermal devices, as well as the development of new ones, it is necessary to adequately predict τ values, particularly at moderate and low temperatures, since many of the low-emissions technologies operate in this range. In order to contribute to this perspective, a literature review encompassing the main approaches and issues in τ prediction when using RGFs was performed. From said review, the following conclusions can be drawn.

Current DRMs and zero-dimensional models exhibit issues in adequately predicting τ for RGFs under moderate and low temperature conditions. The disagreement between experimental and numerical values increases when H_2 is present in the fuel mixture. This, in comparison with the exclusive use of hydrocarbon fuels. The development of specific DRMs for RGFs can improve τ prediction, taking into account the changes in kinetic pathways that the DRMs for conventional fuels cannot capture.

There is no general way to improve τ predictions using a zero-dimensional model. Furthermore, many improvement approaches require previous experimental measurements, which is not possible during the design and evaluation process. Additional studies on the autoignition phenomenon at low and intermediate temperatures, combining thermodynamic and kinetic approaches, are required to establish a more general method that allows improving τ predictions.

Thermal and composition inhomogeneities contribute to weak/mixed autoignition regimes and may be one of the reasons for the disagreement in τ predictions. There is a significant gap in the phenomenological approach to modeling inhomogeneous autoignition regimes. It is necessary to consider the flame fronts prior to volumetric autoignition in order to improve τ predictions. It is necessary to conduct research focused on autoignition under inhomogeneous temperature and concentration conditions while using RGFs in moderate and low temperature ranges, in order to identify differences with high-temperature autoignition and develop new models to improve τ prediction.

There is a relationship between the MN and the autoignition regime, although the behavior can vary according to thermodynamic conditions and combustion properties. More detailed studies on autoignition during knocking are required in order to explain the effect of mixed/weak regimes on abnormal combustion in ICEs. Furthermore, it is necessary to establish a better connection between the S_{ap} and the MN for ICE modeling using RGFs.

Acknowledgements

The authors gratefully acknowledge the financial support provided by ANID - Chile through the *Doctorado Nacional 2020 - Folio: 21201139* scholarship; the Graduate and Programs Directorate for the *Doctorado USM* scholarship, provided by Universidad Técnica Federico Santa María;

and the Research and Extension Division of Universidad Francisco de Paula Santander Ocaña for funding the project with code 158-19-004.

CRedit author statement

Hernando A. Yepes: conceptualization, formal analysis, funds acquisition, investigation, methodology, writing (original draft), visualization and supervision. *Adalberto Salazar:* writing (original draft) and visualization. *José D. Yepes:* visualization.

Conflicts of interest

The authors declare no conflict of interest.

References

Amador Díaz, G. J. (2017). *Auto-ignition modeling in a spark ignition internal combustion engine fueled with gaseous fuels with variable methane number* [Doctoral thesis, Universidad del Norte. https://manglar.uninorte.edu.co/bitstream/handle/10584/7860/11.pdf?sequence=1&isAllowed=y

Arunachalam, A., and Olsen, D. B. (2012). Experimental evaluation of knock characteristics of producer gas. *Biomass and Bioenergy*, *37*, 169-176. https://doi.org/10.1016/j.biombioe.2011.12.016

Baigmohammadi, M., Patel, V., Nagaraja, S., Ramalingam, A., Martinez, S., Panigrahy, S., Mohamed, A. A. E. S., Somers, K. P., Burke, U., Heufer, K. A., Pekalski, A., and Curran, H. J. (2020). Comprehensive experimental and simulation study of the ignition delay time characteristics of binary blended methane, ethane, and ethylene over a wide range of temperature, pressure, equivalence ratio, and dilution. *Energy and Fuels*, *34*(7), 8808-8823. https://doi.org/10.1021/acs.energyfuels.0c00960

Bansal, G., and Im, H. G. (2011). Autoignition and front propagation in low temperature combustion engine environments. *Combustion and Flame*, *158*(11), 2105-2112. https://doi.org/10.1016/j.combustflame.2011.03.019

Barak, S., Pryor, O., López, J., Ninnemann, E., Vasu, S., and Koroglu, B. (2017). High-speed imaging and measurements of ignition delay times in oxy-syngas mixtures with high CO_2 dilution in a shock tube. *Journal of Engineering for Gas Turbines and Power*, *139*(12), 1-7. https://doi.org/10.1115/1.4037458

Beerer, D. J., and McDonell, V. G. (2008). Autoignition of hydrogen and air inside a continuous flow reactor with application to lean premixed combustion. *Journal of Engineering for Gas Turbines and Power*, *130*(5), 051507. https://doi.org/10.1115/1.2939007

Bourque, G., Healy, D., Curran, H., Zinner, C., Kalitan, D., de Vries, J., Aul, C., and Petersen, E. (2009). Ignition and flame speed kinetics of two natural gas blends with high levels of heavier hydrocarbons. *Journal of Engineering for Gas Turbines and Power*, *132*(2), 21504. https://doi.org/10.1115/1.3124665

Burdack, A., Duarte-Herrera, L., López-Jiménez, G., Polklas, T., and Vasco-Echeverri, O. (2023). Techno-economic calculation of green hydrogen production and export from Colombia. *International Journal of Hydrogen Energy*, *48*(5), 1685-1700. https://doi.org/10.1016/J.IJHYDENE.2022.10.064

Burke, U., Somers, K. P., O'Toole, P., Zinner, C. M., Marquet, N., Bourque, G., Petersen, E. L., Metcalfe, W. K., Serinyel, Z., and Curran, H. J. (2015). An ignition delay and kinetic modeling study of methane, dimethyl ether, and their mixtures at high pressures. *Combustion and Flame*, *162*(2), 315-330. https://doi.org/https://doi.org/10.1016/j.combustflame.2014.08.014

Cavaliere, D. E. E., De Joannon, M., Sabia, P., Sirignano, M., and D'Anna, A. (2010). A comprehensive kinetic modeling of ignition of syngas-air mixtures at low temperatures and high pressures. *Combustion Science and Technology*, *182*(7), 692-701. https://doi.org/10.1080/00102200903466525

Chen, C., Pinar, M., and Stengos, T. (2020). Renewable energy consumption and economic growth nexus: Evidence from a threshold model. *Energy Policy*, *139*, 111295. https://doi.org/https://doi.org/10.1016/j.enpol.2020.111295

Cheshmehzangi, A. (2020). COVID-19 and household energy implications: What are the main impacts on energy use? *Heliyon*, *6*(10), e05202. https://doi.org/https://doi.org/10.1016/j.heliyon.2020.e05202

Chintala, V., and Subramanian, K. A. (2017). A comprehensive review on utilization of hydrogen in a compression ignition engine under dual fuel mode. *Renewable and Sustainable Energy Reviews*, *70*, 472-491. https://doi.org/10.1016/J.RSER.2016.11.247

Davidson, D. F., Shao, J. K., Choudhary, R., Mehl, M., Obrecht, N., and Hanson, R. K. (2019). Ignition delay time measurements and modeling for gasoline at very high pressures. *Proceedings of the Combustion Institute*, *37*(4), 4885-4892. https://doi.org/10.1016/j.proci.2018.08.032

de Vries, J., and Petersen, E. L. L. (2007). Autoignition of methane-based fuel blends under gas turbine conditions. *Proceedings of the Combustion Institute*, *31*(2), 3163-3171. https://doi.org/10.1016/j.proci.2006.07.206

Dimitriou, P., and Javaid, R. (2020). A review of ammonia as a compression ignition engine fuel. *International Journal of Hydrogen Energy*, *45*(11), 7098-7118. https://doi.org/10.1016/J.IJHYDENE.2019.12.209

Donato, N., Aul, C., Petersen, E., Zinner, C., Curran, H., and Bourque, G. (2010). Ignition and oxidation of 50/50 butane isomer blends. *Journal of Engineering for Gas Turbines and Power*, *132*(5), 051502. https://doi.org/10.1115/1.3204654

Donohoe, N., Heufer, A., Metcalfe, W. K., Curran, H. J., Davis, M. L., Mathieu, O., Plichta, D., Morones, A., Petersen, E. L., and Güthe, F. (2014). Ignition delay times, laminar flame speeds, and mechanism validation for natural gas/hydrogen blends at elevated pressures. *Combustion and Flame*, *161*(6), 1432-1443. https://doi.org/10.1016/J.COMBUSTFLAME.2013.12.005

Elsemary, I. M. M., Attia, A. A. A., Elnagar, K. H., and Elaraqy, A. A. M. (2016). Experimental investigation on performance of single cylinder spark ignition engine fueled with hydrogen-gasoline mixture. *Applied Thermal Engineering*, *106*, 850-854. https://doi.org/10.1016/j.applthermaleng.2016.05.177

Goldsborough, S. S., Hochgreb, S., Vanhove, G., Wooldridge, M. S., Curran, H. J., and Sung, C. J. (2017). Advances in rapid compression machine studies of low- and intermediate-temperature autoignition phenomena. *Progress in Energy and Combustion Science*, *63*, 1-78. https://doi.org/10.1016/j.pecs.2017.05.002

Gómez Montoya, J. P., Amell, A. A., and Olsen, D. B. (2016). Prediction and measurement of the critical compression ratio and methane number for blends of biogas with methane, propane and hydrogen. *Fuel*, *186*, 168-175. https://doi.org/10.1016/J.FUEL.2016.08.064

Grangeia, C., Santos, L., Ferreira, D. V., Guimarães, R., de Magalhães Ozorio, L., and Tavares, A. (2023). Energy transition scenarios in the transportation sector in Brazil: Contributions from the electrical mobility. *Energy Policy*, *174*, 113434. https://doi.org/10.1016/J.ENPOL.2023.113434

Healy, D., Donato, N. S., Aul, C. J., Petersen, E. L., Zinner, C. M., Bourque, G., and Curran, H. J. (2010a). Isobutane ignition delay time measurements at high pressure and detailed chemical kinetic simulations. *Combustion and Flame*, *157*(8), 1540-1551. https://doi.org/10.1016/j.combustflame.2010.01.011

Healy, D., Donato, N. S., Aul, C. J., Petersen, E. L., Zinner, C. M., Bourque, G., and Curran, H. J. (2010b). n-Butane: Ignition delay measurements at high pressure and detailed chemical kinetic simulations. *Combustion and Flame*, *157*(8), 1526-1539. https://doi.org/10.1016/j.combustflame.2010.01.016

Healy, D., Kalitan, D. M., Aul, C. J., Petersen, E. L., Bourque, G., and Curran, H. J. (2010). Oxidation of C1−C5 alkane quinternary natural gas mixtures at high pressures. *Energy & Fuels*, *24*(3), 1521-1528. https://doi.org/10.1021/ef9011005

Healy, D., M. Kopp, M., L. Polley, N., L. Petersen, E., Bourque, G., and J. Curran, H. (2010). Methane/n-butane ignition delay measurements at high pressure and detailed chemical kinetic simulations. *Energy & Fuels*, *24*(3), 1617-1627. https://doi.org/10.1021/ef901292j

Hu, E., Pan, L., Gao, Z., Lu, X., Meng, X., and Huang, Z. (2016). Shock tube study on ignition delay of hydrogen and evaluation of various kinetic models. *International Journal of Hydrogen Energy*, *41*(30), 13261-13280. https://doi.org/10.1016/j.ijhydene.2016.05.118

Huang, J., Bushe, W. K., Hill, P. G., and Munshi, S. R. (2006). Experimental and kinetic study of shock initiated ignition in homogeneous methane–hydrogen–air mixtures at engine-relevant conditions. *International Journal of Chemical Kinetics*, *38*(4), 221-233. https://doi.org/https://doi.org/10.1002/kin.20157

Im, H. G., Pal, P., Wooldridge, M. S., and Mansfield, A. B. (2015). A regime diagram for autoignition of homogeneous reactant mixtures with turbulent velocity and temperature fluctuations. *Combustion Science and Technology*, *187*(8), 1263-1275. https://doi.org/10.1080/00102202.2015.1034355

International Energy Agency (2016). *World energy outlook 2016*. US Department of Energy. https://doi.org/10.1787/weo-2016-en

International Energy Agency (2018). *Energy policies beyond iea countries: Chile 2018*. OECD. https://doi.org/10.1787/9789264290242-en

Jiang, Y., del Alamo, G., Gruber, A., Bothien, M. R., Seshadri, K., and Williams, F. A. (2019). A skeletal mechanism for prediction of ignition delay times and laminar premixed flame velocities of hydrogen-methane mixtures under gas turbine conditions. *International Journal of Hydrogen Energy*, *44*(33), 18573-18585. https://doi.org/10.1016/j.ijhydene.2019.05.068

Kalitan, D. M., Mertens, J. D., Crofton, M. W., and Petersen, E. L. (2007). Ignition and oxidation of lean CO/H_2 fuel blends in air. *Journal of Propulsion and Power*, *23*(6), 1291-1301. https://doi.org/10.2514/1.28123

Kéromnès, A., Metcalfe, W. K., Heufer, K. A., Donohoe, N., Das, A. K., Sung, C. J., Herzler, J., Naumann, C., Griebel, P., Mathieu, O., Krejci, M. C., Petersen, E. L., Pitz, W. J., and Curran, H. J. (2013). An experimental and detailed chemical kinetic modeling study of hydrogen and syngas mixture oxidation at elevated pressures. *Combustion and Flame*, *160*(6), 995-1011. https://doi.org/10.1016/J.COMBUSTFLAME.2013.01.001

Kuppa, K., Goldmann, A., and Dinkelacker, F. (2018). Predicting ignition delay times of C1-C3 alkanes/hydrogen blends at gas engine conditions. *Fuel*, *222*, 859-869. https://doi.org/10.1016/J.FUEL.2018.02.064

Kuzemko, C., Bradshaw, M., Bridge, G., Goldthau, A., Jewell, J., Overland, I., Scholten, D., Van de Graaf, T., and Westphal, K. (2020). Covid-19 and the politics of sustainable energy transitions. *Energy Research & Social Science*, *68*, 101685. https://doi.org/https://doi.org/10.1016/j.erss.2020.101685

Lamoureux, N., and Paillard, C.-E. (2003). Natural gas ignition delay times behind reflected shock waves: Application to modelling and safety. *Shock Waves*, *13*(1), 57-68. https://doi.org/10.1007/s00193-003-0188-z

Lamoureux, N., Paillard, C.-E., and Vaslier, V. (2002). Low hydrocarbon mixtures ignition delay times investigation behind reflected shock waves. *Shock Waves*, *11*(4), 309-322. https://doi.org/10.1007/s001930100108

Lee, H. C., Jiang, L. Y., and Mohamad, A. A. (2014). A review on the laminar flame speed and ignition delay time of syngas mixtures. *International Journal of Hydrogen Energy*, *39*(2), 1105-1121. https://doi.org/10.1016/J.IJHYDENE.2013.10.068

Lee, H. C., Mohamad, A. A., and Jiang, L. Y. (2017). A detailed chemical kinetics for the combustion of $H_2/CO/CH_4/CO_2$ fuel mixtures. *Fuel*, *193*, 294-307. https://doi.org/10.1016/J.FUEL.2016.12.062

Li, H., Owens, Z. C., Davidson, D. F., and Hanson, R. K. (2008). A simple reactive gasdynamic model for the computation of gas temperature and species concentrations behind reflected shock waves. *International Journal of Chemical Kinetics*, *40*(4), 189-198. https://doi.org/https://doi.org/10.1002/kin.20305

Li, J., Zhao, Z., Kazakov, A., Chaos, M., Dryer, F. L., and Scire, J. I. (2007). A comprehensive kinetic mechanism for CO, CH_2O, and CH_3OH combustion. *International Journal of Chemical Kinetics*, *39*(3), 109-136. https://doi.org/10.1002/kin.20218

Li, W., Zou, C., and Yao, H. (2023). A shock-tube and modeling study of syngas ignition delay times in rich CO_2 environment at elevated pressures. *Combustion and Flame*, *251*, 112695. https://doi.org/10.1016/J.COMBUSTFLAME.2023.112695

Lieuwen, T., Yang, V., and Yetter, R. (2009). *Synthesis gas combustion: Fundamentals and applications*. CRC Press.

Lifshitz, A., Scheller, K., Burcat, A., and Skinner, G. B. (1971). Shock-tube investigation of ignition in methane-oxygen-argon mixtures. *Combustion and Flame*, *16*(3), 311-321. https://doi.org/10.1016/S0010-2180(71)80102-5

Liu, C., Song, H., Zhang, P., Wang, Z., Wooldridge, M. S., He, X., and Suo, G. (2018). A rapid compression machine study of autoignition, spark-ignition and flame propagation characteristics of H_2/CH_4/CO/air mixtures. *Combustion and Flame*, *188*, 150-161. https://doi.org/10.1016/j.combustflame.2017.09.031

Luong, M. B., Hernández Pérez, F. E., and Im, H. G. (2020). Prediction of ignition modes of NTC-fuel/air mixtures with temperature and concentration fluctuations. *Combustion and Flame*, *213*, 382-393. https://doi.org/10.1016/j.combustflame.2019.12.002

Malenshek, M., and Olsen, D. B. (2009). Methane number testing of alternative gaseous fuels. *Fuel*, *88*(4), 650-656. https://doi.org/10.1016/J.FUEL.2008.08.020

Man, X., Tang, C., Wei, L., Pan, L., and Huang, Z. (2013). Measurements and kinetic study on ignition delay times of propane/hydrogen in argon diluted oxygen. *International Journal of Hydrogen Energy*, *38*(5), 2523-2530. https://doi.org/10.1016/j.ijhydene.2012.12.020

Mansfield, A. B., and Wooldridge, M. S. (2014). High-pressure low-temperature ignition behavior of syngas mixtures. *Combustion and Flame*, *161*(9), 2242-2251. https://doi.org/10.1016/J.COMBUSTFLAME.2014.03.001

McDonell, V., and Dunn-Rankin, D. (2008). Chapter 5 – Lean combustion in gas turbines. In D. Dunn-Rankin (Ed.), *Lean Combustion* (pp. 121-160). Academic Press. https://doi.org/http://dx.doi.org/10.1016/B978-012370619-5.50006-6

Medvedev, S. P., Agafonov, G. L., and Khomik, S. V. (2016). Low-temperature ignition delay for hydrogen–air mixtures in light of a reaction mechanism with quantum correction. *Acta Astronautica*, *126*, 150-153.

Metcalfe, W. K., Burke, S. M., Ahmed, S. S., and Curran, H. J. (2013). A hierarchical and comparative kinetic modeling study of C1 − C2 hydrocarbon and oxygenated fuels. *International Journal of Chemical Kinetics*, *45*(10), 638675. https://doi.org/https://doi.org/10.1002/kin.20802

Murugesan, R., Sirmas, N., and Radulescu, M. I. (2019). Non-equilibrium effects on thermal ignition using hard sphere molecular dynamics. *Combustion and Flame*, *205*, 457-465. https://doi.org/10.1016/j.combustflame.2019.04.037

Pan, K., and Wallace, J. S. (2019). A low temperature natural gas reaction mechanism for compression ignition engine application. *Combustion and Flame*, *202*, 334-346. https://doi.org/10.1016/j.combustflame.2019.01.024

Pang, G. A., Davidson, D. F., and Hanson, R. K. (2009). Experimental study and modeling of shock tube ignition delay times for hydrogen–oxygen–argon mixtures at low temperatures. *Proceedings of the Combustion Institute*, *32*(1), 181-188. https://doi.org/10.1016/J.PROCI.2008.06.014

Petersen, E. L., Hall, J. M., Smith, S. D., de Vries, J., Amadio, A. R., and Crofton, M. W. (2007). Ignition of lean methane-based fuel blends at gas turbine pressures. *Journal of Engineering for Gas Turbines and Power*, *129*(4), 937-944. https://doi.org/10.1115/1.2720543

Petersen, E. L., Kalitan, D. M., Barrett, A. B., Reehal, S. C., Mertens, J. D., Beerer, D. J., Hack, R. L., and McDonell, V. G. (2007). New syngas/air ignition data at lower temperature and elevated pressure and comparison to current kinetics models. *Combustion and Flame*, *149*(1-2), 244-247. https://doi.org/10.1016/j.combustflame.2006.12.007

Prince, J. C., and Williams, F. A. (2012). Short chemical-kinetic mechanisms for low-temperature ignition of propane and ethane. *Combustion and Flame*, *159*(7), 2336-2344. https://doi.org/10.1016/J.COMBUSTFLAME.2012.02.012

Rönn, K., Swarts, A., Kalaskar, V., Alger, T., Tripathi, R., Keskiväli, J., Kaario, O., Santasalo-Aarnio, A., Reitz, R., and Larmi, M. (2023). Low-speed pre-ignition and super-knock in boosted spark-ignition engines: A review. *Progress in Energy and Combustion Science*, *95*, 101064. https://doi.org/10.1016/J.PECS.2022.101064

Shao, J., Choudhary, R., Susa, A., Davidson, D. F., and Hanson, R. K. (2019). Shock tube study of the rate constants for H + O_2 + M → HO_2 + M (M = Ar, H_2O, CO_2, N_2) at elevated pressures. *Proceedings of the Combustion Institute*, *37*(1), 145-152. https://doi.org/10.1016/j.proci.2018.05.077

Simsek, Y., Lorca, Á., Urmee, T., Bahri, P. A., and Escobar, R. (2019). Review and assessment of energy policy developments in Chile. *Energy Policy*, *127*, 87-101. https://doi.org/10.1016/J.ENPOL.2018.11.058

Sirmas, N., and Radulescu, M. I. (2017). Thermal ignition revisited with two-dimensional molecular dynamics: Role of fluctuations in activated collisions. *Combustion and Flame*, *177*, 79-88. https://doi.org/10.1016/j.combustflame.2016.12.010

Smith, G. P., Golden, D. M., Frenklach, M., Moriarty, N. W., Eiteneer, B., Goldenberg, M., Bowman, C. T., Hanson, R. K., Song, S., Gardiner, W. C., Jr., V. V. L., and Qin, Z. (2000). *GRI-Mech 3.0*. http://combustion.berkeley.edu/gri-mech/

Sung, C.-J., and Curran, H. J. (2014). Using rapid compression machines for chemical kinetics studies. *Progress in Energy and Combustion Science*, *44*(0), 1-18. https://doi.org/10.1016/j.pecs.2014.04.001

Szwaja, S., Kovacs, V. B., Bereczky, A., and Penninger, A. (2013). Sewage sludge producer gas enriched with methane as a fuel to a spark ignited engine. *Fuel Processing Technology*, *110*, 160-166. https://doi.org/10.1016/j.fuproc.2012.12.008

Tang, C., Man, X., Wei, L., Pan, L., and Huang, Z. (2013). Further study on the ignition delay times of propane–hydrogen–oxygen–argon mixtures: Effect of equivalence ratio. *Combustion and Flame*, *160*(11), 2283–290. https://doi.org/10.1016/j.combustflame.2013.05.012

Thi, L. D., Zhang, Y., and Huang, Z. (2014). Shock tube study on ignition delay of multi-component syngas mixtures – Effect of equivalence ratio. *International Journal of Hydrogen Energy*, *39*(11), 6034-6043. https://doi.org/10.1016/j.ijhydene.2014.01.170

US Energy Information Administration (2016). *International energy outlook 2016*. US Department of Energy. https://www.eia.gov/outlooks/ieo/pdf/0484(2016).pdf

Vallabhuni, S. K., Lele, A. D., Patel, V., Lucassen, A., Moshammer, K., AlAbbad, M., Farooq, A., and Fernandes, R. X. (2018). Autoignition studies of liquefied natural gas (LNG) in a shock tube and a rapid compression machine. *Fuel*, *232*, 423-430. https://doi.org/10.1016/J.FUEL.2018.04.168

Walton, S. M., He, X., Zigler, B. T., and Wooldridge, M. S. (2007). An experimental investigation of the ignition properties of hydrogen and carbon monoxide mixtures for syngas turbine applications. *Proceedings of the Combustion Institute*, *31*(2), 3147-3154. https://doi.org/10.1016/j.proci.2006.08.059

Wang, B. L., Olivier, H., and Grönig, H. (2003). Ignition of shock-heated H_2-air-steam mixtures. *Combustion and Flame*, *133*(1-2), 93-106. https://doi.org/10.1016/S0010-2180(02)00552-7

Wang, B., Yang, Z., Xuan, J., and Jiao, K. (2020). Crises and opportunities in terms of energy and AI technologies during the COVID-19 pandemic. *Energy and AI*, *1*, 100013. https://doi.org/https://doi.org/10.1016/j.egyai.2020.100013

Wang, Q., and Wang, S. (2020). Preventing carbon emission retaliatory rebound post-COVID-19 requires expanding free trade and improving energy efficiency. *Science of The Total Environment*, *746*, 141158. https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.141158

Wang, S., Davidson, D. F., and Hanson, R. K. (2019). Shock tube techniques for kinetic target data to improve reaction models. In T. Favarelli, F. Manenti, and E. Ranzi (Eds.), *Computer Aided Chemical Engineering* (vol. 45, pp. 169-202). Elsevier.

Wang, Z., Liu, H., and Reitz, R. D. (2017). Knocking combustion in spark-ignition engines. *Progress in Energy and Combustion Science*, *61*, 78–-12. https://doi.org/10.1016/j.pecs.2017.03.004

Warnatz, J., Maas, U., and Dibble, R. W. (1996). *Combustion*. Springer. https://doi.org/10.1007/978-3-642-97668-1

Wolde-Rufael, Y., and Mulat-Weldemeskel, E. (2022). The moderating role of environmental tax and renewable energy in CO_2 emissions in Latin America and Caribbean countries: Evidence from method of moments quantile regression. *Environmental Challenges*, *6*, 100412. https://doi.org/10.1016/J.ENVC.2021.100412

Yu, G., Askari, O., and Metghalchi, H. (2018). Theoretical prediction of the effect of blending JP-8 with syngas on the ignition delay time and laminar burning speed. *Journal of Energy Resources Technology, Transactions of the ASME*, *140*(1), 012204. https://doi.org/10.1115/1.4037376

Zádor, J., Taatjes, C. A., and Fernandes, R. X. (2011). Kinetics of elementary reactions in low-temperature autoignition chemistry. *Progress in Energy and Combustion Science*, *37*(4), 371-421. https://doi.org/10.1016/J.PECS.2010.06.006

Zhang, Y., Huang, Z., Wei, L., Zhang, J., and Law, C. K. (2012). Experimental and modeling study on ignition delays of lean mixtures of methane, hydrogen, oxygen, and argon at elevated pressures. *Combustion and Flame*, *159*(3), 918-931. https://doi.org/10.1016/j.combustflame.2011.09.010

Research Article/Generic Engineering

http://doi.org/10.15446/ing.investig.106659

Development and Validation of an Affordable Calibration Method for Surface Plates

Desarrollo y validación de un método de calibración asequible para superficies de referencia

Daniela I. Garzón[1], Jorge L. Galvis[2], David A. Plazas[3], Victor H. Gil[4], and Ovidio Almanza[5]

ABSTRACT

The measurement and calibration of flat surfaces is highly relevant for precision engineering, length metrology, and optical systems. Hence, many National Metrology Institutes (NMIs) tend to offer calibration services in this regard. Typically, mechanical, electromechanical, and optical measuring techniques are applied, with uncertainties in the order of micrometers. However, these techniques necessitate expensive equipment that requires periodical calibration and maintenance, which not many laboratories can afford. This work presents the validation of an affordable and simple calibration technique for surface plates through the evaluation of metrological compatibility with a reference calibration method. A surface plate was calibrated with both methods under the same conditions to validate our proposal. The 2.9 μm uncertainty obtained with the new method demonstrates its reliability and usability for laboratories with surface plates up to 300 mm in length that have a grade AA to B flatness accuracy. Due to its low initial cost, reliability, and ease of implementation, the proposed calibration method can be recommended to all laboratories and industries that need to constantly verify their surface plates.

Keywords: length metrology, surface plates, flatness

RESUMEN

La medición y calibración de superficies planas es de gran relevancia para la ingeniería de precisión, la metrología de longitud y los sistemas ópticos. Por lo tanto, muchos Institutos Nacionales de Metrología (NMIs) tienden a ofrecer servicios de calibración en este aspecto. Típicamente se aplican técnicas de medición mecánica, electromecánica y óptica, con incertidumbres del orden de micrómetros. Sin embargo, estas técnicas requieren de equipos costosos que necesitan calibración y mantenimiento periódicos, lo cual no es asequible para muchos laboratorios. Este trabajo presenta la validación de una técnica de calibración asequible y sencilla para placas de superficie mediante la evaluación de la compatibilidad metrológica con un método de calibración de referencia. Una placa de superficie fue calibrada con ambos métodos bajo las mismas condiciones para validar nuestra propuesta. La incertidumbre de 2.9 μm obtenida con el nuevo método demuestra su fiabilidad y usabilidad para laboratorios con placas de superficie de hasta 300 mm de longitud que tienen una precisión de planitud de grado AA a B. Debido a su bajo costo inicial, fiabilidad y facilidad de implementación, se puede recomendar el método de calibración propuesto a todos los laboratorios e industrias que necesiten verificar constantemente sus placas de superficie.

Palabras clave: metrología de longitud, superficies de referencia, planitud

Received: January 10th 2023
Accepted: April 24th 2024

Introduction

A surface plate can regarded as the initial point for most calibrations using precise dimensional and geometrical measurement instruments. Most laboratories, workshops, and industries have granite, steel, or ceramic surface plates, making the calibration of these surfaces an important task for dimensional metrology. The periodic calibration of surface plates typically occurs in intervals from six months to a year, depending on the grade, resistance, conditions, and frequency of use. Since surface plates have become a primary tool for dimensional laboratories, many methods have been developed to measure their flatness and straightness, such as the two-point connecting method (given i=K., n.d.), the Moody method (Moody, 1955), multi-point methods (Lakota and Gorog, 2011; Mikó, 2021), and scanning methods (Lakota and Gorog, 2011; Ju HUO, 2018; Ehret *et al.*, 2011; Schulz, Ehret, and Křen, 2013). Measurement can be performed with different devices, such as laser interferometers (Azaryan *et al.*, 2017; Ehret, Reinsch, and Schulz, 2019), electronic levels (Yellowhair and Burge, 2008; Glubokov, Glubokova, Afonina, Zelensky, and Semenishchev, 2022), spirit levels (Yang, Wang, and Zhu, 2021), autocollimators (Geckeler *et al.*, 2022; Heikkinen, Byman, Palosuo, Hemming, and Lassila, 2017), and coordinate measuring machines (CMMs) (Ali and Buajarern, 2013). Many of these works have achieved different levels of accuracy due to their main measurement principle, as is the case of the methods that use optical techniques such as interferometry or CMMs.

[1]BSc Physics, Unisersidad Nacional de Colombia, Colombia. E-mail: digarzonp@unal.edu.co
[2]Physics engineer, Universidad Nacional de Colombia, Colombia. MSc Physics, Universidad Nacional de Colombia, Colombia. Affiliation: Specialized professional, Instituto Nacional de Metrología, Colombia. E-mail:jlgalvis@inm.gov.co
[3]Mechanical engineer, Universidad Antonio Nariño, Colombia. Affiliation: Bachelor professional, Instituto Nacional de Metrología, Colombia. E-mail: daplazas@inm.gov.co
[4]Mechanical engineer, Fundación Universidad de America, Colombia. Affiliation: Specialized professional, Instituto Nacional de Metrología, Colombia. E-mail: vgil@inm.gov.co
[5]PhD Physics, Universidad de Valladolid, Spain. MSc Physics, Universidad Nacional de Colombia, Colombia. Affiliation: Associate professor, Universidad Nacional de Colombia, Colombia. E-mail: oaalmanzam@unal.edu.co

However, these techniques have many disadvantages, *e.g.*, the time required for a full surface plate calibration due to the alignment of the lasers or the uncertainty of the CMMs associated with the mechanical positioning system. In light of the above, the Moody method, coupled with the use of electronic levels, is the most commonly used calibration technique for surface plates. When an electronic level is employed for surface plate calibration, the measuring device is moved over a grid that is traced over the surface. Straightness is measured for each grid line, and the results are combined into a plane to represent the flatness deviations of the surface. However, interferometers, CMMs, and electronic levels are expensive for many factories and laboratories to obtain and maintain in the long term, causing them to look for onsite calibrations provided by NMIs. This work presents and validates the roto-translation method for flatness calibration on surface plates. The traditional Moody method using electronic levels was implemented to validate the results. The accuracy of the electronic levels and the similarity between the methods allowed for a direct comparison (Meijer and Heuvelman, 1990; Gusel, Acko, and Sostar, 2000a, 2000b). Thus, this work proposes an affordable method for the calibration of surface plates.

Calibration method

Many flatness calibration methods use the eight-line grid pattern proposed by Moody in 1955 (Moody, 1955; Drescher, 2003). All readings from the grid-line pattern form a mathematical surface to describe the flatness deviations of the points selected in a surface plate of interest. However, in the measuring process, each grid line has a different height as a consequence of variations in the initial point of each grid line used as reference to measure the deviations of the remaining points. Therefore, each grid line has to be mathematically treated to form a surface that describes the flatness deviations of the plate. This surface is delimited by three parallel planes: the first is called *roof plane* and is located at the same height as the highest deviation of the surface plate; the second is the *base plane* and is located at the same height as the lowest deviations of the surface plate; and the third is called *datum plane* and is located between the roof and the base planes (Figure 1). Usually, the calibration methods derived from Moody use the two diagonals of the grid line to form a reference plane for the rest of the lines. This plane is formed by the four ends and the intersection point of the two diagonals in the center of the surface plate.

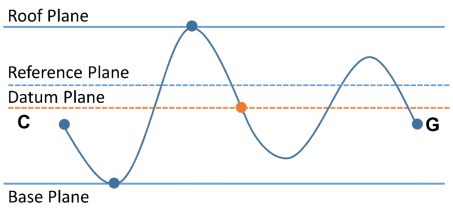


Figure 1. Identification of the principal planes for flatness calibration **Source:**Authors

The calibration method presented in this work is called *roto-translation* and is derived from the Moody method.

Roto-translation method

The roto-translation method divides the surface plate into an eight-line grid. The measurement points must be equally spaced along the lines, with a minimum of five points for each one. The grid includes four perimeter lines, two diagonals and two central ones (Figure 2).

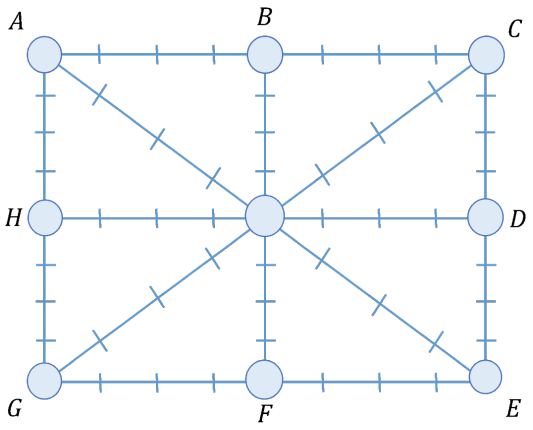


Figure 2. Grid line diagram **Source:** Authors

The calibration setup for this method uses a pin gauge set from 4.9 to 5.1 mm with increments of 0.001 mm per pin, two 5 mm gauge blocks to generate a GAP between the surface plate, and a steel precision straight edge, as shown in Figure 3.

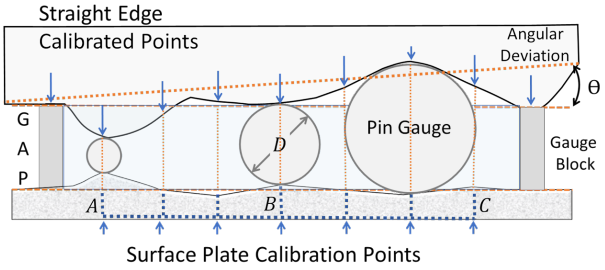


Figure 3. Roto-translation calibration setup **Source:** Authors

Figure 3 shows the measurement setup for surface plate calibration. This setup requires placing the two-gauge block outside the calibration line at the same distance from its limits, placing the straight edge over the gauge blocks and aligning each point of the edge with the calibration points traced on the surface plate. As the gap between the surface and the straight edge is expected to be 5 mm (*i.e.*, the length of the gauge blocks), any deviation from this value will be caused by the straight edge or the surface. As shown in Figure 3, any deviations in the straight edge affect the results of the calibration, causing two important issues: i) the angular deviation caused by the points used to support the edge over the gauge blocks and ii) the effect of the deviations of the straight edge on the readings taken with the gauge pins along the calibration path. In order to solve these issues, the straight edge-calibrated points have to be properly numbered and marked, thereby enabling proper corrections to measurement results. Finally, the accuracy of any measurement will depend on the stability of the setup and the care taken during calibration, as any misalignment

concerning the straight edge and the surface plate will introduce errors into the readings.

Methodology. The general procedure for this surface plate calibration technique is derived from Moody's method, but, instead of taking measurements by steps or stations, the straightness deviation of each grid line is taken directly from each measurement. This procedure is presented below.

1. Align the points of the straight edge with the calibration points of the surface and establish a direction of calibration (*e.g.*, from A to E). Then, center a 5 mm pin between the first point of the straight edge and the point of the surface plate, in order to measure the height of the gap G. If the pin does not fit, smaller pins have to be placed in a sequential manner until one fits in the gap (Figure 3). Otherwise, if the 5 mm pin is smaller than the gap, larger pins have to be placed until one fits. It is important to note that the smaller the increments or decrements of the pins, the more precise and accurate the measurements will be.

2. Store the readings, specifying the number of the straight edge point used for measuring each surface plate point, and classify them based on the two diagonals, the four-perimeter lines, and the two central lines. These readings are the combination between the deviation of the surface plate, the deviations of the straight edge, and the possible tilt of the straight edge caused by the height difference between the supporting points. For the sake of simplicity, the stored readings of each grid line are represented as a linear vector. This is shown in Equations (1), (2), and (3).

$$AE = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_N \end{bmatrix} CG = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_N \end{bmatrix} \tag{1}$$

$$AC = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_N \end{bmatrix} CE = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_N \end{bmatrix} GE = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_N \end{bmatrix} AG = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_N \end{bmatrix} \tag{2}$$

$$BF = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_N \end{bmatrix} HD = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_N \end{bmatrix} \tag{3}$$

3. To eliminate the influence of the straight edge on all the stored readings, make a vector SE that represents the deviations of the points along of the straight edge used for measuring each grid line, including the supporting points that are aligned with the gauge blocks. Then, translate the vector to its datum line by subtracting the deviation of the first point of the vector, as presented in Equation (4).

$$SE'_i = x_i - x_0 = x'_i \text{ for i = 0 to N} \tag{4}$$

As a consequence of the translation, the first value of the vector SE' is $x'_0 = 0$. Afterwards, rotate the vector as suggested in Equation (5) to eliminate the tilt caused by placing the straight edge over the gauge blocks.

$$SE''_i = x'_i + \left(x'_0 - x'_N\right)(i/N) = x''_i \text{ for i = 0 to N} \tag{5}$$

Both ends of the rotated vector SE'' will be $x''_0 = x''_N = 0$ (Figure 4).

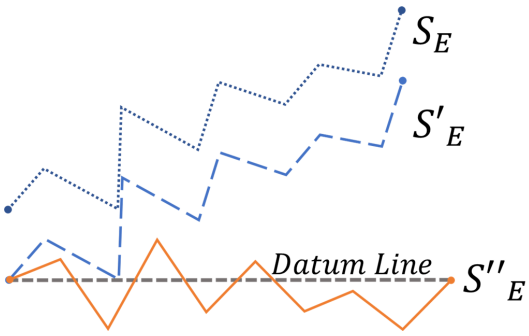


Figure 4. Translation and rotation of the vector of the straight edge **Source:** Authors

As a consequence of the translation and rotation of vector SE, the points of the edge will suffer a horizontal displacement with respect to their original position. In a geometrical analysis of this case, it was found that the maximum permissible horizontal displacement of the straight edge points is $\pm 1\mu m$ with respect to their original position, and that their maximum permissible straightness deviation is $\pm 40\mu m$. If the horizontal displacement of the straight edge points exceeds the $\pm 1\mu m$ interval, it is proven that, due to the alignment with respect to the points of the surface plate, the error in the readings increases. After the rotation, the gap formed by the gauge blocks at the supporting points will be 5 mm.

4. Based on Figure 4, the rotated vector SE'' will only consider the deviations of the straight edge around the datum line. To obtain the real deviation d at each calibration point of the surface plate, the deviations of SE'' have to be added to or subtracted from the diameter D of the pin that fit in the gap G between the surfaces (Figure 5).

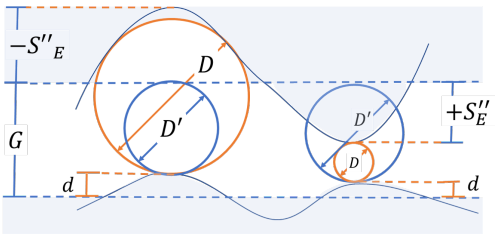


Figure 5. Corrections to the measurements taken with the gauge pins **Source:** Authors

Here, $-S''_E$ and $+S''_E$ are the deviations inward and outward of the edge, respectively, and D' is the real diameter of the pin, calculated as $D' = D + S''_E$ and $D' = D - S''_E$. Thus, the deviation at each point of the surface plate is $d = G - D'$. This correction has to be done for every point of the grid line in Figure 2.

5. Proper measurement of the two diagonals is of great importance since they establish the reference plane on which the rest of the grid lines is laid. After correcting the measurement readings, rotate both diagonals AE and CG until both their ends have the same deviation. To this effect, use Equation (6).

$$\begin{aligned} AE_i^r &= x_i + (x_0 - x_N)\,(i/N) = x^r{}_i \text{ for i = 0 to N} \\ CG_i^r &= x_i + (x_0 - x_N)\,(i/N) = x^r{}_i \text{ for i = 0 to N} \end{aligned} \tag{6}$$

where $x_0^r = x_N^r = \alpha^r$ for CG^r and $x_0^r = x_N^r = \beta^r$ for AE^r. Subsequently, translate the center of both diagonals to the datum plane by subtracting the deviation of the central point of each vector AE^r and CG^r, as shown in Equation (7).

$$\begin{aligned} AE_i'' &= x_i^r - x_{N/2}^r = x''{}_i \text{ for i = 0 to N} \\ CG_i'' &= x_i^r - x_{N/2}^r = x_i'' \text{ for i = 0 to N} \end{aligned} \tag{7}$$

Due to the translations, $x_0'' = x_N'' = \alpha, x_{N/2}'' = 0$ for CG'' and $x_0'' = x_N'' = \beta, x_{N/2}'' = 0$ for AE''. After the rotation and translation of the diagonals, they will form a reference plane for the rest of grid lines, which have to be rotated and translated to coincide with this plane. It is important to make a diagram with the values found for the center and both ends of the diagonals, as shown in Figure 6.

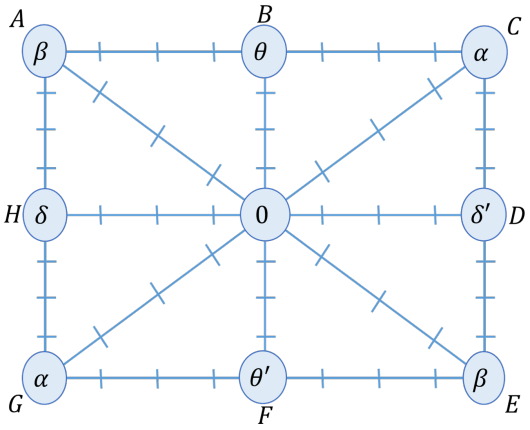


Figure 6. Diagram guide for the rotation and translation of grid lines **Source:** Authors

This diagram provides a guide for the operations that have to be performed in order to make all the grid lines coincide on the datum plane.

6. To make them coincide with the diagonals, translate the perimeter lines to the datum plane by subtracting their first values, as shown in Equation 8.

$$\begin{aligned} AC_i' &= x_i - x_0, \text{ for } i = 0 \text{ to } N \\ CE_i' &= x_i - x_0, \text{ for } i = 0 \text{ to } N \\ GE_i' &= x_i - x_0, \text{ for } i = 0 \text{ to } N \\ AG_i' &= x_i - x_0, \text{ for } i = 0 \text{ to } N \end{aligned} \tag{8}$$

Then, translate and rotate the perimeter lines to make them match at both ends with the diagonals. Equation (9) presents two roto-translation factors, $k_{\alpha\beta}$ and $k_{\beta\alpha}$, for the rotation and translation of each perimeter line that has to begin in α and end in β and of those that have to begin in β and end in α. To translate and rotate each perimeter line, use the roto-translation factor as indicated in Equation (10).

$$\begin{aligned} k_{\alpha\beta} &= \left[\alpha - \left(\beta - x_N'\right)\right]/N \\ k_{\beta\alpha} &= \left[\beta - \left(\alpha - x_N'\right)\right]/N \end{aligned} \tag{9}$$

$$\begin{aligned} AC_i'' &= x_i' + \left[\left(\alpha - x_N'\right) + \left(k_{\beta\alpha} \cdot (N - i)\right)\right] \\ AG_i'' &= x_i' + \left[\left(\alpha - x_N'\right) + \left(k_{\beta\alpha} \cdot (N - i)\right)\right] \\ CE_i'' &= x_i' + \left[\left(\beta - x_N'\right) + \left(k_{\alpha\beta} \cdot (N - i)\right)\right] \\ GE_i'' &= x_i' + \left[\left(\beta - x_N'\right) + \left(k_{\alpha\beta} \cdot (N - i)\right)\right] \end{aligned} \tag{10}$$

It is important to note that the roto-translation factors will be different for each perimeter line due to their dependence on the last value of each line x_N'. These rotations and translations position the perimeter lines onto the datum plane (Figure 6).

7. Subsequently, translate the central lines to the datum plane, as presented in Equation (11).

$$\begin{aligned} BF_i' &= x_i - x_0, \text{ for } i = 0 \text{ to } N \\ HD_i' &= x_i - x_0, \text{ for } i = 0 \text{ to } N \end{aligned} \tag{11}$$

Then, translate and rotate the central lines to make them coincide with the central deviation of the perimeter lines (Figure 6). As in the previous step, use the roto-translation factors $k_{\theta\theta'}$ and $k_{\delta\delta'}$, presented in Equations (12) and (13), to rotate and translate the central lines.

$$\begin{aligned} k_{\theta\theta'} &= \left[\theta - \left(\theta' - x_N'\right)\right]/N \\ k_{\delta\delta'} &= \left[\delta - \left(\delta' - x_N'\right)\right]/N \end{aligned} \tag{12}$$

$$\begin{aligned} BF_i'' &= x_i' + \left[\left(\theta' - x_N'\right) + \left(k_{\theta\theta'}.(N - i)\right)\right] \\ HD_i'' &= x_i' + \left[\left(\delta' - x_N'\right) + \left(k_{\delta\delta'}(N - i)\right)\right] \end{aligned} \tag{13}$$

If everything is done correctly, the middle point of the central lines should be zero. However, Moody established a closure error for the middle point of the central lines, which allows for a maximum deviation of 2.5 μm from 0 (Moody, 1955). When the absolute value of the deviations of the middle points of the central lines is greater than 0 and smaller than 2.5 μm, the central lines BF_i'' and HD_i'' have to be corrected, with the exception of both ends. To correct the central lines, subtract the deviation of the middle point from each central line.

8. Translate all the lines to the reference plane by subtracting the lowest value x_L'' of all grid lines, as shown in Equation (14).

$$\begin{aligned} AE_i^f &= x_i'' - x_L'' = x_i^f \quad CE_i^f = x_i'' - x''_L = x_i^f \\ CG_i^f &= x_i'' - x_L'' = x_i^f \quad GE_i^f = x_i'' - x''_L = x_i^f \\ AC_i^f &= x_i'' - x_L'' = x_i^f \quad BF_i^f = x_i'' - x_L'' = x_i^f \\ AG_i^f &= x_i'' - x_L'' = x_i^f \quad HD_i^f = x_i'' - x_L'' = x_i^f \end{aligned} \tag{14}$$

Finally, all the grid lines will coincide on the reference plane and represent the flatness deviations of the surface plate.

Roto-translation uncertainty

For this method, uncertainty is calculated according to the recommendations of the *Guide to the expression of uncertainty in measurement* (GUM) (Espinosa, Diaz, Baca, Allison, and Shilling, 2008). Equation (15) presents the combined standard uncertainty u_c for the roto-translation method as an estimate of the standard deviation of the possible flatness deviation values. This uncertainty contains all the influence factors that affect the flatness deviations of the surface plate.

$$u_c^2 = u_{res}^2 + u_{rep}^2 + u_{se}^2 + u_{gp}^2 + u_{gb}^2 \tag{15}$$

Here, u_{res} is the uncertainty regarding the resolution of the pin gauges, u_{rep} represents the repeatability of the measurements taken for each grid line, u_{se} denotes the uncertainty of the straight edge, u_{gp} corresponds to the uncertainty of the gauge pins, and u_{gb} is the uncertainty of the gauge blocks.

Method validation

To validate a new method, it is important to determine whether the measurement results fulfill the requirements established for the proposed use. Because of the above, establishing the performance characteristics and limitations related to the measurement process, such as precision, repeatability, limit of detection, and robustness, among others, is the most important task. Measurement methods for flatness calibration have to be capable of ensuring that the results are framed within the tolerances set for surface plates, from grade AA to grade B, and guarantee traceability to the International System of Units (SI) (Drescher, 2003). For the validation process, the roto-translation method was fully developed and optimized; the measurement instruments (the gauge blocks, the gauge pins, and the straight edge) were regularly calibrated, technically controlled, and well maintained (ISO, 2015).

The metrological compatibility model proposed by the VIM3 was employed to validate the roto-translation method (JCGM-BIPM, 2008) because it allows evaluating uncertainty and the precision of each calibrated point. This model calculates the significance of the differences between the results of multiple measurements performed on the same measurand (surface plate). The compatibility test determined whether the results obtained using the roto-translation method could be satisfactorily compared to those of the validated Moody method and could thus be considered acceptable.

To apply the compatibility test, the same surface plate was calibrated using the Moody method via electronic levels. These electronic levels were operated on the pendulum principle, which involves sensing and quantifying the tilt of the level based on the imbalance of the pendulum regarding the reference position. For the flatness calibration of the surface plate, one electronic level remained stationary, while the second one was moved along the grid line presented in Figure 2.

Moody uncertainty

The standard combined uncertainty for the Moody method is presented in Equation 16.

$$u_c^2 = u_{res}^2 + u_{rep}^2 + u_r^2 + u_{el}^2 \tag{16}$$

where u_{res} is the uncertainty regarding the resolution of the electronic levels, u_{rep} denotes the repeatability of the measurements taken for each grid line, u_r corresponds to the uncertainty regarding the rule used to make the grid line diagram on the surface plate, and u_{el} represents the uncertainty of the electronic levels.

Metrological compatibility

NMIs and accredited laboratories report uncertainties with a coverage factor k for the risk of committing an error. Typically, $k = 2$ indicates a confidence of 95% that the measurement results will be within the estimated interval for the reported value (Gromczak *et al.*, 2016). The VIM established that there is metrological compatibility when \the absolute value of the difference of any pair of measured quantity values from two different measurement results is smaller than some chosen multiple of the standard measurement uncertainty of that difference" (Heydorn, 2010). Equation (17) presents the mathematical expression for this definition.

$$|x_i - x_R| \leq \kappa \cdot u_{x_i - x_R} \tag{17}$$

where $|x_i - x_R|$ denotes the absolute error of the method under validation (x_i) with respect to the reference method (x_R), k is the coverage factor and has to be the same for both uncertainties, and $u_{x_i - x_R}$ is the standard uncertainty of the difference between any two quantity values. Equation (18) presents the expression for $u_{x_i - x_R}$.

$$u_{x_i - x_R} = \sqrt{u^2(x_i) + u^2(x_R) - 2r(x_i, x_R)\, u(x_i)\, u(x_R)} \tag{18}$$

where $u(x_i)$ and $u(x_R)$ represent the uncertainty of the under-validation method and the reference method; and the correlation factor between the quantities is denoted as $r(x_i, x_R)$. The calibration methods are independent, making the correlation factor equal to 0. Since the standard expanded uncertainty is $U(x_i) = ku(x_i)$, Equation (17) is usually simplified as presented in Equation (19). This expression is usually known as a *normalized error* (ISO, 2006).

$$E_n = \frac{x_i - x_R}{\sqrt{U^2(x_i) + U^2(x_R)}} \Upsilon \tag{19}$$

A normalized error value between ±1 indicates an acceptable degree of compatibility between the calibration methods. The nearest the value of E_n to 0, the better the compatibility.

Results

To validate the roto-translation method, an A-grade surface plate with a working area of 650×650 mm (ASME-B89.3.7, 2013) was employed as the measurand. For the comparison between calibration methods, we had to ensure that the influence variables affecting the calibration were controlled, so that they did not affect the repeteability and reproducibility of the results. Given the above, the laboratory's environmental conditions were controlled, maintaining a temperature of 20±0.5 °C and a relative humidity of about 50 ±10% Rh (Skattum, 2015).

Moody and electronic levels

The Moody method measurements were taking using two Mahr electronic levels with a foot spacing of L =104.5 mm and expanded uncertainty of U = 0.25 μm. The electronic levels measured the difference between the signals produced by the pendulums to reduce the influence of the vibrations and tilt of the surface plate. Thus, one of the levels acted as a reference and remained stationary at the start of each line of the diagram, and the second was displaced stepwise along the grid lines. The results were stored manually by the calibration operators and were computed in MATLAB, in order to obtain the Moody diagram and its uncertainty. Figure 7 presents all the flatness deviations of the granite surface plate obtained with the Moody method. The advantage of using electronic levels and the Moody method to validate a new measuring method lies in its compatibility and comparability with other, more precise methods, such as those involving auto-colimators and interferometers (Skattum, 2015; Yellowhair and Burge, 2008; Glubokov *et al.*, 2022; Espinosa *et al.*, 2008). Nevertheless, this approach is constrained by the fixed foot spacing of the levels, which limits the measurements to medium and large surface plates (Skattum, 2015; Drescher, 2003).

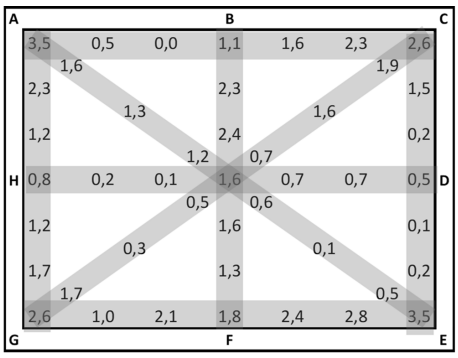


Figure 7. Measurement results obtained with the electronic levels and the Moody method **Source:** Authors

Roto-translation method

To implement the roto-translation method, it is important to properly select the straight edge, since its main purpose is to verify the straightness and flatness of mechanical parts (DIN874-1, 2003). Nevertheless, many of these straight edges are suitable for the verification, measurement, and calibration of surface plates due to their small straight deviations. In light of the above, we used a grade-00 Mahr steel straight edge with a length of 1000 mm, as well as a working edge with a flatness of 7.7 μm and an expanded uncertainty of U=2.7 μm, with a coverage factor k = 2. As a consequence of using different steel measurement instruments (*i.e.*, the gauge pins, the gauge blocks, and the straight edge), it is important to remove the oils used to prevent oxidation, in addition to letting the instruments reach the laboratory temperature on the surface plate. Additionally, during calibration, it is important for the calibrated points of the straight edge to be identified with ruler marks and be properly aligned with the reference calibration points drawn on the surface, in order to prevent measurement errors. Since the surface plate is square, the same calibrated points of the straight edge were used to calibrate the central and perimeter lines, as well as both diagonals (Table 1).

Table 1. Transformations for each vector SE used for measuring the surface plate (μ m)

Grid lines	i	0	1	2	3	4	5	6	7	8
Diagonal lines	SE'	0.0	0.6	3.3	-0.3	-3.2	-0.1	2.5	0.2	0.0
	SE''	0.0	0.6	3.3	-0.3	-3.2	-0.1	2.5	0.2	0.0
Central and perimeter lines	SE'	0.0	0.6	3.8	2.2	-0.8	-3.7	-3.0	-0.1	1.1
	SE''	0.0	0.5	3.5	1.7	-1.4	-4.4	-3.8	-1.1	0.0

Source: Authors

A shown in Table 1, for the diagonal lines, both reference ends of the straight edge were used, thereby avoiding the need to perform any translation or rotation of the vector SE. Following the roto-traslation method, the data for the grid lines, obtained using the gauge pins, was corrected with the vectors SE'' and then rotated and translated (Tables 2,3, and 4). The factors obtained for the rotation and translation of the perimeter lines were $k_{\alpha\beta}$ = 0.1μm and $k_{\alpha\beta}$ -0.3 μm for CE and GE, respectively; and $k_{\beta\alpha}$ =0.1 μm and $k_{\beta\alpha}$ =0.2 μm for AC and AG. For the central lines, BF $k_{\theta\theta'}$ =-0.4 μm and HD $k_{\delta\delta'}$ =0.1 μm were used.

Table 2. Diagonal lines

i	AE	AE^r	AE''	AE^f	CG	CG^r	CG	CG^f
0	0.5	0.5	1.0	3.5	-1.5	-1.5	0.0	2.5
1	-1.3	-1.1	-0.6	1.9	-0.2	-1.9	-0.4	2.2
2	-2.0	-1.7	-1.2	1.4	-1.6	-2.3	-0.8	1.8
3	-1.0	-0.5	0.0	2.5	-0.5	-1.5	0.0	2.5
4	-2.5	-1.8	-1.3	1.2	-1.5	-2.8	-1.3	1.2
5	-3.0	-2.2	-1.7	0.9	0.1	-1.6	-0.1	2.5
6	-0.5	0.5	1.0	3.5	0.5	-1.5	0.0	2.5

Source: Authors

Once the rotations and translations had been performed. We evaluated whether the ends of the diagonals coincided with those of the perimeter lines; whether both ends of the central lines coincided with the central point of the perimeter lines; and whether the center of the central lines coincided with the central point of the diagonals or at least met the criteria of the closure error. Finally, a Moody diagram was elaborated with the results of the final vectors (Figure 8).

Table 3. Central lines

i	BF	BF'	BF''	BF^f	HD	HD'	HD''	HD^f
0	1.1	0.0	-1.6	0.9	0.2	0.0	-1.5	1.0
1	0.5	-0.6	-0.6	1.9	-0.5	-0.7	-1.1	1.5
2	1.1	0.1	0.4	2.9	-0.7	-0.9	-1.4	1.2
3	0.4	-0.7	0.0	2.5	0.7	0.5	0.0	2.5
4	-0.3	-1.4	-0.3	2.2	0.3	0.1	-0.5	2.0
5	-1.9	-3.0	-1.6	0.9	-0.2	-0.4	-1.0	1.5
6	-0.5	-1.5	-1.0	1.6	0.1	-0.1	-2.1	0.4

Source: Authors

Table 4. Perimeter lines

i	AC	AC'	AC''	AC^f	AG	AG'	AG''	AG^f
0	0.5	0.0	1.0	3.5	1.0	0.0	1.0	3.5
1	-1.5	-2.1	-1.2	1.3	0.0	-1.1	-0.2	2.3
2	-1.7	-2.3	-1.6	1.0	-0.4	-1.5	-0.8	1.7
3	-1.7	-2.2	-1.6	0.9	-1.,0	-2.0	-1.5	1.0
4	-0.6	-1.,2	-0.7	1.8	0.4	-0.7	-0.3	2.2
5	0.8	0.3	0.6	3.1	0.8	-0.2	-0.1	2.5
6	0.4	-0.1	0.0	2.5	1.,1	0.,0	0.0	2.5
i	CE	CE'	CE''	CE^f	GE	GE'	GE''	GE^f
0	-0.5	0.0	0.0	2.5	1.2	0.,0	0.0	2.5
1	-0.5	-0.,1	-0.2	2.4	-0.5	-1.7	-1.4	1.1
2	-1.7	-1.3	-1.5	1.1	0.8	-0.4	0.2	2.8
3	-2.3	-1.8	-2.1	0.4	-0.7	-1.9	-1.0	1.6
4	-1.6	-1.2	-1.5	1.0	1.4	0.2	1.4	3.9
5	-2.5	-2.0	-2.5	0.0	1.4	0.2	1.7	4.3
6	1.1	1.6	1.0	3.5	0.,4	-0.,8	1.0	3.5

Source: Authors

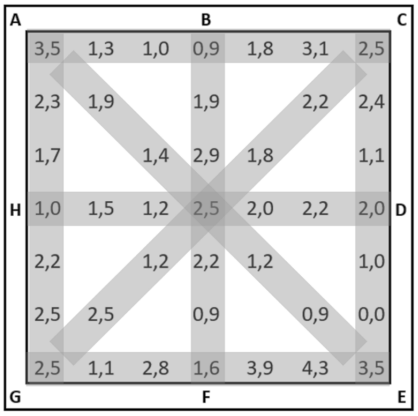


Figure 8. Measurement results obtained with the roto-translation method **Source:** Authors

Evaluating measurement uncertainty

Table 5 presents the uncertainty evaluation of the roto-translation method, which was carried out using Equation (15. Note that the straight edge ($u(se)$) is the factor with the greatest influence on the roto-translation's expanded uncertainty. In this work, the straight edge was calibrated using a CMM with high uncertainty. However, if required, this parameter can be reduced by calibrating the straight edge through a different measuring system or method with less uncertainty.

For the uncertainty evaluation of the Moody method, Equation (16) was used. The results are presented in Table 6.

Table 5. Uncertainty estimation for the roto-translation method

Source of uncertainty	$u(x_i)$	Units	c_i	ν_i	$(u(x_i)\times c_i)$
Resolution	$u(res)=\frac{res}{2\sqrt{3}}$	μm	1	50	0.29
Repeatability	$u(rep)=\frac{s_p}{\sqrt{n}}$	μm	1	2	0.33
Straight Edge	$u(SE)=\frac{U_{SE}}{k}$	μm	1	200	1.35
Gauge Pin	$u(gp)=\frac{U_{gp}}{k}$	μm	1	200	0.40
Gauge Block	$u(gb)=\frac{U_{gb}}{k}$	μm	1	200	0.02
Expanded uncertainty, ($k=2$)					**2.9 μm**

Source: Authors

Table 6. Uncertainty estimation for the Moody method

Source of uncertainty	$u(x_i)$	Units	c_i	ν_i	$(u(x_i)\times c_i)$
Resolution	$u(res)=\frac{res}{2\sqrt{3}}$	μm	1	50	0.01
Repeatability	$u(rep)=\frac{s_p}{\sqrt{n}}$	μm	1	2	0.30
Electronic Levels	$u(SE)=\frac{U_{el}}{k}$	μm	1	200	0.13
Ruler	$u(r)=\frac{\sqrt{(u_c c_c)^2+(u_\alpha c_\alpha)^2+(u_\theta c_\theta)^2}}{L}$	μm	1	200	0.0001
Expanded uncertainty, ($k=2$)					**1,1 μm**

Source: Authors

Table 6 shows that the Moody expanded uncertainty is influenced by the ruler used to trace the lines of the diagram, and that it is composed of the uncertainty of the ruler u_c, the uncertainty of the ruler coefficient of thermal expansion u_α, the laboratory temperature u_θ, and the distance between the centers of the electronic level feet L. It is important to consider this uncertainty since it directly influences the steps in which the electronics levels take measurements to estimate the deviation of each calibration point in the grid line diagram.

For both methods, *flatness* was defined as the distance between the base plane and the roof plane (Novyanto and Pratiwi, 2016; Skattum, 2015).

Table 7. Uncertainty estimation for each measurement method

Measurement Method	Flatness [μm]	Closure error		Uncertainty [μm]
		BF[μm]	HD[μm]	
Roto-translation	4.3	-1.2	-1.1	2.9
Moody	3.5	-2.3	-1.9	1.1

Source: Authors

Based on Table 7, it is important to note that the flatness results obtained with both methods are comparable, despite the fact that the resolution of the gauge pins ($\pm 1\mu$m) is 20 times higher than that of the electronic levels (0.05μm). Nevertheless, uncertainty entailed a great difference between both methods; the uncertainty of the roto-traslation was higher than that of the Moody method. This uncertainty is suitable for the calibration and verification of surface plates in the industry and in small laboratories, as has been previously reported (ASME-B89.3.7, 2013; ISO8512-2, 1990; Zahwi, Amer,

Abdou, and Elmelegy, 2013; Skattum, 2015). The closure error obtained for both methods is an important criterion for correcting the measurements, as it indicates potential issues associated with personnel, the measurement system, or the surface plate. Finally, the flatness values obtained with both methods range from 3.5 μm to 4.3 μm, meeting the grade-A flatness specifications for surface plates of this size (ASME-B89.3.7, 2013).

Validating the roto-translation method

To compare the results shown in Figures 7 and 8, each grid line result was represented with its uncertainty bars (Figures 9 and 10). These Figures demonstrate that both methods describe the same surface plate point by point, regardless of the differences in resolution and uncertainty.

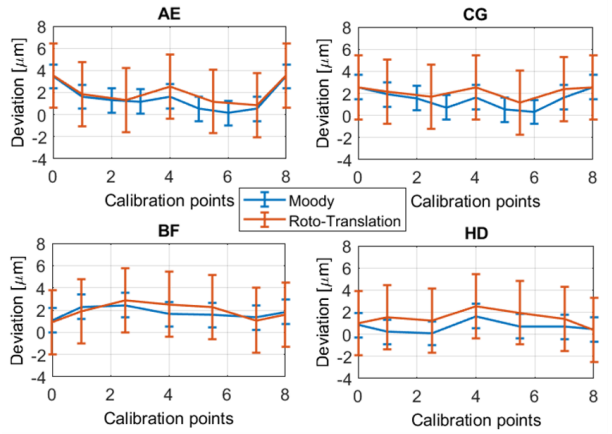


Figure 9. Validation of the roto-translation method with regard to the diagonals and central lines **Source:** Authors

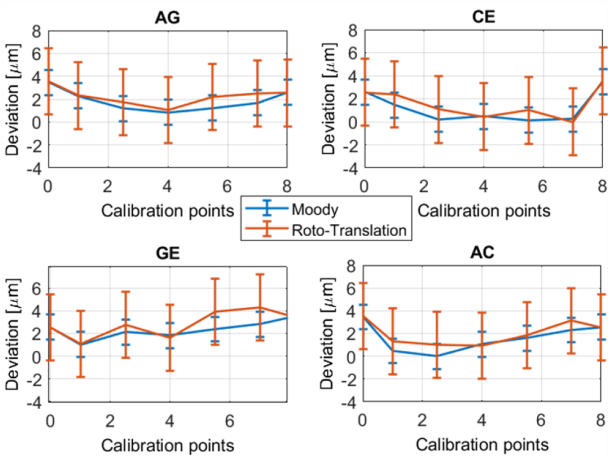


Figure 10. Validation of the roto-translation method with regard to perimeter lines **Source:** Authors

As shown in Figures 9 and 10, the flatness deviation variations between methods are most likely associated with the differences between the measurement systems, with the error introduced by the individual carrying out the calibration, and with the fact that the measurement points can slightly differ between methods because the gauge pins have to be placed exactly in each point of the grid lines. Meanwhile, for the electronic levels, the deviation of each point is estimated via geometric calculations.

To validate the results obtained with the roto-translation method, the normalized error E_n was calculated, using Equation (19) to evaluate the consistency of each calibration point of the Moody diagram (Zahwi *et al.*, 2013), as shown in Table 8.

Table 8. Estimation of the normalized error

	Normalized error							
i	AE	CG	AC	CE	GE	AG	BF	HD
0	0.01	-0.02	0.03	-0.01	-0.01	0.03	-0.04	0.06
1	0.08	0.09	0.26	0.27	0.02	0.00	-0.11	0.38
2	0.05	0.19	0.29	0.26	0.18	0.17	0.14	0.32
3	0.26	0.26	-0.04	-0.01	-0.07	0.06	0.27	0.27
4	0.25	0.23	0.06	0.26	0.45	0.29	0.19	0.38
5	0.11	0.25	0.24	-0.07	0.44	0.23	-0.12	0.25
6	0.01	-0.02	-0.01	0.03	0.03	-0.01	-0.07	-0.01

Source: Authors

As observed in Table 8, the E_n values are within the range of ±1 for all measured points. It could be said that there are no significant differences between the measured flatness deviations found by both methods. Given the above, the roto-traslation method is compatible with and comparable to the Moody method, and it can be used for calibrating and verifying grade-AA surface plates with dimensions over 609×1219 mm, grade-A plates with dimensions over 457×609 mm, and grade-B surface plates with dimensions over 305×457 mm (ASME-B89.3.7, 2013).

Conclusions

In validating the proposed roto-traslation method, we found that it is important to allow the surface plate, gauge pins, straight edge, and blocks to reach thermal stability, as this can directly affect the repeatability and reproducibility of the results (Skattum, 2015). Additionally, the uncertainty of this method is suitable for calibrations and intermediate verifications; it meets the requirements for the measurement and calibration of surface plates of grades AA to B (ASME-B89.3.7, 2013; Drescher, 2003). In Table 8, the normalized error reveals a bias that is sufficiently small in relation to the Moody method, demonstrating a point-by-point metrological compatibility between methods. Given the above, the roto-translation method can be recommended to laboratories that require constant verification of the state of surface plates, surfaces, large lenses, and mirrors (Yellowhair and Burge, 2008), as it implies a lower economic investment (*i.e.*, instrument purchasing and technical maintenance).

The expanded uncertainty obtained with the Moody method was U= 1.1μ m, with a coverage factor k = 2. According to the documentation (ASME-B89.3.7, 2013) for a grade-A surface plate of 24 × 24 (*i.e.* 609.6 × 609.6 mm), the overall flatness has to be 160μin (*i.e.*, 4,06 μm). Thus, the flatness results obtained with the Moody method are within the standard specifications.

As part of the calibration procedure with the roto-translation method, we found that the straightness ruler marks have to be perfectly aligned with the reference surface calibration points, as this directly influences the results and can cause the closing error to be greater than 2.5 μm. This was proven by different technical professionals of the laboratory, and it directly affected the reproducibility of the method.

Since the accuracy of the measurement instruments employed by the roto-translation method is lower than that of electronics levels, there is a magnitude difference between the uncertainties presented in Table 7 and Figure 8. Since the straight edge was calibrated with a CMM, its uncertainty is the most significant component in the uncertainty model of the roto-translation method.

With the results presented in this work, the research team means to encourage new and small laboratories to use this method to calibrate and characterize their surface plates.

Acknowledgments

This work was carried out within the framework of the project titled *Plan for strengthening the INM as a research center* (code 9932100271370), funded by the Ministry of Science, Technology, and Innovation (Minciencias) and jointly executed by the INM and Universidad Nacional de Colombia through the Special Cooperation Agreement 001 of 2020.

CRedit author statement

Daniela I. Garzon carried out most of the experiments, collected all the data for the formal analysis, and developed software programs and applications for data processing and results visualization. Jorge L. Galvis conceived the idea and did the background research, supervised the research, provided critical feedback, and led the research and the validation process. David A. Plazas carried out all the calibrations of the measurement instruments used during the experiments and participated in the validation process. Victor H. Gil designed the first approximation to the measurement method and participated in the validation process. Ovidio Almanza reviewed, supported, and oversaw the research process.

Conflicts of interest

The authors declare no conflict of interest.

References

Ali, S. H., and Buajarern, J. (2013, Dec). New measurement method and uncertainty estimation for plate dimensions and surface quality. *Advances in Materials Science and Engineering*, *2013*, 1-10 https://doi.org/10.1155/2013/918380.

ASME-B89.3.7. (2013). *B89.7.3. Granite Surface Plates*. ASME.

Azaryan, N., Budagov, J., Gayde, J.-C., Girolamo, B. D., Glagolev, V., Lyablin, M., ... Shirkov, G. (2017, 1). The innovative method of high accuracy interferometric calibration of the precision laser inclinometer. *Physics of Particles and Nuclei Letters*, *14*, 112-122 http://link.springer.com/10.1134/S154747711701006X.

DIN874-1. (2003). Geometrical product specifications (gps) - straight edges - part 1: Steel straight edges; dimensions, technical delivery conditions. *DIN 874-1:2003-11*, *2003*.

Drescher, J. (2003). Analytical estimation of measurement uncertainty in surface plate calibration by the moody method using differential levels. *Precision engineering*, *27*(3), 323–332.

Ehret, G., Reinsch, H., and Schulz, M. (2019). Interferometric and deflectometric flatness metrology with nanometre measurement uncertainties for optics up to 1 metre at PTB. In S. Han, T. Yoshizawa, S. Zhang, and B. Chen (Eds.), *Optical metrology and inspection for industrial applications vi* (Vol. 11189, p. 1118905 https://doi.org/10.1117/12.2538872.). SPIE.

Ehret, G., Schulz, M., Fitzenreiter, A., Baier, M., Jockel, W., Stavridis, M., and Elster, C. (2011). Alignment methods for ultraprecise deflectometric flatness metrology. In P. H. Lehmann, W. Osten, and K. Gastinger (Eds.), *Optical measurement systems for industrial inspection vii* (Vol. 8082, p. 808213 https://doi.org/10.1117/12.889325.). SPIE.

Espinosa, O. C., Diaz, P., Baca, M. C., Allison, B. N., and Shilling, K. M. (2008). *Comparison of calibration methods for a surface plate*. (Tech. Rep.). Sandia National Lab.(SNL-NM), Albuquerque, NM (United States).

Geckeler, R. D., Schumann, M., Just, A., Krause, M., Lassila, A., and Heikkinen, V. (2022, feb). A comparison of traceable spatial angle autocollimator calibrations performed by ptb and vtt mikes. *Metrologia*, *59*(2), 024002 https://dx.doi.org/10.1088/1681-7575/ac42b9.

given i=K., f., given=Kouyu. (n.d.). On a method for measuring flatness by two-point connecting method. , *8*(30), 274–280 http://dx.doi.org/10.1299/jsme1958.8.274.

Glubokov, A., Glubokova, S., Afonina, I., Zelensky, A., and Semenishchev, E. A. (2022, 12). Automated measuring system for straightness and flatness deviations of extended surfaces. *Optical Metrology and Inspection for Industrial Applications IX* http://dx.doi.org/10.1117/12.2646259..

Gromczak, K., Gąska, A., Kowalski, M., Ostrowska, K., Sładek, J., Gruza, M., and Gąska, P. (2016, dec). Determination of validation threshold for coordinate measuring methods using a metrological compatibility model. *Measurement Science and Technology*, *28*(1), 015010 https://dx.doi.org/10.1088/1361-6501/28/1/015010.

Gusel, A., Acko, B., and Sostar, A. (2000a). Assuring the traceability of electronic levels for calibration of granite surface plates. In *Proc. xvi imeko world congress*.

Gusel, A., Acko, B., and Sostar, A. (2000b). Assuring the traceability of electronic levels for calibration of granite surface plates. In *Proc. xvi imeko world congress*.

Heikkinen, V., Byman, V., Palosuo, I., Hemming, B., and Lassila, A. (2017, 6). Interferometric 2d small angle generator for autocollimator calibration. *Metrologia*, *54*, 253-261 https://iopscience.iop.org/article/10.1088/1681-7575/aa648d.

Heydorn, K. (2010, Nov 01). Metrological compatibility a key issue in future accreditation. *Accreditation and Quality Assurance*, *15*(11), 643-645 https://doi.org/10.1007/s00769-010-0691-8.

ISO. (2006). Guide iso/cei 99 2007 vocabulaire international de métrologie concepts fondamentaux et généraux et termes associés (vim). *Vocabulary*, *2007*(VIM), 1–150.

ISO. (2015). Statistical methods for use in proficiency testing by interlaboratory comparison. *Iso 13528,*, *2015*.

ISO8512-2. (1990). Surface plates-part 2: Granite. *ISO 8512-2*, *1990*.

JCGM-BIPM. (2008). Evaluation of measurement data | guide to the expression of uncertainty in measurement. *Int. Organ. Stand. Geneva ISBN*, *50*(September), 134.

Ju HUO, M. Y., Yunhui LI. (2018). Multi-channel signal parameters joint optimization for gnss terminals. *Journal of Systems Engineering and Electronics*, *29*(4), 844. https://dx.doi.org/10.21629/JSEE.2018.04.19. Retrieved from https://www.jseepub.com/EN/abstract/article_6421.shtml

Lakota, S., and Gorog, A. (2011). Flatness measurement by multi-point methods and by scanning methods. *Ad Alta: Journal of Interdisciplinary Research*, *1*(1), 124–127.

Meijer, J., and Heuvelman, C. (1990). Accuracy of surface plate measurements|general purpose software for flatness measurement. *CIRP annals*, *39*(1), 545–548.

Mikó, B. (2021, 2). Assessment of flatness error by regression analysis. *Measurement*, *171*, 108720 http://dx.doi.org/10.1016/j.measurement.2020.108720.

Moody, J. (1955). How to calibrate surface plates in the plant. *The Tool Engineer*, *1955*, 85–91.

Novyanto, O., and Pratiwi, E. (2016). A preliminary study to evaluate the topography of narrow surface plate. *Instrumentasi*, *39*(1), 1–8 https://jurnalinstrumentasi.bsn.go.id/index.php/ji/article/view/67.

Schulz, M., Ehret, G., and Křen, P. (2013). High accuracy flatness metrology within the european metrology research program. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, *710*, 37-41. https://doi.org/10.1016/j.nima.2012.10.112. Retrieved from https://www.sciencedirect.com/science/article/pii/S0168900212012855 (The 4th international workshop on Metrology for X-ray Optics, Mirror Design, and Fabrication)

Skattum, G. A. (2015, Mar). Estimating thermal effects for granite surface plate calibration. *NCSLI Measure*, *10*(1), 50–58 https://dx.doi.org/10.1080/19315775.2015.11721716.

Yang, Y., Wang, T., and Zhu, G. (2021, jul). Calibration method for the bubble angle of inclination for the spirit levels based on the calibration device. *Journal of Physics: Conference Series*, *1982*(1), 012160 https://dx.doi.org/10.1088/1742-6596/1982/1/012160.

Yellowhair, J., and Burge, J. H. (2008). Measurement of optical flatness using electronic levels. *Optical Engineering*, *47*(2), 023604 https://doi.org/10.1117/1.2831131.

Zahwi, S., Amer, M., Abdou, M., and Elmelegy, A. (2013). On the calibration of surface plates. *Measurement*, *46*(2), 1019-1028. https://doi.org/10.1016/j.measurement.2011.10.009. Retrieved from https://www.sciencedirect.com/science/article/pii/S0263224111003551

Research Article/Mechanical Engineering, Mechatronics, and Materials Science http://doi.org/10.15446/ing.investig.107624

Influence of Thermal Aging on the Sliding Wear of a Biocomposite Material Reinforced with Bamboo Fibers

Influencia del envejecimiento térmico sobre el desgaste deslizante de un material biocompuesto reforzado con fibras de bambú

Eudi Blanco[1], Jorge I. Fajardo[2], Edwuin J. Carrasquero[3], Caribay Urbina[4], Luis M. López[5], and Luis Cruz[6]

ABSTRACT

This study evaluated the effect of thermal aging on the tribological properties of biocomposites formed by an isotactic polypropylene matrix (iPP) reinforced with 20 wt% (PP/20F), 30 wt% (PP/30F), and 40 wt% (PP/40F) of randomly oriented bamboo fibers. iPP, along with the grafting of maleic anhydride molecules (MAPP), was used as a coupling agent. The accelerated thermal aging involved the continuous heating of the materials at 98 ˚C for 10 days. Wear tests were performed under the Pin-on-Disk configuration to determine the wear factor (K) and the friction coefficient (μ) of the materials. After thermal aging, the μ value of the PP/20F composite increased by 40.5%, while, for raw PP, PP/30F, and PP/40F, the increase was 2.1, 7.5, and 2.2%, respectively. The aged PP/30F composite achieved the highest μ value. The loss of wear resistance due to aging was more prominent in the raw PP. The K factor of the aged and unaged PP/20F was the lowest. The use of scanning electron microscopy allowed identifying that adhesive, abrasive, and fatigue wear mechanisms were the dominant ones.

Keywords: polypropylene, natural fiber, bamboo, biocomposite, tribology, sliding wear, wear rate, coefficient of friction, aging

RESUMEN

En este estudio se evaluó el efecto del envejecimiento térmico sobre las propiedades tribológicas de biocompuestos formados por una matriz de polipropileno isotáctico (iPP) reforzada con 20 wt% (PP/20F), 30 wt% (PP/30F) y 40 wt% (PP/40F) de fibras de bambú orientadas al azar. Se utilizó iPP con injertos de moléculas de anhídrido maleico (MAPP) como agente de acople. El envejecimiento térmico acelerado consistió en un calentamiento continuo de los materiales a 98 ˚C durante 10 días. Se realizaron ensayos de desgaste bajo la configuración *Pin-on-Disk* para determinar el factor de desgaste (K) y el coeficiente de fricción (μ) de los materiales. Después del envejecimiento térmico, el valor de μ del compuesto PP/20F aumentó en un 40.5 %, mientras que, para el PP puro, PP/30F y PP/40F, el aumento fue de 2.1, 7.5 y 2.2 % respectivamente. El compuesto PP/30F envejecido alcanzó el mayor valor de μ. La pérdida de la resistencia al desgaste debido al envejecimiento fue más resaltante en el PP puro. El factor K del PP/20F envejecido y no envejecido fue el más bajo. El uso de microscopía electrónica de barrido permitió identificar que los mecanismos de desgaste adhesivo, abrasivo y por fatiga fueron los dominantes.

Palabras clave: polipropileno, fibra natural, bambú, biocompuesto, tribología, desgaste deslizante, tasa de desgaste, coeficiente de fricción, envejecimiento

Received: March 4th, 2023
Accepted: April 24th, 2024

[1] Materials engineer, Universidad Simón Bolívar, Venezuela. MSc in Mechanical Engineering, Universidad Central de Venezuela, Venezuela. Affiliation: Department of Physical Metallurgy, Faculty of Engineering, Universidad Central de Venezuela, Venezuela. E-mail: eudiblanco@gmail.com

[2] Mechanical engineer, Universidad Politécnica Salesiana, Ecuador. PhD in Engineering, Universidad Pontificia Bolivariana, Colombia. Affiliation: New Materials and Transformation Processes Research Group (GiMaT), Mechanical Engineering Faculty, Universidad Politécnica Salesiana, Ecuador. E-mail: jfajardo@ups.edu.ec

[3] Metallurgical engineer, Universidad Central de Venezuela, Venezuela. PhD in Mechanics, University of Sciences and Technologies of Lille, France, and PhD in Engineering Sciences, Universidad Central de Venezuela, Venezuela. Affiliation: Research Group on Characterization, Processing and Protection of Materials, Faculty of Science and Engineering, Universidad Estatal de Milagro, Ecuador. E-mail: ecarrasqueror@unemi.edu.ec

[4] Degree in Chemistry, Universidad Central de Venezuela, Venezuela. PhD in Science, with an emphasis on Chemistry, Universidad Central de Venezuela, Venezuela. Affiliation: Dr. Mitsuo Ogura Center for Electron Microscopy, Faculty of Sciences, Universidad Central de Venezuela, Venezuela. E-mail: caribayurbina@gmail.com

[5] Mechanical engineer, Universidad Politécnica Salesiana, Ecuador. PhD in Industrial Engineering, Universidad San Marcos, Perú. Affiliation: New Materials and Transformation Processes Research Group (GiMaT), Mechanical Engineering Faculty, Universidad Politécnica Salesiana, Ecuador. E-mail: llopez@ups.edu.ec

[6] Mechanical engineer, Universidad Pontificia Bolivariana, Colombia. PhD in Industrial Engineering, Polytechnic University of Madrid, Spain. Affiliation: Research Group on New Materials, School of Engineering, Universidad Pontificia Bolivariana, Colombia. E-mail: luis.cruz@upb.edu.co

Introduction

The tribological properties of composites play an important role in the selection of materials for different purposes. As the composite materials used in industries are exposed to different types of wear (*e.g.*, adhesive or abrasive wear) during their service life, it is very important to consider their tribological performance. Moreover, by considering the tribology of composite materials, we can save a lot of the energy wasted in overcoming the friction between two moving surfaces (Kerni *et al*., 2020). Recent research has focused on studying the friction and wear properties of polymers reinforced with natural fibers. Progress has been made towards improving these properties in composites, in light of the fact that 90% of mechanical part failures are due to tribological loading conditions (Yousif *et al*., 2007). Polymers' wear and friction behavior can be improved or deteriorated by the incorporation of fibers as reinforcement (Bajpai *et al*., 2012).

Polymers are the most commonly used matrices in natural fiber composites due to their light weight. They can be processed at low temperatures. Both thermoplastic and thermoset polymers have been used for matrices with natural fibers. Matrix materials have temperature limitations because most natural fibers are thermally unstable above 200 °C. Thermoplastics (such as polyethylene, polypropylene, polyvinyl chloride, and polystyrene) and thermosets (such as epoxy resin, unsaturated polyester, PF, and vinyl ester) are mainly used in the matrix. Thermoplastics can be softened by applying heat and hardened by cooling. They can therefore be easily recycled (Jariwala and Jain, 2019).

In some applications such as bearings, where one part slides over another, wear is often significant. As low friction is essential, thermoplastics like polypropylene perform well in these applications. These thermoplastics offer a number of additional benefits in addition to wear resistance, such as good corrosion resistance, lower weight, and less noise. Parts made from these thermoplastics also tend to generate less friction and heat, which increases service life and ultimately reduces maintenance costs.

The effect of accelerated thermal aging on the tribological behavior of polypropylene reinforced with bamboo fibers is not sufficiently studied in the literature, so this work is a contribution to the knowledge of the effect of this type of aging on the wear and friction behavior of this biocomposite. Noting the enormous advantages and opportunities associated with natural fibers, in addition to the promising application of polymeric biocomposites in the automotive industry, it is necessary to further investigate the tribological properties of these materials after aging. For this industry, and in many others, long-term stability is important because thermal and oxidative degradation is known to affect the performance of polymer composites.

Methodology

Materials

Isotactic polypropylene (PP) homopolymer was used as polymeric matrix, whose melt flow index (MFI) was 12 g/10 min (at 230 °C/2.16 kg), measured under ISO 1133-1:2011 (2011). The polypropylene matrix was reinforced with randomly oriented bundles of short *Guadua angustifolia* Kunth (GAK) fibers obtained through the steam explosion technique, with a length distribution of 1.0 ± 0.5 mm and a diameter distribution of 0.4 ± 0.2 mm. The onset temperature of degradation was 200 °C. We used maleic anhydride grafted polypropylene (MAPP), with a weight-average molecular weight (Mw) and number-average molecular weight (Mn) of 9100 and 3900, respectively, as the coupling agent between the fiber and the matrix. Molecular weights were determined through the gel permeation chromatography (GPC) method. A maleic anhydride percentage of 8% was used.

Composite processing

PP/Bamboo fiber (PP/FB) composite pellets were prepared using a Dr. Collin E-20 T single-screw extruder, with the following composition: PP with 20 wt% BF (PP/20F), PP with 30 wt% BF (PP/30F), and PP with 40 wt% BF (PP/40F). Then, type 1B 527 ISO standardized specimens (Figure 1) – as per ASTM D638-14 (ASTM International, 2014) – were injected with the different compositions using a Dr. BOY injection molding machine E35. From the specimens obtained by injection, 15 × 15 × 5 mm samples were cut using a Struers Minitom low-speed precision cutting machine with a diamond disc.

Figure 1. Type 1B 527 ISO standardized specimen (ASTM D638-14) made of PP/20F biocomposite, from which the samples for the study were extracted
Source: Authors

Accelerated thermal aging

A group of raw PP and PP/FB samples were subjected to a thermal aging process by continuous heating at 98 °C for 10 days using a BLUE M SW-11TA-1 gravity convection oven (Figure 2). Subsequently, the samples were removed from the oven to cool to room temperature. The temperature selection for this heat treatment was based on its proximity to the continuous-use temperature of homopolymer polypropylene (100 °C).

Figure 2. Gravity convection oven used to perform the accelerated thermal aging test
Source: Authors

Sliding wear tests

The tribological study of PP/FB composites was carried out via sliding wear tests, according to ASTM G99-17 (ASTM International, 2017). These tests were carried out on a standard CSM Instruments tribometer (Figure 3), under the Pin-on-Disk configuration, which consisted of sliding a stationary pin (which acts as a support for the sphere or ball) against the surface of the sample under study coupled to a rotating disk (Figure 4). This equipment includes softwarolled relative humidity of 70 ± 2%. The sliding conditions were as follows: 31 836 cycles, equivalent to a total sliding distance of 600 m; a linear sliding velocity of 15 cm/s; and an applied normal load of 10 N. The static counterparts used were 6 mm diameter spheres of AISI 52100 steel.

Figure 3. Tribometer used for the study of sliding wear
Source: Authors

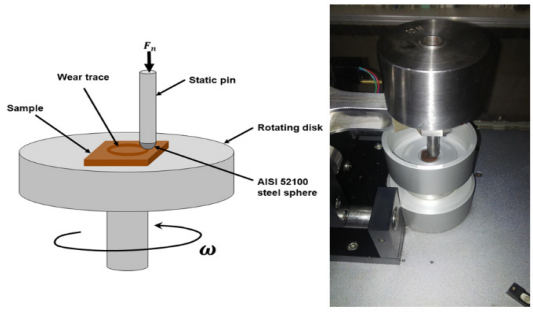


Figure 4. Tribometer used for the study of sliding wear
Source: Authors

For each of the sliding wear tests, the average coefficient of friction values (μ_p) was obtained, and, subsequently, the wear rate (K) was calculated using Equation (1).

$$K = \frac{V_w}{F_n \cdot d} \qquad \left[\text{mm}^3/\text{N}\cdot\text{m}\right] \tag{1}$$

where V_w is the wear volume in [mm^3], F_n is the applied normal load in [N], and d is the total sliding distance in [m]. Equation (2) was used to determine V_w.

$$V_w = \frac{\Delta m}{\rho} \cdot 1000 \quad [\text{mm}^3] \tag{2}$$

where Δm is the mass loss of the specimens in [g] and numerically represents the difference in mass of the specimens before and after the tribology test; ρ is the density of the PP and PP/FB specimens in [g/cm^3], which was obtained by following ASTM D4892-14(2019)e1 (ASTM International, 2019). The values of the densities of the samples are shown in Table 1. The mass difference (Δm) was determined using a Gram digital precision balance, model FR-500.

Table 1. Densities (ρ) of the PP and PP/FB samples

Sample	ρ [g/cm^3]
PP	0.905
PP/20F	0.960
PP/30F	0.990
PP/40F	1.030

Source: Authors

In addition, the friction coefficient achieved by each sample was analyzed and presented in terms of friction fluctuations ($\Delta\mu$), the friction stability coefficient (μ_s), and the coefficient of variability (μ_v). These parameters were calculated through the following formulas (Singh *et al*., 2019):

$$\mu_s = \frac{\mu_p}{\mu_{max}} \tag{3}$$

$$\mu_v = \frac{\Delta\mu}{\mu_p} \tag{4}$$

where μ_p is the average μ recorded for all tests, and $\Delta\mu$ is the difference between the maximum (μ_{max}) and minimum (μ_{min}) values of μ obtained in each test.

Scanning electron microscopy (SEM)

After performing the tribological tests, high-resolution images of the wear traces of all PP and PP/FB samples, with and without aging, were obtained using a scanning electron microscope (FEI, model QUANTA FEG 250) with a secondary electron detector. The surfaces to be visualized were previously metalized by applying a gold (Au) coating using an ion sputter metallizer.

Results and discussion

Effect of fiber content on friction and wear of PP/FB biocomposites

Figure 5 shows the variation of the friction coefficient (μ) with sliding distance for raw PP and PP/FB biocomposites. The friction value increased during the initial phase of wear, which is due to the increase in the contact area between the polymer and the metallic counterpart. Thereafter, it stabilizes and follows a steady state behavior.

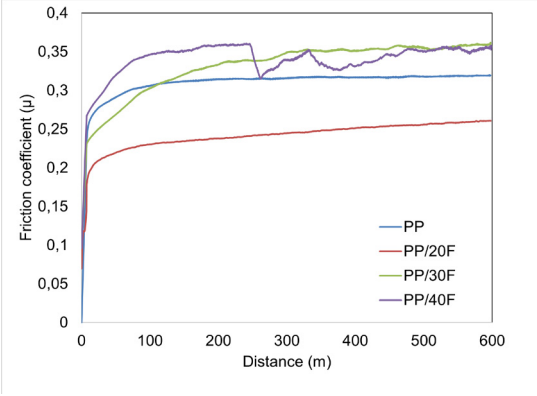


Figure 5. Variation of the friction coefficient (μ) with the sliding distance
Source: Authors

The fluctuations of the friction coefficient (Δμ) are measured as the difference between the highest and lowest μ value reached in each of the tests. This is shown in Figure 6. It can be observed that the fluctuations increase with the increase in bamboo fiber content. Within the group of composite materials, the PP/20F sample showed the lowest fluctuation (0.19).

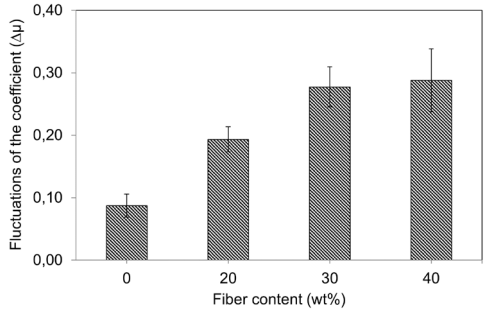


Figure 6. Variation of μ fluctuations (Δμ) with bamboo fiber content
Source: Authors

Moreover, the coefficients of friction stability (μ_s) and variability (μ_v) were calculated for the materials (Figure 7). For a good friction behavior of the material, high μ_s values (close to 1) and low μ_v values (close to 0) are generally needed (Singh *et al.*, 2019). In the tribology tests carried out on the samples, it was found that the μ_s value gradually decreases with the percentage of bamboo fibers, while the μ_v value increases. These results showed that noticeable fluctuations of μ are obtained when PP is reinforced with bamboo fibers, and that these fluctuations increase when the number of fibers in the polymeric matrix is increased. Inadequate fiber distribution or mixing at a higher ratio (40 wt%) may cause structural discontinuities in the composites and thus result in higher friction variability. Meanwhile, at a lower ratio (20 wt%), homogeneity in the mixture could have been achieved, resulting in high friction stability (Singh *et al.*, 2019), as observed in Figure 5.

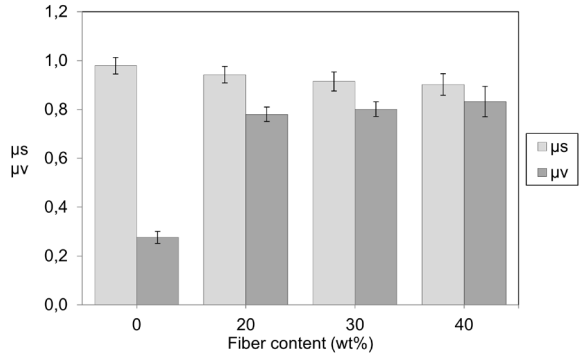


Figure 7. Variation in the coefficients of friction stability (μ_s) and variability (μ_v) with the bamboo fiber content
Source: Authors

The graph in Figure 8 shows the variation of μ with the fiber content of the PP/FB composites, and Table 2 shows the values obtained for this coefficient. PP recorded a decrease of -21.6% of its μ when reinforced with 20 wt% of bamboo fibers. This may be due to the presence of wear particles, commonly called *debris*, *i.e.*, polymer residues that accumulate and plastically deform in the wear trace, creating a soft polymer film that causes a decrease in friction or adheres to its sliding counterpart in a phenomenon called *continuous film transfer* (Nirmal *et al.*, 2012). This film prevents the polymer from coming into direct contact with the metal counterpart material, thus preventing abrasive action and reducing the rate of wear. The thinner and more well distributed the film, the lower the values of the friction and wear coefficients (Mimaroglu *et al.*, 2018). Therefore, the PP/20F composite showed a better performance against friction.

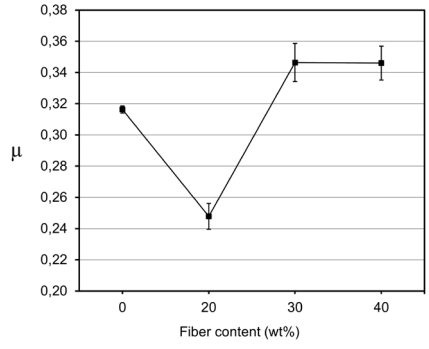


Figure 8. Variation of μ with a fiber content of PP/FB biocomposites
Source: Authors

On the other hand, the μ value of PP shows increases of 9.5 and 9.3% when reinforced with 30 and 40 wt% bamboo fibers, respectively. A higher reinforcement content in turn generates a greater number of fibers exposed during the sliding wear tests. Such fibers tend to detach or fracture and form debris, or a 'third body' between the (metal-polymer) contact surfaces (Nirmal *et al*., 2012; Narish *et al*., 2011). This can cause high relative motion shear strength during sliding wear testing, thereby contributing to a much higher friction coefficient value (Nirmal *et al*., 2012).

Although the plastic component of the biocomposite produces the soft layer of polymer that causes the friction coefficient to decrease, the presence of the fiber is also essential to increasing wear resistance. An amount of fiber greater than 20 wt% causes an increase in the friction coefficient of the biocomposite but could provide the PP with better wear resistance.

Table 2. Coefficient of friction (μ) and wear factor (K) obtained for raw PP and PP/FB biocomposites

Sample	μ	K [×10^{-4} mm^3/N m]
PP	0.3163 ± 0.0023	1.3808 ± 0.0836
PP/20F	0.2478 ± 0.0083	1.0416 ± 0.0220
PP/30F	0.3464 ± 0.0122	1.1784 ± 0.0656
PP/40F	0.3460 ± 0.0108	1.2944 ± 0.0263

Source: Authors

Figure 9 shows the variation of the wear rate (K) with the number of fibers in the composites, whose values are shown in Table 2. In the results obtained for this important tribological parameter, it can be seen that the wear rate of PP decreases with the presence of fibers, with the PP/20F composite showing the lowest K value, representing a decrease of -24.6% in the K value obtained for pure PP. According to the studies carried out by Blanco *et al*. (2020), the stiffness of PP increases with the content of bamboo fiber, given the reinforcing effect (decreasing tenacity) that this fiber achieves in PP. Materials with high stiffness have a better wear behavior, since the contact area between the metal and the polymer is reduced and the shear strength increases (Sınmazçelik and Yılmaz, 2007). Therefore, PP/FB composites exhibited a low wear rate compared to PP without fibers. Specifically, the PP/20F composite achieved a higher wear resistance, which could be due to the better bonding between matrix and fiber (Mimaroglu *et al*., 2018) that occurs during the manufacture of the composite, since at low reinforcement concentrations (20 wt%), a greater contact area between both components or a greater impregnation of the fibers is achieved and, consequently, a better reinforcement, which translates into a higher wear resistance (Sınmazçelik and Yılmaz, 2007). On the other hand, increasing the fiber content to 30 and 40 wt% also resulted – although to a lesser extent – in a decrease in the wear rate of PP (-14.4 and -65.%, respectively). Therefore, fiber reinforcement reduces the wear rate of the composites (Sınmazçelik and Yılmaz, 2007), with PP/20F exhibiting the best performance against wear.

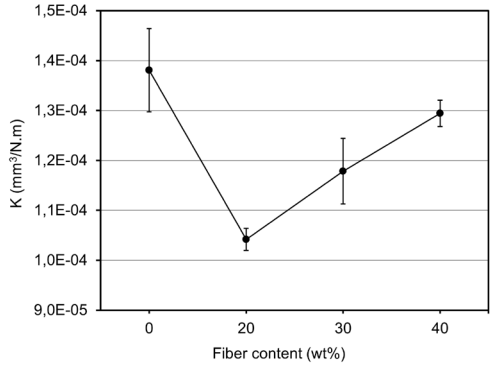


Figure 9. Variation of wear rate (K) with a fiber content of PP/FB biocomposites
Source: Authors

In the group of PP/FB biocomposites, there was a progressive increase in K as the amount of fiber increased from 20 to 40 wt%, with the PP/40F composite producing the highest wear rate within the group, which represented an increase of 24% over that obtained by the PP/20F composite. This means that the amount of wear in the PP/FB samples increased with the fiber content. Such behavior may be due to a loss of reinforcement of the composites, caused by the detachment of the fibers from the polymeric matrix during tribological tests. The increase in the weight percentage of fibers in the polymer matrix produces a greater detachment of the fibers due to a localized temperature increase at the contact surface between the metal and the sample (Sınmazçelik and Yılmaz, 2007). In general, the low tribological performance of polymers may be related to their viscoelasticity and thermal properties. Sliding contact between the specimen and the metal counterface results in heat generation at the asperities, which in turn increases the temperature at the interface. The formation of thermal gradients due to non-uniform temperature distributions creates thermal stresses in the specimen during sliding. These stresses weaken the natural adhesion of the fiber and matrix, causing the fibers to loosen and become easily sheared as a result of repeated axial thrust (Yallew *et al*., 2014). This loosening of the fibers is usually greater the higher the fiber content of the composite (40 wt% in this case). Furthermore, the increase in temperature, generated by the thermomechanical loading, softens the surface of the polymer, which entails decreased hardness. Likewise, there is a dramatic loss in shear strength and, therefore, the penetration resistance of the metallic counterpart decreases. Consequently, higher wear volumes are produced (Sınmazçelik and Yılmaz, 2007; Nirmal *et al*., 2012).

Effect of thermal aging on friction and wear of PP/FB biocomposites

Figure 10 shows a comparison between the friction curves of thermally aged and non-thermally aged PP without fibers and PP/FB samples. The results show that the aged samples

produced higher friction coefficients than the unaged ones – such behavior will be explained later. The average μ values obtained for the aged samples are shown in Table 3.

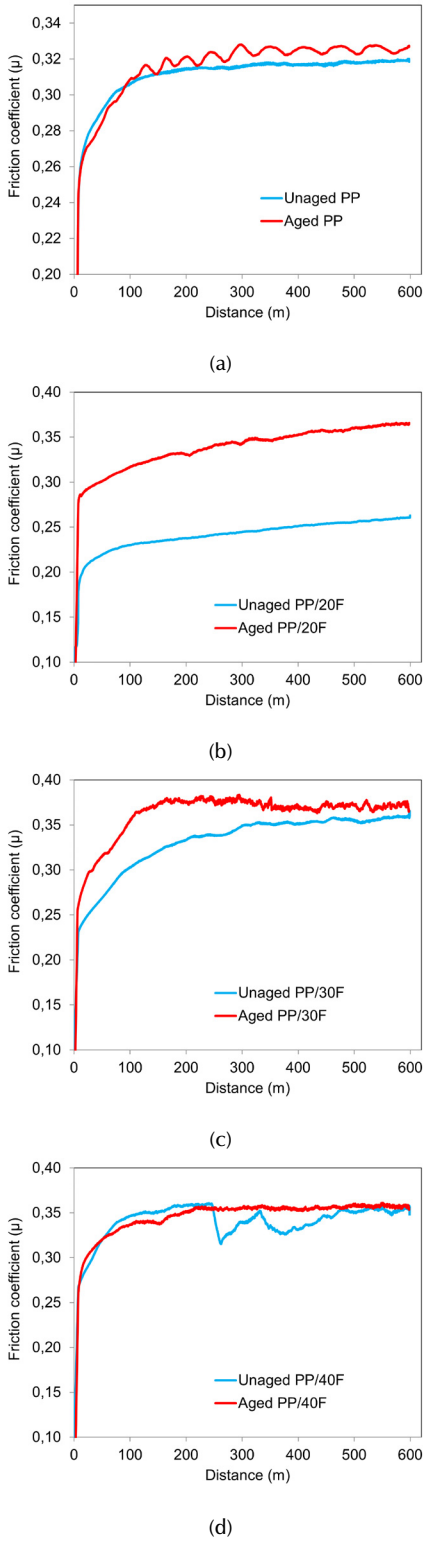


Figure 10. Comparison of the coefficient of friction (μ) *vs.* the sliding distance curves for aged and unaged samples of a) PP, b) PP/20F, c) PP/30F, and d) PP/40F
Source: Authors

Table 3. Coefficient of friction (μ) and wear factor (K) obtained for the aged materials

Sample	μ	K [×10^{-4} mm^3/N m]
PP	0.3230 ± 0.0039	2.8589 ± 0.0362
PP/20F	0.3482 ± 0.0124	1.1285 ± 0.1228
PP/30F	0.3726 ± 0.0042	1.2630 ± 0.0738
PP/40F	0.3538 ± 0.0052	1.3751 ± 0.0586

Source: Authors

It is important to note that the comparative analyses presented below are based on a comparison of the friction parameters (μ_v, μ_s, and μ) and the wear rate (K) obtained for the aged and non-aged samples. All this, while keeping the fiber content fixed.

Figure 11 shows a comparative plot of the friction coefficient of variability (μ_v) of the aged and unaged samples. The aged PP/40F composite exhibited a greater decrease in μ_v (-65.8%) than the unaged sample, *i.e.*, for the PP/40F composite, aging significantly improved the stability of the friction coefficient values, which is evident in Figure 10d.

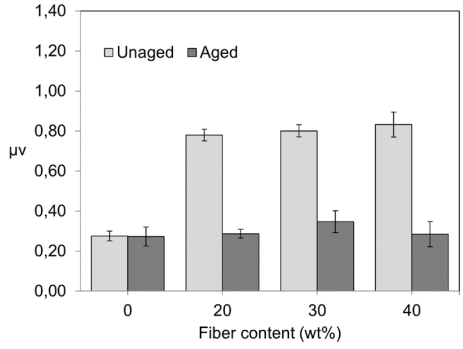


Figure 11. Comparison of the coefficient of friction variability (μ_v) of aged and unaged PP and PP/FB samples
Source: Authors

Figure 12 shows a comparison of the coefficient of friction stability (μ_s) obtained for the aged and unaged PP and PP/FB samples. The highest stability in μ_s values during sliding wear tests occurred with the aged PP/30F and PP/40F composites. This improvement in the friction stability of the composites with a higher percentage of fibers could be due to the increase in the stiffness obtained when they are previously subjected to isothermal heating for some time (Blanco *et al.*, 2020; Blanco-Sánchez *et al.*, 2022).

Figure 13 depicts a graph comparing the value of μ obtained for the aged and unaged samples. This graph confirms what was observed in Figure 10: this parameter increased as a consequence of aging in all the samples. This increase was more noticeable in the PP/20F composite, and less marked in raw PP and PP/40F. After thermal aging, the coefficient of friction of the PP/20F composite increased by 40.5%, while, for the raw PP and the PP/30F and PP/40F composites, the

increase was 2.1, 7.5, and 2.2%, respectively. The aged PP/30F composite showed the highest μ value (0.37).

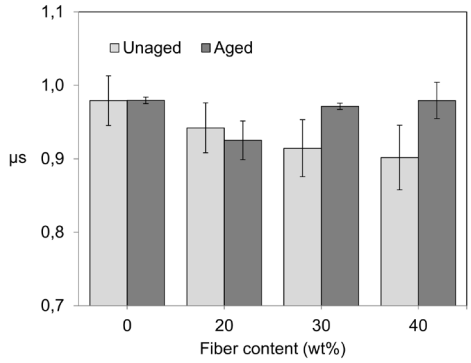


Figure 12. Comparison of the coefficient of friction stability (μ_s) of aged and unaged PP and PP/FB samples
Source: Authors

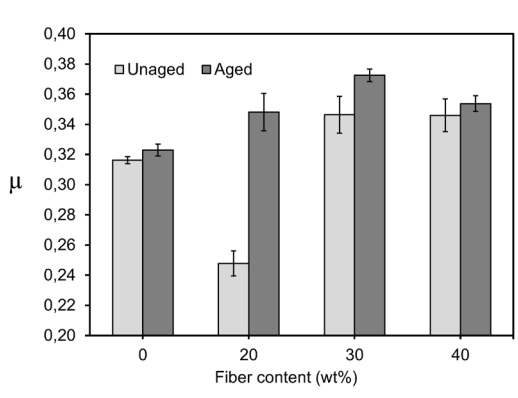


Figure 13. Effect of thermal aging on the coefficient of friction (μ) of raw PP and PP/FB biocomposites
Source: Authors

The increase in the friction parameter of the samples, due to thermal aging, could be due to the secondary crystallization process of PP (annealing) that occurs during heating, which can lead to an increase in the degree of crystallinity or the size of the pre-existing polymer crystals. This makes these materials stiffer and represents a significant decrease in toughness, increasing shear strength especially when there is a low amount of fiber in the matrix (PP/20F) (Inácio *et al.*, 2018; Sınmazçelik and Yılmaz, 2007; Blanco *et al.*, 2020; Blanco-Sánchez *et al.*, 2022, Law *et al.*, 2008), and, as a consequence, an increase in the coefficient of friction of the material occurs. The increase of μ in the composites with a higher bamboo fiber content (PP/30F and PP/40F) could be mostly due to the presence of a higher number of exposed fibers in the surface layer during sliding wear tests, which are easily detached from the matrix given the possible weakening and/or rupture of the polymer/compatibilizer bonds due to aging (Inácio *et al.*, 2018). These shed fibers tend to form material debris in the space between the metal and the polymer, thus causing increased friction between the two surfaces (Nirmal *et al.*, 2012).

Figure 14 compares the K values obtained for the aged and non-thermally aged samples. In this graph, it can be observed that the aged samples reached a low wear resistance compared to the unaged ones. This behavior could be attributed to an embrittlement of the samples, a product of the increase in the crystallinity of the polymer during thermal aging (Inácio *et al.*, 2018; Sınmazçelik and Yılmaz, 2007). Consequently, the material tends to fracture at very low deformations (Blanco *et al.*, 2020) during the sliding wear process. Furthermore, as noted above, thermal aging also weakens the bonds between the polymer and the coupling agent (MAPP). Thus, a loss of the reinforcing effect of the bamboo fiber in the PP matrix takes place, making the aged composites less wear-resistant than the unaged ones.

The loss of wear resistance due to aging was more noticeable in the unreinforced PP, which can be attributed to the high brittleness that could have been reached by the PP without fibers due to the increase in crystallinity during thermal aging (Sınmazçelik and Yılmaz, 2007; Blanco *et al.*, 2020; Law *et al.*, 2008), wherein the polymer chains have greater freedom of movement, since there are no bamboo fibers present in their volume, to increase the size of the pre-existing crystals or to form new crystals within the amorphous regions (Blanco-Sánchez *et al.*, 2022). On the other hand, the wear rate of aged and unaged PP/20F was the lowest compared to that of the rest of the composites (Figure 14).

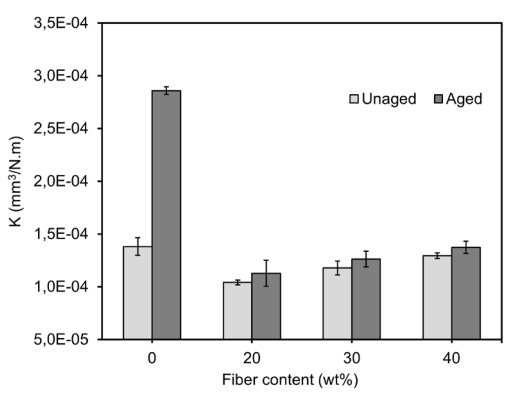
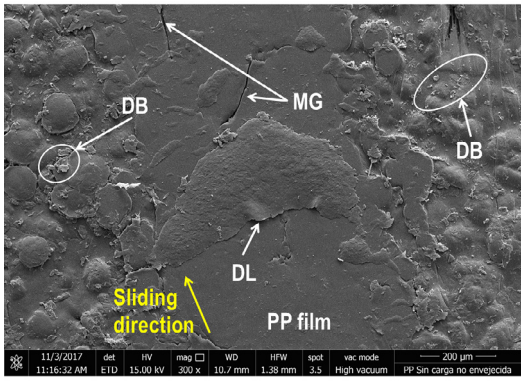


Figure 14. Effect of thermal aging on the wear rate (K) of raw PP and PP/FB biocomposites
Source: Authors

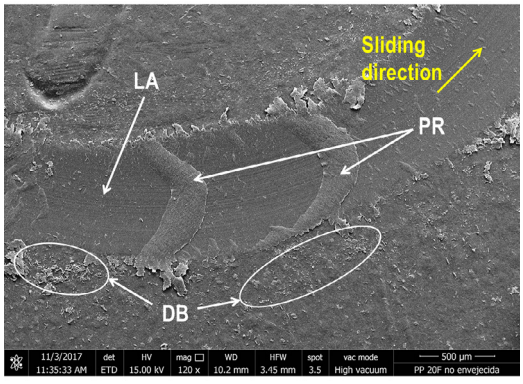
SEM analysis

The tribological behavior of the PP/FB composites can also be explained by analyzing the SEM images of their wear surfaces. Figure 15 presents the micrographs of the wear traces belonging to the pure PP samples and the PP/FB composites without thermal aging. Figure 15a corresponds to the wear surface of the PP without fibers, showing the formation of a film of plastically deformed polymeric material, as well as the presence of micro-cracks, delamination, and polymer residues (debris). In Figures 15c and 15d, corresponding to the wear traces of PP/30F and PP/40F, respectively, a great deterioration of the surface can be seen, which evinces the low wear resistance of

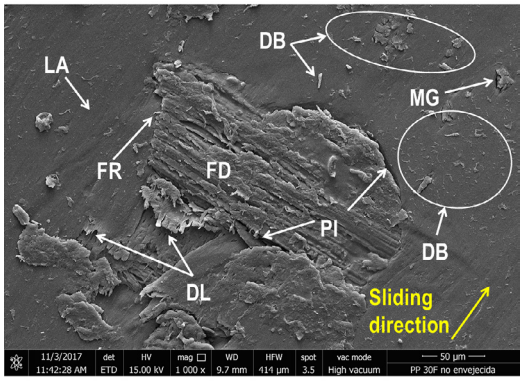
both composites. In these Figures, the four severe wear mechanisms that normally occur in polymeric biocomposites are observed: (i) matrix wear (debris, plastic deformation, plowing, shearing, and cracking); (ii) fiber wear by sliding; (iii) fiber pullout or fracture; and (iv) loss of fiber-matrix interaction (Sınmazçelik and Yılmaz, 2007). Figure 15b shows the polymer film on the wear surface of PP/20F, with the presence of finely divided and loose polymer particles at the edge of the footprint, as produced by matrix wear. The wear surface is relatively smooth, indicating that the wear damage was slight (Bhushan, 2013). However, it was noted that the plastically deformed polymer film on the tread showed uniformly distributed 'prow' marks (Martínez *et al*., 2010) and did not show cracking or appreciable fiber detachment, which constitutes evidence of PP/20F's better wear performance compared to the rest of the composites, in agreement with the results shown in Figure 9.



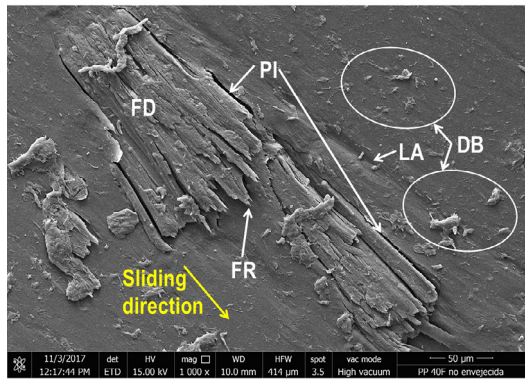
(a)



(b)



(c)

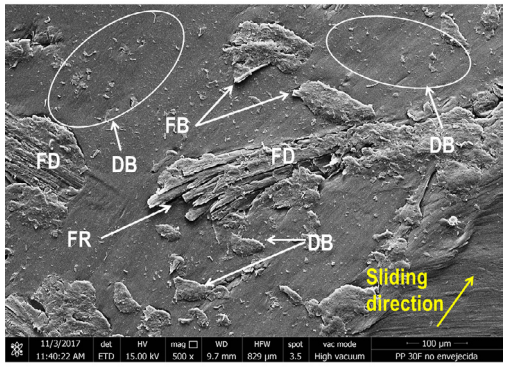


(d)

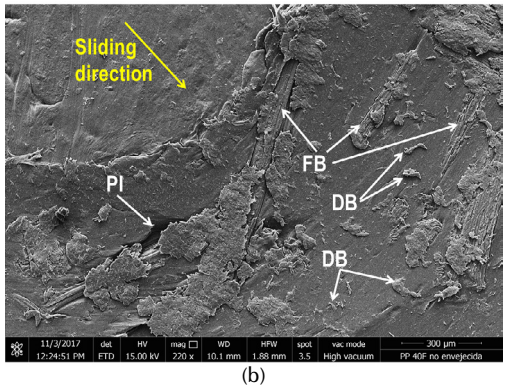
Figure 15. Worn and unaged surfaces of: a) PP without fibers, b) PP/20F, c) PP/30F, and d) PP/40F. Abbreviations: DB: debris; DL: delamination; FD: worn fiber bundle; FR: fiber fracture; MG: microcrack; LA: plow lines; PI: fiber-matrix debonding; PR: prow.
Source: Authors

Further micrographs of the wear traces of the PP/30F and PP/40F composites are shown in Figures 16a and 16b, respectively. In both composites, polymer debris was observed, and many remains of detached or fractured fibers embedded in plastically deformed polymer debris and adhered to the wear surface. Such remains of fibers possibly formed a 'third body' between the contact surfaces (metal-polymer), producing a high shear strength during the sliding wear test, which produced higher friction between the tribological pair and a higher wear of the composites. This behavior corresponds to the results of the friction coefficient and wear rate, as shown in the graphs of Figures 8 and 9, respectively, wherein the PP/30F and PP/40F composites reached the highest values of μ and K compared to the PP/20F composite.

Thin wear particles (debris) in the form of plaque or flakes were observed in Figure 17a, which are commonly found during the sliding of dry and lubricated interfaces. These particles are produced as a result of plowing (abrasive wear), which can also be seen in Figure 17a, due to the rubbing of the asperities of the metal part against the polymer surface. After the surface has been plowed several times, material removal can occur by a low-cycle fatigue mechanism. When plowing occurs, ridges form along the sides of the plowed grooves. These ridges flatten and eventually fracture after repeated loading and unloading cycles. The plowing process also causes plastic deformation of the subsoil and may contribute to the nucleation of surface and subsurface cracks. Additional loading and unloading (high-stress, low-cycle fatigue) causes these cracks, voids, and pre-existing cracks to propagate (subsurface cracks propagate parallel to the surface to some depth) and coalesce with neighboring cracks that eventually shear at the surface and result in thin wear platelets (Bhushan, 2013; Martínez *et al*., 2010).

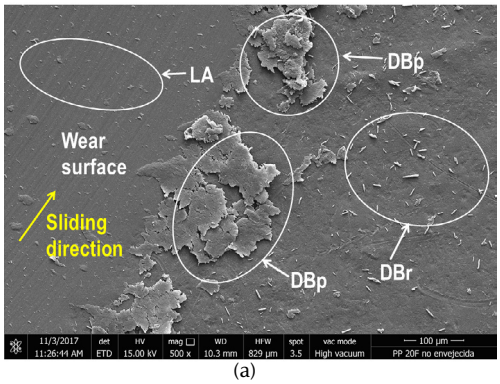


(a)

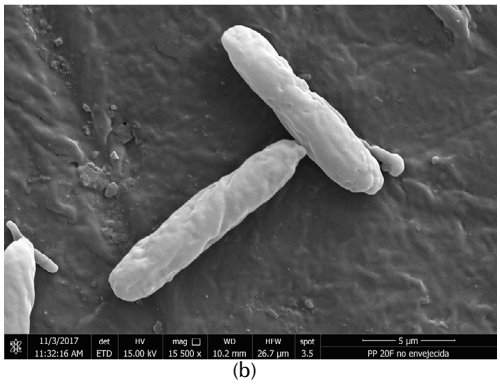


(b)

Figure 16. Worn and unaged surfaces of a) PP/30F and b) PP/40F. Abbreviations: DB: debris; FB: detached fibers; FD: fiber bundle worn; FR: fiber fracture; PI: fiber-matrix unbinding.
Source: Authors



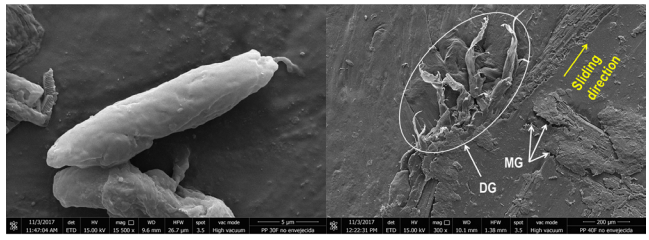
(a)



(b)

Figure 17. Worn surface of unaged PP/20F with the presence of wear debris a) in the form of plates or flakes and b) in the form of rollers. Abbreviations: DBp: plate or flake debris; DBr: roller debris; LA: plow lines.
Source: Authors

A characteristic of fatigue wear is material damage under the repetitive action of compressive, tensile, and shear deformations during the cyclic loading caused by the interaction between the metallic counterpart and the rough surface of the polymer during sliding, leading to the generation and development of cracks, which may be favored by the presence of defects. Some authors modify the term *fatigue wear* to *frictional wear* if the polymer exhibits a low resistance to tearing, causing the formation of roller-like particles at the sliding interface and the tearing of the laminated fragment (Martínez *et al*., 2010). In Figures 17 and 18a, the roller-like wear particles can be seen, in addition to the tearing of the polymer component in Figure 18b. This corroborates that the fatigue or frictional wear mechanism was also present during the tribology tests of the PP/FB composites.

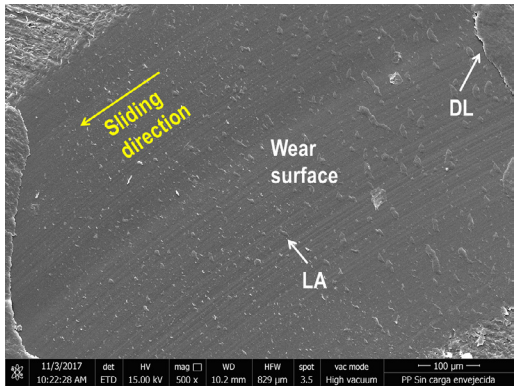


(a) (b)

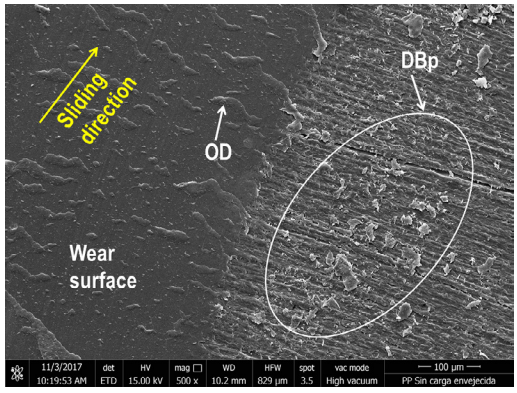
Figure 18. (a) Polymer wear particles (with roller morphology) observed after wear testing of unaged PP/30F composite; (b) Polymer tearing and cracking observed after wear testing of unaged PP/40F composite. Abbreviations: DG: polymer tearing; MG: microcrack.
Source: Authors

In general terms, according to the results obtained for the μ and K parameters (Figures 13 and 14, respectively), the thermal aging process applied in this work caused an increase in the friction of the materials, as well as a loss of the wear resistance of the PP/FB biocomposites. This loss is more significant in raw PP. The worn surface of the aged fiber-free PP showed signs of delamination, in addition to significant evidence of abrasive wear, as many plow lines were observed (Figure 19a), and a large amount of flake-like polymer particles (Figure 19b). Any wear particles generated on the surface will cause further plowing and increase the frictional force, thereby accelerating the delamination process (Martínez *et al*., 2010). The low temperature at which most polymers melt, as well as their low thermal conductivity, ensures that frictional contact temperatures can easily reach the melting point of the polymer, causing its surface to melt, entailing the formation of a 'prow'. The effect of 'prow' formations over the entire wear trace of the polymer is known as fatigue wave formation (Bartenevev and Lavrentev, 1981), which was also observed, very noticeably, on the wear surface of the aged raw PP (Figures 19b and 19c) and the unaged PP/20F (Figure 15b). These waves are associated with the fatigue mechanism, given the cyclic passage of metal counterbody asperities on the polymer surface, as well as with a melt wear mechanism (da Silva *et al*., 2007). All this evidence on the contact surface demonstrates

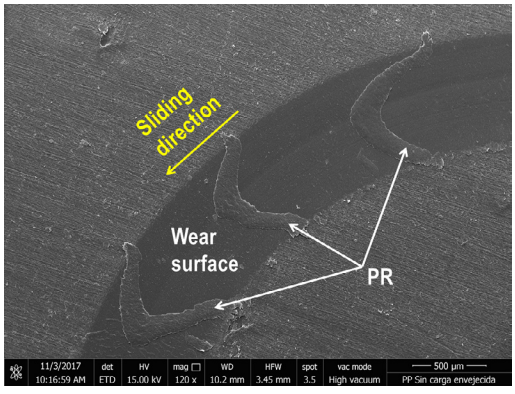
the significant loss of wear resistance of the PP without fibers due to thermal aging. Nevertheless, this loss of wear resistance was found to be lower in the PP/FB composites (Figure 14).



(a)



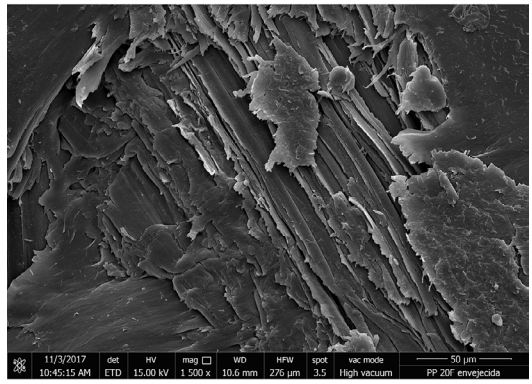
(b)



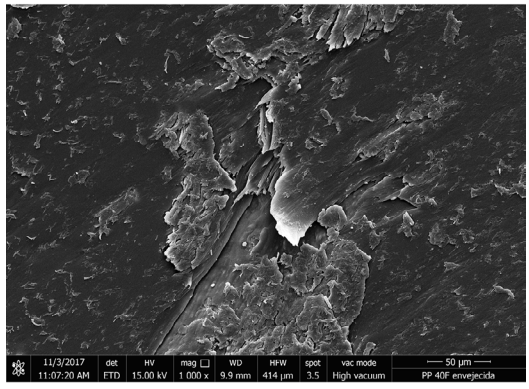
(c)

Figure 19. Worn surfaces of aged PP with the presence of a) plow lines (abrasive wear), b) flake-like polymer debris (adhesive wear), and c) prows. Abbreviations: DL: delamination; DBp: plate or flake debris; LA: plow line; OD: waves; PR: prow.
Source: Authors

The wear mechanism exhibited by the biocomposites (Figure 20) could be related to delamination. This phenomenon occurred due to the stratified nature of the stress state in the Hertzian zone, as a result of the frictional or tangential load that generates a shear stress field on the surface and subsoil. The maximum values of these shear stresses are reached at a certain distance from the surface. Likewise, the stress state generated by the different cyclic loads acting in the contact zone favored the cracks to nucleate below the surface (where shear stresses are maximal) and propagate to the surface. This propagation of cracks and their extension to neighboring cracks and surface shear cause the wear lamellae to delaminate after a large number of asperities pass through each point on the surface of the material (da Silva *et al*., 2007).



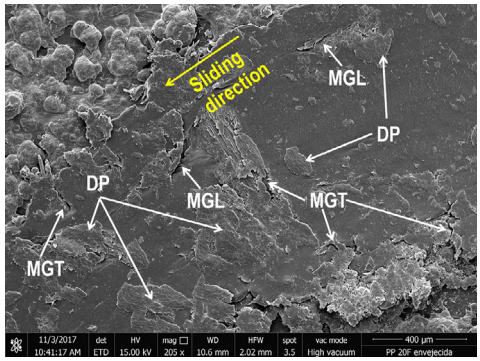
(a)



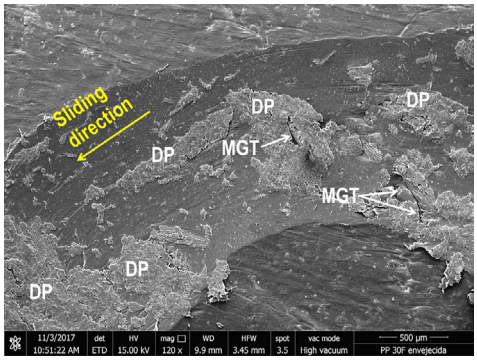
(b)

Figure 20. Wear traces of a) thermally aged PP/20F and b) PP/40F, which show delamination
Source: Authors

In the case of brittle materials with low fracture toughness, wear occurs by brittle fracture. In these cases, the worn zone is characterized by significant cracking (Bhushan, 2013). Figure 21 shows the wear traces of PP/20F and PP/30F composites with the presence of longitudinal and transverse microcracks towards the sliding direction. The transverse microcracks observed on the surface are consistent with a repeated plowing mechanism causing surface fatigue, while the microcracks observed in the wear direction are formed due to micro-cutting produced by the asperities of the metallic counterpart on the polymer surface (Harsha and Tewari, 2003). Polymer tribology studies by authors such as da Silva *et al*. (2007) state that plowing can cause the formation of cracks on the worn surface as a result of the high tensile stresses generated in the contact zone. In addition, the propagation of these cracks results in surface tearing.



(a)

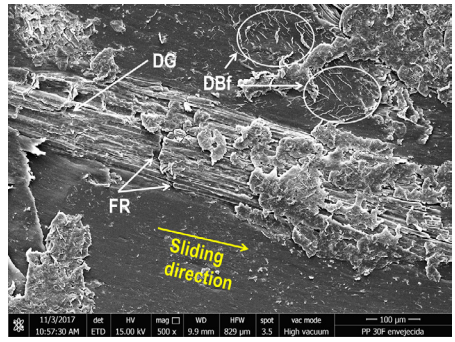


(b)

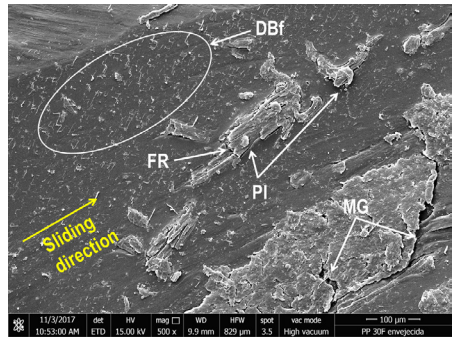
Figure 21. Wear traces of thermally aged a) PP/20F and b) PP/30F, showing cracking. Abbreviations: PD: plastic deformation; MGL: microcrack longitudinal to the sliding direction; MGT: microcrack transverse to the sliding direction.
Source: Authors

In adhesive wear, the localized cold welding of asperities takes place due to the very strong adhesive forces that occur when two surfaces slide against each other. However, as the yield strength of these cold-welded joints is exceeded, these asperities deform plastically, which typically occurs in a softer material when it slides against a harder surface (Narish *et al*., 2011). Figure 21 shows the abraded surfaces of aged PP/20F and PP/30F biocomposites. These surfaces have attached patches of polymeric material, characterized as *plastic deformation*. This surface characteristic is common in the predominant adhesive wear mechanism in all the sliding wear tests performed in this work. However, these surface defects were mostly observed in the wear trace of the aged PP/30F (Figure 21b), generating higher friction between the tribological pair, which is why that composite reached the highest value of μ (Figure 13).

Another reason why thermal aging caused the biocomposites to have lower wear resistance when compared to the samples without aging was the damage caused to the fiber by the action of an abrasive wear mechanism. Figure 22a presents a micrograph showing the abrasive wear of a bundle of bamboo fibers, characterized by the fracture and tearing of several fibers. In addition, the wear debris appears as fine fibrils cut from the polymeric matrix that adheres to the surface (abrasive wear) (Harsha and Tewari, 2003).



(a)



(b)

Figure 22. Wear trace of thermally aged PP/30F showing a) fiber abrasion wear and (b) loss of fiber-matrix interaction. Abbreviations: DG: fiber tearing; DBf: fibril polymer debris; FR: fiber fracture; MG: microcrack; PI: fiber-matrix debonding.
Source: Authors

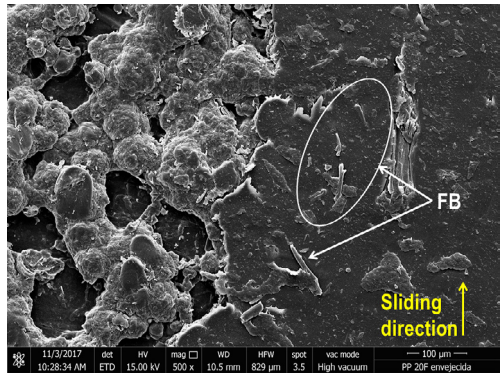


Figure 23. Remnants of bamboo fibers and delamination were observed in the wear trace of thermally aged PP/20F. Abbreviations: FB: detached fiber.
Source: Authors

The prolonged exposure of PP/FB composites to environments with temperatures close to the continuous-use temperature of PP (98 °C) causes a decrease in fiber-matrix interaction, which results in a loss of wear resistance. This decreased interaction is likely attributed to a loss of moisture from the fiber during thermal aging, resulting in micro-embrittlement around the fibers, which reduces the area of the fiber-matrix interface, thus decreasing the shear strength at the interface (Blanco *et al*., 2020). This effect, together with the thermal gradients that occur during sliding wear, given the

non-uniform temperature distribution that creates thermal stresses on the sliding surface, weakens the fiber-matrix adhesion (Figure 22b), which causes bamboo fibers to loosen and become easily sheared as a result of repeated axial thrust during sliding (Yallew *et al*., 2014). This could be the cause for the loosening and fracture of the fibers in the matrix during sliding wear tests, which was also observed in the wear of the thermally aged PP/20F composite (Figure 23) and could have been the reason why this composite exhibited a remarkable increase in its friction coefficient in comparison with the unaged sample (Figure 13).

Conclusions

PP reinforced with 20 wt% of bamboo fibers showed better friction and wear performance, as this percentage of fibers enabled a greater reduction in the coefficient of friction and wear rate of PP (-21.6 and -24.6%, respectively). After thermal aging, the friction coefficient of the PP/20F composite increased by 40.5%, while, for raw PP, PP/30F, and PP/40F, the increase was 2.1, 7.5, and 2.2%, respectively. Likewise, aging caused a loss in the wear resistance of PP and PP/FB composites. This effect was highly noticeable in PP without fibers and less marked in the biocomposites. Greater effectiveness was achieved when the bamboo fiber content was 20 wt%. The use of SEM made it possible to identify the worn surfaces of the raw PP and PP/FB biocomposites. After the tests, some of the wear mechanisms were evidenced, such as prow formation, plowing, cutting, tearing, melting, wave formation, and adhesion. Adhesive, abrasive, and fatigue wear were the dominant ones in the tribology tests of these materials.

CRedit author statement

Eudi Blanco: conceptualization, investigation, methodology, resources, data curation, formal analysis, validation, writing (original draft), visualization, and writing (review and editing). **Jorge Fajardo and Caribay Urbina**: methodology, validation, writing (review and editing), and resources. **Edwuin Carrasquero, Luis López, and Luis Cruz:** validation and writing (review and editing). All authors contributed to the writing of the manuscript and approved its final version for publication.

References

ASTM International (2014). *Standard test method for tensile properties of plastics* (*ASTM D638-14*). ASTM. https://doi.org/10.1520/D0638-14

ASTM International (2017). *Standard test method for wear testing with a pin-on-disk apparatus* (*ASTM G99-17*). ASTM. https://doi.org/10.1520/G0099-17

ASTM International (2019). *Standard test method for density of solid pitch* (*helium pycnometer method*) (*ASTM D4892-14*(*2019*)*e1*). ASTM. https://doi.org/10.1520/D4892-14R19E01

Bajpai, P. K., Singh, I., and Madaan, J. (2012). Frictional and adhesive wear performance of natural fibre reinforced polypropylene composites. *Journal of Engineering Tribology, 227*(4), 385-392. https://doi.org/10.1177/1350650112461868

Bhushan, B. (2013). *Principles and applications of tribology*. John Wiley & Sons Ltd.

Blanco, E., Fajardo, J., Carrasquero, E., Urbina, C., and León, J. B. (2020). Estudio de las propiedades a tensión de un material biocompuesto reforzado con haces de fibras cortas de bambú. *Revista UIS Ingenierías, 19*(3), 163-176. https://doi.org/10.18273/revuin.v19n3-2020016

Blanco-Sánchez, E., Madera-Mujica, A., Pérez-Castillo, M., Fajardo-Seminario, J., Carrasquero-Rodríguez, E., López-López, L., and Cruz-Riaño, L. (2022). Influencia del contenido de fibra y del recocido sobre las propiedades térmicas de un material biocompuesto reforzado con fibras de bambú. *Revista UIS Ingenierías, 21*(2), 39-52. https://doi.org/10.18273/revuin.v21n2-2022004

Bartenevev, G. M., and Lavrentev, V. V. (1981). *Friction and wear of polymers*. Elsevier Scientific Publishing Company.

da Silva, R. L. C., da Silva, C. H., Medeiros, J. T. N. (2007). Is there delamination wear in polyurethane? *Wear, 263*, 974-983. https://doi.org/10.1016/j.wear.2007.01.082

Harsha, A. P., and Tewari, U. S. (2003). Two-body and three-body abrasive wear behaviour of polyaryletherketone composites. *Polymer Testing, 22*(4), 403-418. https://doi.org/10.1016/S0142-9418(02)00121-6

Inácio, A. L. N, Nonato, R. C., and Bonse, B. C. (2018). Mechanical and thermal behavior of aged composites of recycled PP/EPDM/talc reinforced with bamboo fiber. *Polymer Testing, 72*, 357-363. https://doi.org/10.1016/j.polymertesting.2018.10.035

International Organization for Standardization (ISO) (2011). *Determination of the melt mass-flow rate* (*MFR*) *and melt volume-flow rate* (*MVR*) *of thermoplastics - Part 1: Standard method* (*ISO 1133-1:2011*). ISO.

Jariwala, H., and Jain, P. (2019). A review on mechanical behavior of natural fiber reinforced polymer composites and its applications. *Journal of Reinforced Plastics and Composites, 38*(10), 441-453. https://doi.org/10.1177/0731684419828524

Kerni, L., Singh, S., Patnaik, A., and Kumar, N. (2020). A review on natural fiber reinforced composites. *Materials Today: Proceedings, 28*(3), 1616-1621. https://doi.org/10.1016/j.matpr.2020.04.851

Law, A., Simon, L., and Lee-Sullivan, P. (2008). Effects of thermal aging on isotactic polypropylene crystallinity. *Polymer Engineering and Science, 48*, 627-633. https://doi.org/10.1002/pen.20987

Martínez, F. J., Canales, M., Bielsa, J. M., and Jiménez, M. A. (2010). Relationship between wear rate and mechanical fatigue in sliding TPU–metal contacts. *Wear, 268*, 388-398. https://doi.org/10.1016/j.wear.2009.08.026

Mimaroglu, A., Unal, H., and Yetgin, S. H. (2018). Tribological properties of nanoclay reinforced polyamide-6/polypropylene blend. *Macromolecular Symposia, 379*, 1700022. https://doi.org/10.1002/masy.201700022

Narish, S., Yousif, B. F., and Rilling, D. (2011). Adhesive wear of thermoplastic composite based on kenaf fibres. *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology 225*(2), 101-109. https://doi.org/10.1177/2041305X10394053

Nirmal, U., Hashim, J., and Low, K. O. (2012). Adhesive wear and frictional performance of bamboo fibres reinforced epoxy composite. *Tribology International, 47*, 122-133. https://doi.org/10.1016/j.triboint.2011.10.012

Singh, T., Kumar, N., Ashok Raj, J., Grewal, J. S., Patnaik, A., and Fekete, G. (2019). Natural fiber reinforced non-asbestos brake friction composites: influence of ramie fiber on physico-mechanical and tribological properties. *Materials Research Express, 6*, 115701. https://doi.org/10.1088/2053-1591/ab45a4

Sınmazçelik, T., and Yılmaz, T. (2007). Thermal aging effects on mechanical and tribological performance of PEEK and short fiber reinforced PEEK composites. *Materials and Design, 28*(2), 641-648. https://doi.org/10.1016/j.matdes.2005.07.007

Yallew, T. B., Kumar, P., and Singh, I. (2014). Sliding wear properties of jute fabric reinforced polypropylene composites. *Procedia Engineering 97*, 402-411. https://doi.org/10.1016/j.proeng.2014.12.264

Yousif, B. F., and El-Tayeb, N. S. M.; (2007). Tribological evaluations of polyester composites considering three orientations of CSM glass fibres using BOR machine. *Applied Composite Materials, 14*(2), 105-116. https://doi.org/10.1007/s10443-007-9034-2

Research Article/System and Computer Engineering http://doi.org/10.15446/ing.investig.106069

Facial Cryptograms Classification through their Local Texture Features

Clasificación de criptogramas faciales a través de sus características de textura local

José T. Guillen-Bonilla[1], Jorge Aguilar-Santiago[2], Juan C. Estrada-Gutiérrez[3], and Maricela Jiménez-Rodríguez[4]

ABSTRACT

With the increasing use of social networks, unauthorized individuals have become able to detect or intercept personal data, which could be used improperly, thereby causing personal damage. Therefore, it is essential to utilize a security mechanism that aids in protecting information from malicious attacks. In this work, facial recognition is proposed, using the local textural features of cryptograms. Red-Green-Blue (RGB) facial images were encrypted by applying the mathematical Logistic Map model, which generated a cryptogram. These cryptogram's local textural features were extracted via Coordinated Cluster Representation (CCR) transform. The high classification efficiency (97-100%) of the encrypted facial images was experimentally validated using two databases: the first one was generated by controlling parameters such as rotation, scale, and lighting; and the second one is a public database. This technique is suitable for a wide range of applications related to user authentication, and it safeguards the identity of authorized users when accompanied by additional layers of security involving images of interest, such as those employed by the medical field, enhancing the security of users whose diseases are graphically studied in hospitals. In addition, this technique can be deployed to protect new product launches where images are important, such as clothing, footwear, mosaics, etc., since one does not need to decrypt the images in order to classify them.

Keywords: facial recognition, classification efficiency, chaos, cryptography, coordinated clusters representation, local textural features

RESUMEN

Con el uso creciente de las redes sociales, personas no autorizadas han conseguido detectar o interceptar datos personales, que podrían utilizarse de manera inapropiada, causando así daños personales. Por lo tanto, es esencial utilizar un mecanismo de seguridad que ayude a proteger la información de ataques maliciosos. En este trabajo se propone el reconocimiento facial, utilizando las características texturales locales de los criptogramas. Se cifraron imágenes faciales en formato Red-Green-Blue (RGB) aplicando el modelo matemático de Mapa Logístico, lo que generó un criptograma. Las características texturales locales de estos criptogramas se extrajeron mediante la transformación de representación de *cluster* coordinado (CCR). La alta eficiencia de clasificación (97-100%) de las imágenes faciales cifradas fue validada experimentalmente utilizando dos bases de datos: la primera fue generada controlando parámetros como la rotación, escala e iluminación; y la segunda es una base de datos pública. Esta técnica es adecuada para una amplia gama de aplicaciones relacionadas con la autenticación de usuarios, y protege la identidad de los usuarios autorizados cuando se acompaña de capas adicionales de seguridad que involucran imágenes de interés, como las utilizadas en el campo médico, mejorando la seguridad de los usuarios cuyas enfermedades se estudian gráficamente en los hospitales. Además, esta técnica puede desplegarse para proteger lanzamientos de nuevos productos donde las imágenes son importantes, como ropa, calzado, mosaicos, etc., ya que no es necesario descifrar las imágenes para clasificarlas.

Palabras clave: reconocimiento facial, eficiencia de la clasificación, caos, criptografía, representación de cluster coordinado, características texturales locales

Received: November 29th, 2022
Accepted: February 20th, 2024

Introduction

Currently, facial recognition is a very active area of research worldwide, given its large number of applications. Among these, the following can be mentioned: access to the control of mobile devices and computers, video surveillance, and access to facilities, among others (Kalech, 2019; Bello-Cerezo et al., 2019; Peng et al., 2019). However, all of these applications are prone to information theft (hacking) (fliphtml5.com, n.d.). Therefore, it is advisable to encrypt facial data, avoiding possible harm to people due

[1] Doctor of Sciences, Centro de Investigaciones en Óptica. Affiliation: full professor, Centro Universitario de Ciencias Exactas e Ingenierías, Universidad de Guadalajara, México. E-mail: trinidad.guillen@academicos.udg.mx

[2] Student of the Doctorate in Technology Sciences, Centro Universitario de la Ciénega. Affiliation: subject professor, Centro Universitario de la Ciénega, Universidad de Guadalajara, México. E-mail: jorge.asantiago@academicos.udg.mx

[3] Doctor of Sciences, Universidad de Guadalajara. Affiliation: full professor, Centro Universitario de la Ciénega, Universidad de Guadalajara, México. E-mail: jcarlos.estrada@academicos.udg.mx

[4] Doctor of Sciences and Technology, Universidad de Guadalajara. Affiliation: full professor, Centro Universitario de la Ciénega, Universidad de Guadalajara, México. Email: maricela.jrodriguez@academicos.udg.mx

to inadequate information handling. A facial image can be encrypted by applying a mathematical model governed by a nonlinear dynamic equation whose behavior is complex-chaotic when its control parameters are adequate. Commonly, it would be necessary to decrypt these images when aiming for facial recognition. To prevent this, it should be possible to apply a recognition technique with the cryptograms generated during the encoding process by using a descriptor. This paper proposes the combination of chaotic cryptography, a texture descriptor, and a multi-class classifier to perform facial recognition through cryptograms.

The remainder of this paper is divided into the following sections. In the *Literature review*, several related works are detailed. The section titled *Classification and encryption system* outlines the steps to encrypt and classify images. The *Results* section presents the experimental tests conducted and the results obtained, and the final section provides the conclusions reached in this research.

Literature review

The literature has proposed various dynamic nonlinear mathematical models in studying and applying chaos theory to information encryption. Some examples include the Lorenz attractor, the Rössler attractor (Kumar and Girdhar, 2021; Rodríguez *et al*., 2016; Jiménez-Rodríguez *et al*., 2018), the Henon attractor (Afifi, 2019), the Logistic Map, and the Sine Map (Xiang and Liu, 2020; Pan *et al*., 2018; Suman *et al*., 2022). This is due to the fact that, because of their properties, chaotic systems are closely related to cryptography. These properties include a high sensitivity to initial conditions and system parameters that are utilized as encryption keys, making it more difficult for an attacker to decode encrypted data and thereby increasing the security of the information that travels through the network. This helps to prevent malicious users from corrupting the integrity of said information and ensure that it can only be retrieved by those who have the right keys.

The Logistic Map approach is based on a mathematical function in parabolic form (a degree-two polynomial). This model can exhibit very wide dynamics by merely varying the value of a parameter, since there may be trajectories that are periodic or chaotic, tending towards a fixed point. Although the Logistic Map was first applied in studying the evolution of populations, it has been efficiently employed in the encryption of digital images (Pan *et al*., 2018). Chaotic systems have been widely utilized in ciphering methods, as is the case of the research presented herein, which proposes the asynchronous Boolean network encryption technique, implementing a new two-dimensional (2D) chaotic system with a codification rule (Gao *et al*., 2023a). Furthermore, a ciphering method based on three dimensions (3D) was developed in order to implement 2D-chaotic systems while employing Logistic Maps (Gao *et al*., 2023c). Another facial encryption technique deploys a HASH key generator and a 2D-logistic tent modular map. For facial recognition, the Histogram of Oriented Gradients (HOG) has also been applied (Gao *et al*., 2023b). Another work proposed an audio encryption system that combines a one-dimensional (1D)-infinite collapse map (1D-ICM) and a Logistic Map (Wu *et al*., 2022).

A generated cryptogram is a random field that contains the information of an original image, together with a series of pseudo-random numbers. Furthermore, a cryptogram contains texture information that can be extracted by applying a specialized statistical technique. In this type of approach (Alaei *et al*., 2019; Nanni *et al*., 2019), the digital image is regarded as a source of detectable patterns, with an observation window of size I×J (commonly I×J=3×3 pixels) (Leyferman *et al*., 2023). For each position, a texture unit is calculated and assigned to a function-of- probability distribution in terms of texture units. Such a probability function is interpreted as a texture spectrum and is used as a multi-dimensional characteristic vector in classifiers.

Multiple works have employed the textural characteristics of images in order to perform biometric recognition (*e.g.*, visual cryptography) and to distribute facial images in separate databases (Ren and Zhang, 2022). A facial recognition and authentication system was developed by Ibrahim *et al*. (2021). A visual authentication and facial recognition method in color was also proposed in the literature (Ibrahim *et al*., 2019), wherein the image is encrypted and divided into two shared segments, with one stored in the memory card and the other one in a database. When the sensor reads the card, the images are overlaid to uncover the original.

Visual cryptography has been employed in facial recognition. Here, photographs are divided into two images that are stored in different servers. Both of these images must be obtained during the recovery process (Mohan and R., 2021). A secure communication system was implemented by Aguilar Santiago *et al*. (2020), employing OpenCV to perform facial detection and recognition. Ahmad Khan *et al*. (2021) developed a facial recognition technique that leverages the homomorphic properties of cryptosystems and the Euclidean distance. The novelty of this technique lies in not revealing the real image during the recognition process.

In this work, the cryptogram generated with the mathematical Logistic Map model is interpreted as a source of texture information, upon which facial recognition is performed through local texture features. The texture is extracted by applying the coordinated clusters representation (CCR) transform, which conducts a local autocorrelation analysis over a binary image. This approach is based on two mathematical theorems (Sánchez-Yáñez *et al*., 2003). The texture spectrum generated with the CCR transform is utilized as a characteristic vector in a multi-class classifier based on the statistics of the image. In most research aiming for recognition through visual cryptography or authentication, different devices have been employed to enhance security (Ren and Zhang, 2022; Ibrahim *et al*., 2019, 2021; Mohan and R., 2021) or to perform recognition prior

to encryption. Conversely, the technique proposed in this paper does not require a distributed storage system; hence, time and resources are spared, since the images must not be decrypted for facial recognition.

This paper offers the following contributions and applications:

- Classification efficiency is increased by employing cryptograms while environmental conditions (such as rotation, scale, and lighting) remain invariant.
- The proposed recognition technique works directly with the stored encrypted photographs, thus avoiding the need for an image decryption algorithm.
- The security in storing or sending the ciphered data is enhanced.
- The computational cost is decreased in scenarios where security and recognition are required.
- This approach prevents the storage of image data across different devices as a security reinforcement.
- This method allows sharing encrypted images in social media while preventing non-authorized users from visualizing them.
- The use of secure databases containing underage user information is enabled, blocking the visualization or the tracking of their data.
- The databases can be stored in cloud services, preventing third parties from accessing and misusing the data.
- This method enables real-time facial identification, employing the cryptogram corresponding to the detected face and safeguarding user privacy.

Classification and encryption system

Figure 1 presents the proposed multi-class classifier system based on the coordinated cluster representation (CCR) transform and the encrypted facial images. In this classifier, RGB images are encrypted, cryptograms are binarized, binary images are represented as a texture spectrum (obtained using the CCR transform), the Hamming distance is employed to measure the similarity between a test image and a database (of images) previously identified by a human expert, and the facial image is finally classified. The classifier requires two phases (Sánchez-Yáñez *et al.*, 2003): learning and recognition.

In the learning phase, the classifier is trained within a five-step process:

1. A database of RGB facial images is generated, or a public database is employed. Each facial image is encrypted by applying the mathematical Logistic Map model.
2. All color cryptograms are converted into gray-level ones using the MATLAB software. Subsequently, these cryptograms are binarized with the fuzzy c-means algorithm (Dunn, 1973).
3. From each binary cryptogram obtained in step 2, Q random sub-cryptograms are extracted, with a size of ¼ of the original area. Their local textural features are extracted using the CCR transform. Thus, each sub-cryptogram is represented by a texture spectrum $F_{I\times J}^{q,m}(b)$, where q indicates the sub-cryptogram number, m indicates the class $(m=1,2,\ldots,M)$, and M is the total number of classes.
4. The prototype vector of the *m-th* cryptogram $F_{I\times J}^{m}(b)$ is determined using the histograms $F_{I\times J}^{q,m}(b)$.
5. Steps 1 and 4 are repeated for each class (cryptogram).

In the recognition phase, all facial cryptograms are classified. This phase consists of seven steps:

1. A facial image is acquired using a digital camera.
2. This test image is encrypted via the Logistic Map.
3. The resulting RGB cryptogram is converted into a gray-level one in MATLAB, and it is transformed into a binary facial cryptogram utilizing the fuzzy c-means algorithm.
4. P test random cryptograms, similar in size to those of the learning stage, are extracted from the facial cryptogram, and their local textural features are extracted from each sub-cryptogram, such that each sub-cryptogram is represented as a texture spectrum $F_{I\times J}^{Test,p}(b)$, where p $(p=1,2,\ldots,P)$ indicates the *p-th* sub-cryptogram.
5. Using the Hamming distance, the similarity between $F_{I\times J}^{Test,p}$ and $F_{I\times J}^{m}(b)$ is measured (in the CCR feature space). Subsequently, the test sub-cryptogram is assigned to class m if and only if the Hamming distance is minimal.
6. The cryptogram is decoded.
7. Steps 1-6 are repeated for each test cryptogram.

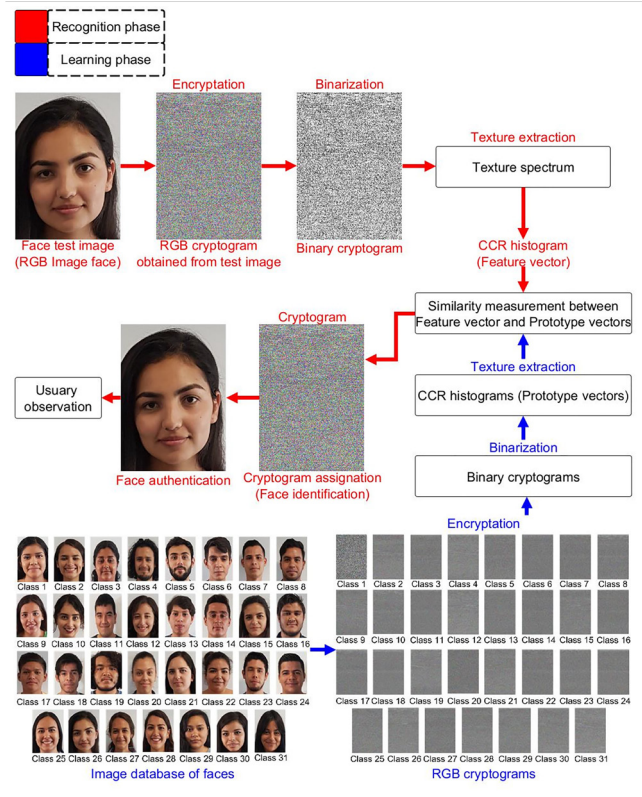


Figure 1. Proposed multi-class classifier system for facial identification
Source: Authors

Encryption

The mathematical Logistic Map model is governed by a nonlinear dynamic equation whose behavior is complex-chaotic (Xiang and Liu, 2020):

$$x_{n+1} = bx_n\left(1-x_n\right) \quad (1)$$

where b is a control parameter $\left(0 \le b \le 4\right)$, x_0 is the initial condition of the chaotic map $\left(0 \le x_0 \le 1\right)$, and x_{n+1} denotes the subsequent values obtained in the n^{ava} iteration of x_0. This system exhibits a chaotic behavior when the control parameter b takes a value within the interval $b \in [3.57, 4]$. In Jiménez-Rodríguez et al. (2015), we proposed an algorithm to generate a cryptogram from a color image. In this work, we adapted said algorithm for our application. The algorithm works with the RGB subpixels (red, green, blue) that make up each pixel of the image. These have an integer value between 0 and 255. The encryption keys are $b1$, the $x1_0$ parameter, and the initial condition. These keys are used to obtain orbit 1, which is applied in the diffusion process. Moreover, $b2$ and $x2_0$ correspond to a parameter and an initial condition that allow generating orbit 2, used in the confusion technique.

Encoding algorithm

The encoding process is presented below.

1. Store the subpixels in a vector called SP, which has a length $long = SP_n^R + SP_n^G + SP_n^B$, where $SP = \left[SP_1^R, SP_1^G, SP_1^B, \ldots, SP_n^R, SP_n^G, SP_n^B\right]$.
2. Divide each element of SP by 255 to obtain values between 0 and 1: $SP = \frac{\left[SP_1^R, SP_1^G, SP_1^B, \ldots, SP_n^R, SP_n^G, SP_n^B\right]}{255}$.
3. Create a vector named mix with length $long$ to store SP values chaotically.
4. Solve Equation (1) using the encryption keys $b1$ and $x1_0$, with each value of the orbit stored in $orb_{log} = [0.3452, 0.8970, 0.5673, \ldots, 0.6787]$.
5. Generate a position between 1 and $long$ $pos = round\left(long * orb_{log}[i]\right)$.
6. Verify whether the pos position in the mix vector is empty. Then, store an SP value; otherwise, save the subpixel in a vector called loc.
7. Repeat steps 5 and 6.
8. Proceed through the mix vector and, in each empty location, store a value of loc.
9. Solve Equation (1) using keys $b2$ and $x2_0$ to generate $rblog2 = [0.6958, 0.7968, 0.9854, \ldots, 0.8456]$, a vector with $long$ chaotic values.
10. Add the values of vectors $orblog2$ and mix in an orderly fashion. Store the result in a vector denominated $cipher$, $cipher = orblog2 + mix$.

Figure 2a displays an RGB facial image, and Figure 2b shows its cryptogram (steps 1-10).

(a) (b)

Figure 2. a) RGB facial image; b) RGB cryptogram obtained by applying the Logistic Map
Source: Authors

As shown in Figure 2, the cipher algorithm efficiently encrypts the facial information, as the cryptogram is completely chaotic and the face is not recognizable.

Coordinated cluster representation

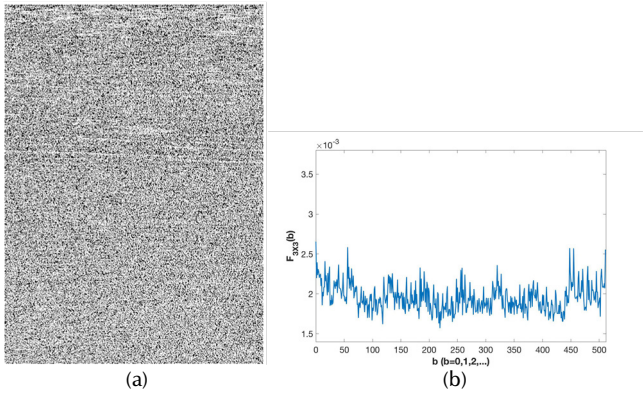
Considering that the cryptogram is a random field (Figure 2b), and that the color channels retain the information of the original image, the RGB color cryptogram can be transformed into a gray-level one, losing only its color information. If the gray-level cryptogram is transformed by applying a binarization algorithm, its lighting information is eliminated, but the binary cryptogram retains sufficient information for its characterization through local textural features. This is possible because the binary cryptogram is a 2D random field that can be interpreted as a source of texture information. In this work, CCR is used to extract local textural features. To extract the texture of the facial cryptograms, the following steps are required:

1. The facial image is encrypted with the mathematical Logistic Map model. Thus, an RGB color cryptogram is obtained.
2. The RGB cryptogram is transformed into a gray- level one.
3. The gray-level cryptogram is transformed into a binary one. This binary cryptogram is interpreted as a source of texture information.
4. An observation window of $I \times J$ is established.
5. The observation window is moved pixel by pixel over the entire binary cryptogram. For each position, a texture unit is calculated, which is a discrete random variable b, whose value depends on the binary state within the observation window.

6. The texture unit is used as an index in the discrete histogram $H_{I\times J}(b)$, whose length is $2^{I\times J}$.
7. The histogram $H_{I\times J}(b)$ is divided by the total of binary patterns, obtaining a function-of-probability distribution $F_{I\times J}(b)$ in terms of texture units. This is a discrete equalized histogram $F_{I\times J}(b)$, and it is interpreted as a texture spectrum.

By applying the described algorithm (steps 1-7), the texture spectrum of an RGB facial image can be calculated. Figure 3a shows the binary facial cryptogram calculated from the RGB one (Figure 2b), and Figure 3b presents its texture spectrum $F_{3\times3}(b)$, where $I\times J=3\times3$ pixels.

In Figure 3b, the texture spectrum fully represents the cryptogram (Figure 3a). This texture spectrum has two important characteristics: a) the histogram occupies a low-dimensional space $\left(R^{512}\ dimension\right)$, and b) it can be used as a characteristic vector in texture classifiers.



(a) (b)

Figure 3. a) Binary facial cryptogram obtained using MATLAB and fuzzy c-means, b) texture spectrum F(3×3) calculated with an observation window of I×J=3×3 pixels
Source: Authors

Decoding algorithm

As shown in Figure 1, decoding is the last stage of the multi-class classifier system. The cryptogram (Figure 2b) preserves the information of the original image. This information can be recovered by following these steps (Jiménez-Rodríguez et al., 2015):

1. Using keys $b2$ a $x2_0$, generate $orblog2=[0.6958,0.7968,0.9854,\ldots,0.8456]$, a vector with $long$ chaotic values.
2. Perform the reverse process of step 10 in the encoding algorithm to encrypt $mix=cipher-orblog2$.
3. Regenerate orb_{log} as in step 4 of Encoding algorithm.
4. Calculate pos performing step 5 of the encoding algorithm.
5. Verify whether the location pos in the vector mix is not empty. Then, take the value and store it in the next position of the vector, called $original$. Otherwise, skip a location in the *original* vector.
6. Repeat steps 4 and 5 for each value of orb_{log}.
7. Proceed through the mix vector, and store each element in the empty positions of $original$ in an orderly fashion.
8. Multiply each original value by 255 and round out the result – in this way, the integers corresponding to the subpixels are retrieved.

Figure 4 presents the facial image retrieved from the cryptogram shown in Figure 2b, by following steps 1-8.

Figure 4. Facial image retrieved from the cryptogram
Source: Authors

Note that the decryption algorithm recovers the original image from the cryptogram, demonstrating the effectiveness of our proposal in retrieving information. This confirms that a facial image can be classified in the image space or in that of the cryptogram. Hence, the result must be the same.

Results

Database of facial images

To perform the corresponding experiments, two databases of facial images were used. The first database was generated by taking photographs of 31 students from La Ciénega University Center. During the capture process, the scale, rotation, lighting, and resolution of the camera were controlled. This simplified our facial classification issues, but also limited the potential applications. All facial images were in RGB color, their size was 480 x 576 pixels, and each facial image represented a mood: serious, smiling, happy, and merry. Each image was identified as a class. The classes can be seen in Figure 5.

The second database was the FEI face database (the Brazilian facial emotion identification database). This is a public database created by the Artificial Intelligence Laboratory of

the FEI in São Bernardo do Campo, São Paulo, Brazil (Carlos Eduardo Thomaz–Personal Web Page, n.d.). We used the FEI database's facial images of 200 photographed individuals. All images were in RGB color and had the same size (260 x 360 pixels), and their background was white. Figure 6 shows 100 facial images taken from the FEI face database.

Encryption and decryption

The mathematical Logistic Map model was applied, as well as the encryption algorithm, with the encryption keys $b1 = 3.87756$, $x1_0 = 0.34342$, $b2 = 3.77$, $x2_0 = 0.8990$. Both of the databases used were encrypted, obtaining RGB cryptograms. The cryptograms and the original facial images were identical in size (480 x 576 and 260 x 360 pixels). Figure 7 presents the RGB cryptograms obtained from the first database (Figure 5). Note that these cryptograms are chaotic fields, and that the facial information is unrecognizable. It can thus be stated that the encryption process efficiently protects user information.

By applying the decryption algorithm described in the *Decoding algorithm* section, the RGB cryptograms were decrypted, once again obtaining the facial images shown in Figures 5 and 6. Subsequently, via a correlation analysis, the similarity between the original and the decrypted images was compared, with a correlation coefficient always equal to one. This proved that each original image can be efficiently recovered from its cryptogram.

Classification efficiency

Using the previously described classifier system, the facial image cryptograms were classified. Four experiments were conducted, classifying the sub-cryptograms into their corresponding classes. The CCR histogram was utilized as a multi-dimensional characteristic vector, the facial images were encrypted with the mathematical Logistic Map model, and Q and P finally took a value of 80. In the first and second experiments, the generated database (Figure 5) was classified while calculating the CCR histogram with observation windows of $I \times J = 3 \times 3$ and $I \times J = 4 \times 4$ pixels. In the third and fourth experiments, the public database (Figure 6) was once again classified with the same CCR histogram size. The experimental results were in the form of a confusion matrix, where the guesses are in the main diagonal and the identification errors correspond to all elements outside it. In this vein, the efficiency (%) is given by Equation (2).

$$Ef\left(\%\right) = \frac{\sum_m diag\left(A\right)}{\sum_{mm} A} \times 100 \qquad (2)$$

where A is the confusion matrix, $diag\left(A\right)$ indicates the main diagonal of said matrix, and $Ef\left(\%\right)$ is the percent classification efficiency of the cryptograms. Figure 7 presents our experimental results regarding sub-cryptogram classification.

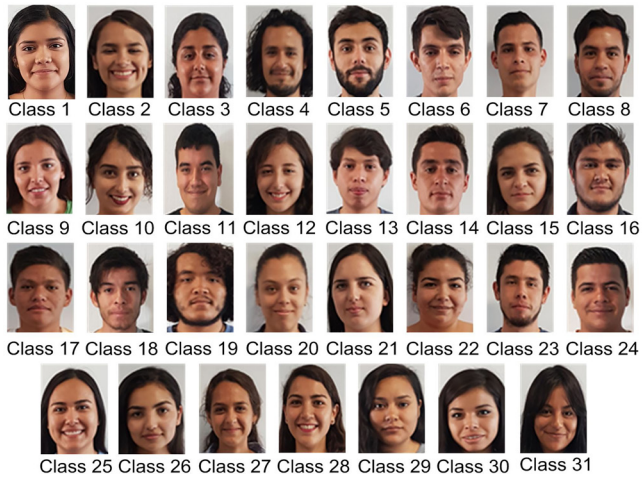


Figure 5. Database of facial images generated from La Ciénega University Center
Source: Authors

Figure 6. Public database used (FEI face database)
Source: (Carlos Eduardo Thomaz - Personal Web Page, s. f.)

As seen in Figure 8, the classification efficiency is very high: it is 100% when the size of the observation window is 4 x 4 pixels (Figures 8b and 8d), and it reaches 97% when the size is 3 x 3 pixels (Figures 8a and 8c). These results are the same for the two databases. In the first case, the efficiency is 100% because the CCR transform extracts

sufficient texture information from the cryptograms, and the classifier considers sufficient characteristics. In the second case, the efficiency is 97% because the CCR transform, and the classifiers do not extract sufficient information from the cryptogram. This reduces the values obtained, which are high for both databases. The results agree with those of other studies in the literature that employ the CCR histogram as a multi-dimensional characteristic vector in classifier systems (Kurmyshev *et al*., 2007; Kurmyshev and Sánchez-Yáñez, 2005; Guillen-Bonilla *et al*., 2007).

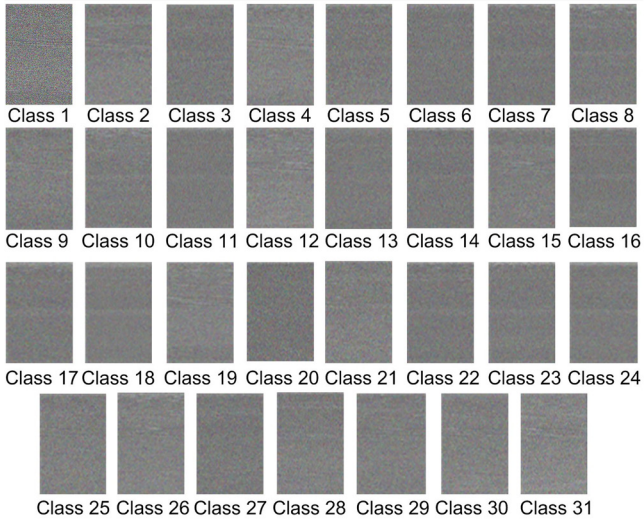


Figure 7. Cryptograms generated from the facial images in Figure 5
Source: Authors

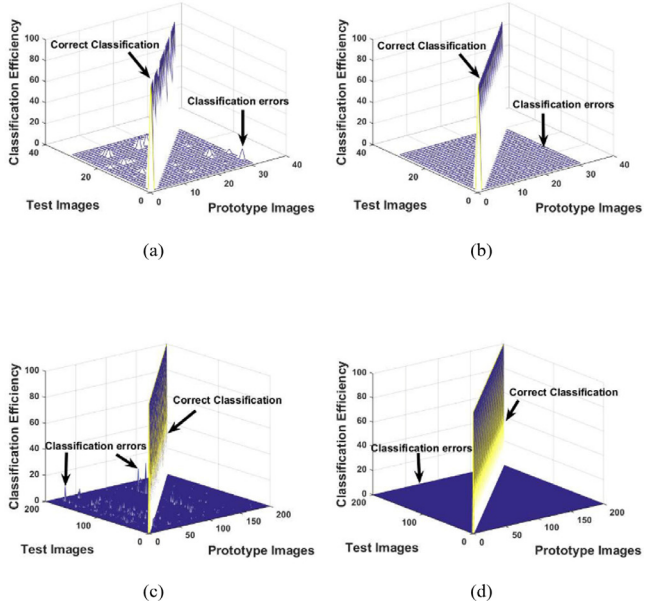


Figure 8. Experimental results regarding cryptogram classification efficiency: a) I x J = 3 x 3 (generated database); b) I x J = 4 x 4 (generated database); c) I x J = 3 x 3 (public database); d) I x J = 4 x 4 (public database)
Source: Authors

Comparison

Diverse texture extraction techniques have been reported in the literature and applied in facial recognition. Each technique has its own classification efficiency, which depends on the textural features extracted from the digital images and the classifier system. Nevertheless, when comparing CCR against other techniques (Table 1), our proposal shows a higher classification than the DLFace, MDLFR, CRC, H-CRC, LPP, LBP, and CS-LGC methods, demonstrating its promise for facial-image classification. Another important consideration is that CCR classified cryptograms (facial images encrypted using the mathematical Logistic Map model), while other texture-extraction techniques classified the actual photographs in the image domain.

Table 1. Comparison between CCR and other texture extraction techniques applied in facial recognition

Classification efficiency (%)		
Operators	*Minimum*	*Maximum*
DLFace (Peng *et al*., 2019)	86.12	98.68
MDLFR (Luo *et al*., 2019)	82.19	95.81
CRC (Zhang *et al*., 2014)	84.16	92.73
H-CRC (Shi *et al*., 2019)	95.41	99.17
LPP (Bansal *et al*., 2014)	80.93	94.33
LBP (Yang *et al*., 2019)	36.11	96.47
CS-LGC (Yang *et al*., 2019)	39.51	98.33
CCR (our proposal)	96.39	100

Note: DLFace: deep local descriptor for cross-modality face recognition; MDLFR: multi-resolution dictionary learning for face recognition; CRC: collaborative representation-based classification; H-CRC: histogram-based CRC; LPP: locality preserving projection; LBP: local binary pattern; CS-LGC: central symmetric local gradient coding; CCR: coordinated cluster representation.
Source: Authors

In this work, we performed facial identification using the local textural features extracted from binary cryptograms. In our methodology (Figure 1), facial images are encrypted by applying the mathematical Logistic Map model, the CCR transform is used in textural feature extraction from the facial cryptogram, and a multi-class classifier is applied for cryptogram classification, wherein the CCR histogram is used as a multi-dimensional characteristic vector. During the classification of cryptograms, color and lighting information are eliminated via digital image processing techniques. However, CCR boasts a high classification efficiency because the binary facial cryptogram preserves sufficient texture information to be characterized through CCR (Sánchez-Yáñez *et al*., 2003; Dunn, 1973; Jiménez-Rodríguez et al., 2015; Carlos Eduardo Thomaz–Personal Web Page, n.d.; Kurmyshev *et al*., 2007; Kurmyshev and Cervantes, 1996), obtaining an classification accuracy in the order of 96.39-100% (Table 1).

Four numerical experiments were conducted on two facial image databases, employing observation window sizes of 3 x 3 and 4 x 4 pixels. These experiments confirmed the high

cryptogram classification accuracy of our proposal (Figure 1). When the CCR histogram was calculated with a 3 x 3 pixel window size, the resulting classification efficiencies are 97.70% (generated database) and 96.39% (public database). If the window size is 4 x 4 pixels, both efficiencies are 100%. This demonstrates that facial cryptograms can be classified through their local textural features or in the CCR feature space, and the obtained results agree with others reported in the literature (Sánchez-Yáñez *et al*., 2003; Kurmyshev *et al*., 2007; Guillen-Bonilla *et al*., 2007).

Based on Table 1, combining the CCR transform and the encryption algorithm with the mathematical Logistic Map model offers two very important advantages. First, our proposal provides computer security, as chaotic mathematical models are used during the encryption of facial information, implying greater security for users. Second, the classification efficiency is high under controlled conditions (*i.e.*, rotation, scale, lighting, and resolution). If the conditions are not controlled, the efficiency can be reduced due to other physical variables.

The encryption system can be deployed to cipher any type of data. Once this data type is divided into blocks of bytes, it can be applied to a computing system for real-time encryption. Table 2 provides the ciphering times (s) for different image sizes.

Table 2. Encryption times in seconds

Image size	Time (s)
128 x 128	0.1368629
256 x 256	0.2126618
512 x 512	0.3767965
1024 x 1024	0.9764167

Source: Authors

Our proposal has potential applications in computer security because facial information is encrypted with chaotic mathematical models. This user-directed data protection scheme prevents any unintended use of personal information and increases data security during the information transfer through social networks and/or media. Therefore, future work shall consider the following directions: the development of security schemes in portable systems and the attainment of increased security in information transfer systems and social networks. In addition, we will consider making the classification system more robust by employing variables related to scale, rotation, lighting, and resolution.

Conclusions

Because our aim is user security, this work proposed a methodology for identifying facial images encrypted via the Logistic Map method. During the facial recognition process, local textural features were extracted from binary cryptograms, texture extraction was performed using the CCR transform, and classification was conducted by applying a multi-class classifier. The classifier was based on facial cryptogram statistics and utilized the CCR histogram as a characteristic vector. Our proposal was validated with four numerical experiments, wherein the CCR histogram was calculated using 3 x 3 and 4 x 4 pixel observation windows. Two databases of encrypted facial images were employed, and the efficiency increased from 96.39 to 100%. During image acquisition, the conditions (scale, rotation, camera resolution, and lighting) were controlled. If the conditions had not been controlled, the classification error would have increased, as more variables and more noise sources would have been involved in our experiments. Even so, our proposal shows a high potential for application in the field of computer security.

Acknowledgments

The authors thank Mexico's National Council of Science and Technology (CONACyT) for the support granted. J. Aguilar- Santiago expresses his gratitude to CONACyT for the scholarships received. The authors thank all students for their collaboration.

Author contributions

José T. Guillen-Bonilla: investigation, methodology, and software. Jorge Aguilar-Santiago: software. Juan C. Estrada-Gutiérrez: formal analysis. Maricela Jiménez-Rodríguez: investigation, methodology, and software. All authors participated in the writing, review, and editing of this manuscript.

Conflicts of interest

The authors declare no conflicts of interest regarding the publication of this paper.

Data availability

The algorithms, figures, and pixel data used to support the findings of this study are included within the article.

References

Afifi, A. (2019). A chaotic confusion-diffusion image encryption based on Henon map. *International Journal of Network Security & Its Applications*, *11*(4), 19-30. https://doi.org/10.5121/ijnsa.2019.11402

Aguilar Santiago, J., Flores Siordia, O., Guillen Bonilla, J. T., Estrada Gutiérrez, J. C., González Novoa, M. G., and Jiménez Rodríguez, M. (2020). Chaotic cryptosystem for selective encryption of faces in photographs. *Security and Communication Networks*, *2020*, 1-22. https://doi.org/10.1155/2020/8848356

Ahmad Khan, F., Bouridane, A., Boussakta, S., Jiang, R., and Almaadeed, S. (2021). Secure facial recognition in the encrypted domain using a local ternary pattern approach. *Journal of Information Security and Applications, 59*, 102810. https://doi.org/10.1016/j.jisa.2021.102810

Alaei, F., Alaei, A., Pal, U., and Blumenstein, M. (2019). A comparative study of different texture features for document image retrieval. *Expert Systems with Applications*, *121*, 97-114. https://doi.org/10.1016/j.eswa.2018.12.007

Bansal, P., Mittal, S., and Gupta, M. (2014). *Using Locality Preserving Projections in Face Recognition*. *2*(3), 99-108. https://api.semanticscholar.org/CorpusID:212515605

Bello-Cerezo, R., Bianconi, F., Di Maria, F., Napoletano, P., and Smeraldi, F. (2019). Comparative evaluation of hand-crafted image descriptors *vs.* off-the-shelf CNN-based features for colour texture classification under ideal and realistic conditions. *Applied Sciences*, *9*(4), 738. https://doi.org/10.3390/app9040738

Carlos Eduardo Thomaz — Personal Web Page (n.d.). https://fei.edu.br/~cet/facedatabase.html

Dunn, J. C. (1973). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, *3*(3), 32-57. https://doi.org/10.1080/01969727308546046

fliphtml5.com (n.d.). *4243_0819_rp_qtrly-threats-aug-2019_lores*. https://fliphtml5.com/rshui/bhkw/4243_0819_rp_qtrly-threats-aug-2019_lores/

Gao, S., Wu, R., Wang, X., Liu, J., and Li, Q. (2023a). EFR-CSTP: Encryption for face recognition based on the chaos and semi-tensor product theory. *Information Sciences, 621*, 766-781. https://doi.org/10.1016/j.ins.2022.11.121

Gao, S., Wu, R., Wang, X., Liu, J., Li, Q., Wang, C., and Tang, X. (2023b). Asynchronous updating boolean network encryption algorithm. *IEEE Transactions on Circuits and Systems for Video Technology*, *33*(8), 4388-4400. https://doi.org/10.1109/TCSVT.2023.3237136

Gao, S., Wu, R., Wang, X., Wang, J., Li, Q., Wang, C., and Tang, X. (2023c). A 3D model encryption scheme based on a cascaded chaotic system. *Signal Processing, 202*, 108745. https://doi.org/10.1016/j.sigpro.2022.108745

Guillen-Bonilla, J. T., Kurmyshev, E., and Fernández, A. (2007). Quantifying a similarity of classes of texture images. *Applied Optics*, *46*(23), 5562. https://doi.org/10.1364/AO.46.005562

Ibrahim, D. R., Abdullah, R., and Teh, J. S. (2021). Multifactor authentication system based on color visual cryptography, facial recognition, and dragonfly optimization. *Information Security Journal: A Global Perspective, 30*(3), 149-159. https://doi.org/10.1080/19393555.2020.1817633

Ibrahim, D. R., Abdullah, R., Teh, J. S., and Alsalibi, B. (2019). *Authentication for ID cards based on colour visual cryptography and facial recognition* [Conference presentation]. 3rd International Conference on Cryptography, Security and Privacy. https://doi.org/10.1145/3309074.3309077

Jiménez-Rodríguez, M.., Flores-Siordia, O., and González-Novoa, M. G. (2015). Sistema para codificar información implementando varias órbitas caóticas. *Ingeniería, Investigación y Tecnología*, *16*(3), 335-343. https://doi.org/10.1016/j.riit.2015.05.004

Jiménez-Rodríguez, M., Padilla Leyferman, C. E., Estrada Gutiérrez, J. C., González Novoa, M. G., Gómez Rodríguez, H., and Flores Siordia, O. (2018). Steganography applied in the origin claim of pictures captured by drones based on chaos. *Ingeniería e Investigación*, *38*(2), 61-69. https://doi.org/10.15446/ing.investig.v38n2.64509

Kalech, M. (2019). Cyber-attack detection in SCADA systems using temporal pattern recognition techniques. *Computers & Security, 84*, 225-238. https://doi.org/10.1016/j.cose.2019.03.007

Kumar, V., and Girdhar, A. (2021). A 2D logistic map and Lorenz-Rossler chaotic system based RGB image encryption approach. *Multimedia Tools and Applications*, *80*(3), 3749-3773. https://doi.org/10.1007/s11042-020-09854-x

Kurmyshev, E. V., and Cervantes, M. (1996). A quasi-statistical approach to digital binary image representation. *International Journal of e-Navigation and Maritime Security, 42*(1), 104-116. https://doi.org/10.1016/j.enavi.2017.05.007

Kurmyshev, E. V., Poterasu, M., and Guillen-Bonilla, J. T. (2007). Image scale determination for optimal texture classification using coordinated clusters representation. *Applied Optics*, *46*(9), 1467. https://doi.org/10.1364/AO.46.001467

Kurmyshev, E. V., and Sánchez-Yáñez, R. E. (2005). Comparative experiment with colour texture classifiers using the CCR feature space. *Pattern Recognition Letters*, *26*(9), 1346-1353. https://doi.org/10.1016/j.patrec.2004.11.028

Leyferman, C. E. P., Bonilla, J. T. G., Gutiérrez, J. C. E., and Rodríguez, M. J. (2023). A novel technique for texture description and image classification based in RGB compositions. *IET Communications*, *17*(10), 1162-1176. https://doi.org/10.1049/cmu2.12601

Luo, X., Xu, Y., and Yang, J. (2019). Multi-resolution dictionary learning for face recognition. *Pattern Recognition, 93*, 283-292. https://doi.org/10.1016/j.patcog.2019.04.027

Mohan, J., and R., R. (2021). Enhancing home security through visual cryptography. *Microprocessors and Microsystems, 80*, 103355. https://doi.org/10.1016/j.micpro.2020.103355

Nanni, L., Brahnam, S., and Lumini, A. (2019). Texture descriptors for representing feature vectors. *Expert Systems with Applications, 122*, 163-172. https://doi.org/10.1016/j.eswa.2018.12.052

Pan, H., Lei, Y., and Jian, C. (2018). Research on digital image encryption algorithm based on double logistic chaotic map. *EURASIP Journal on Image and Video Processing*, *2018*(1), 142. https://doi.org/10.1186/s13640-018-0386-3

Peng, C., Wang, N., Li, J., and Gao, X. (2019). DLFace: Deep local descriptor for cross-modality face recognition. *Pattern Recognition, 90*, 161-171. https://doi.org/10.1016/j.patcog.2019.01.041

Ren, L., and Zhang, D. (2022). A privacy-preserving biometric recognition system with visual cryptography. *Advances in Multimedia*, *2022*, 1-7. https://doi.org/10.1155/2022/1057114

Rodríguez, M. J., González-Novoa, M. G., Estrada-Gutiérrez, J. C., Acosta-Lúa, C., and Flores-Siordia, O. (2016). Secure point-to-point communication using chaos. *DYNA*, *83*(197), 180. https://doi.org/10.15446/dyna.v83n197.53506

Sánchez-Yáñez, R. E., Kurmyshev, E. V., and Cuevas, F. J. (2003). A framework for texture classification using the coordinated clusters representation. *Pattern Recognition Letters, 24*(1-3), 21-31. https://doi.org/10.1016/S0167-8655(02)00185-X

Shi, L., Song, X., Zhang, T., and Zhu, Y. (2019). Histogram-based CRC for 3D-aided pose-invariant face recognition. *Sensors*, *19*(4), 759. https://doi.org/10.3390/s19040759

Suman, R. R., Mondal, B., and Mandal, T. (2022). A secure encryption scheme using a composite logistic sine map (CLSM) and SHA-256. *Multimedia Tools and Applications*, *81*(19), 27089-27110. https://doi.org/10.1007/s11042-021-11460-4

Wu, R., Gao, S., Wang, X., Liu, S., Li, Q., and Erkan, U. (2022). AEA-NCS: An audio encryption algorithm based on a nested chaotic system. *Chaos, Solitons & Fractals, 165*, 112770. https://doi.org/10.1016/j.chaos.2022.112770

Xiang, H., and Liu, L. (2020). An improved digital logistic map and its application in image encryption. *Multimedia Tools and Applications*, *79*(41-42), 30329-30355. https://doi.org/10.1007/s11042-020-09595-x

Yang, J., Wang, X., Han, S., Wang, J., Park, D. S., and Wang, Y. (2019). Improved real-time facial expression recognition based on a novel balanced and symmetric local gradient coding. *Sensors*, *19*(8), 1899. https://doi.org/10.3390/s19081899

Zhang, L., Yang, M., Feng, X., Ma, Y., and Zhang, D. (2014). *Collaborative representation based classification for face recognition*. https://doi.org/10.48550/arXiv.1204.2358

Research Article/System and computer engineering

http://doi.org/10.15446/ing.investig.101231

Business-IT Alignment Maturity Diagnosis of a Health Organization using Luftman's SAM Model

Diagnóstico de madurez de alineación negocio-TI de una organización de salud utilizando el modelo SAM de Luftman

Gloria Restrepo-Espinel[1], Miguel Torres-Moreno[2], and Jairo Aponte-Melo[3]

ABSTRACT

Information technologies (ITs) provide optimization, service improvement, and a higher level of competitiveness, adding value to organizations' strategic results and enabling the implementation of digital transformation initiatives. However, organizations must achieve high alignment between business and IT strategies in order to maximize these contributions. Researchers have proposed various methods to improve alignment and have created measurement instruments to assess it within an organization. One of the most recognized is Luftman's SAM model (strategic alignment maturity model), which has a scale of five alignment levels and evaluates six dimensions. This work assesses the degree of alignment between business and IT units within a healthcare entity in Villavicencio (Colombia) using the Luftman instrument. The results indicate that the organization has a maturity level of 2, which denotes a low alignment. The organization needs to strengthen its IT unit and increase its participation in strategic planning.

Keywords: strategic planning, technology assessment, information technology, business management, strategic alignment

RESUMEN

Las tecnologías de la información (TI) brindan optimización, mejoras del servicio y un mayor nivel de competitividad, agregando valor a los resultados estratégicos de las organizaciones y posibilitando la implementación de iniciativas de transformación digital. Sin embargo, las organizaciones deben lograr una alta alineación entre las estrategias de negocio y de TI para maximizar estas contribuciones. Los investigadores han propuesto varios métodos para mejorar la alineación y han creado instrumentos de medición para evaluarla dentro de una organización. Uno de los más reconocidos es el modelo SAM de Luftman (modelo de madurez de alineación estratégica), que tiene una escala de cinco niveles de alineación y evalúa seis dimensiones. Este trabajo evalúa el grado de alineación entre unidades de negocios y TI dentro de una entidad de salud en Villavicencio (Colombia) utilizando el instrumento de Luftman. Los resultados indican que la organización tiene un nivel de madurez de 2, lo cual denota una baja alineación. La organización requiere fortalecer su unidad de TI e incrementar la participación de esta en la planeación estratégica.

Palabras clave: planeación estratégica, evaluación de la tecnología, tecnología de la información, administración de empresas, alineación estratégica

Received: February 23th, 2021
Accepted: September 13th, 2023

Introduction

The introduction of technology into our daily lives is an increasingly crucial issue in our society. Information technologies (ITs) allow organizations to optimize, improve, and revamp services and products, adding value to their strategic results. Moreover, digital transformation seeks to fulfill organizational and strategic transformation goals using digitalization strategies, frameworks, processes, and practices (Zhu *et al*., 2021). A recent Gartner publication predicts that 50% of organizations will experience more significant collaboration between business and IT teams by 2025 (Gartner, n.d.-a).

Nevertheless, practical evidence has shown that, for many organizations, it is hard to align business and IT appropriately. This impediment also affects the efforts of organizations that seek to advance digital transformation initiatives, since adequate IT-business alignment is an essential requirement to start a true digital transformation (Lisienkova, 2017). Various approaches have been proposed to improve alignment, and measurement instruments have been created to assess its degree of maturity (Baker and Singh, 2019; Berberat and Baudet, 2019; Chan and Reich, 2007; Darii *et al*., 2020; Haghighi Rad and Rowzan, 2018; Henderson and Venkatraman, 1993; Luftman *et al*., 2017). Henderson and Venkatraman (1993) and Luftman (2000) identified six dimensions that should be evaluated and controlled in order to reach a harmonic IT-business relationship that leads to proper alignment and the achievement of an organization's goals.

[1] ME in Systems and Computer Engineering, Universidad Nacional de Colombia, Colombia. Email: glrestrepoe@unal.edu.co

[2] PhD in Systems and Computer Engineering, Universidad Nacional de Colombia, Colombia. Affiliation: Business analyst, Endava SAS, Colombia. Email: mtorresm@unal.edu.co

[3] PhD in Engineering (Systems and Computing). Affiliation: Associate professor, Department of Systems and Industrial Engineering, Universidad Nacional de Colombia, Colombia. Email: jhapontem@unal.edu.co

In the case of manufacturing companies, the digital transformation of industrial production requires not only the adoption of cyber-physical systems and Internet of Things (IoT), platforms but also a focus on readapting processes and the overall organization in order to deal with a business and added-value problem. Issa *et al*. (2018) presented a framework to both assess and advise Industry 4.0 implementations. This framework takes some elements and concepts from capability maturity models (CMM) and the alignment models regarding ITs and strategies, business processes, and organizational structures. In a more general business framework, Maciá Pérez *et al*. (2021) proposed a model for implementing and operating a portfolio of strategic IT projects, aligning IT with business goals by involving top management levels in decision-making, and promoting cooperation, communication, and leadership. After studying 43 small- medium-sized enterprises (SMEs), Canhoto *et al*. (2021) set out to determine how digital alignment is reached. Their findings, as well as an initial framework and an inductive analysis, were used to propose a model for digital alignment in SMEs. These authors identified five phases of digital aligning, ranging from passive acceptance to transformation.

Using the Luftman instrument, several companies in different economic sectors have been evaluated with the purpose of assessing their degree of alignment. One recent study involved 395 companies in the US Fortune 1 000 list (Luftman *et al*., 2017), 2,87% of which belonged to the healthcare sector. Moreover, it is worth mentioning that, among the organizations in this sector, ITs are given little importance, even though healthcare is a key aspect for developing countries, especially given the current public health conditions around the world.

The past several years have been grueling for the healthcare and life sciences industry, with internal and external dynamics collectively forcing organizations to transform their operation and business models.

The shortage of healthcare workers is a growing problem, and healthcare delivery organizations (HDOs) are exploring digital health alternatives to meet the demand. Digital health products and services have shown good results with regard to enhancing patient experience, improving population health, reducing costs, and improving the work life of healthcare clinicians and staff. The COVID-19 pandemic has accelerated the adoption of digital health technologies in the industry, and HDOs now agree on the need to accelerate digital transformation (Gartner, n.d.-b).

According to Gartner, HDO CIOs (Chief information Officers) should take the following steps to accelerate digital transformation in the healthcare industry (Gartner, n.d.-b):

- Enabling consumer engagement, activation, and the delivery of digital health and healthcare services by adopting digital-first technology domains. This includes telehealth, remote patient monitoring, and patient portals.

- Seizing the opportunity created by the COVID-19 pandemic to deploy enterprise-grade, digital-first technology domains within pandemic-driven digital initiatives. This will help HDOs prepare for the future.

- Prepare for a new normal by adopting a digital-first business strategy that makes digital interactions the first, the preferred, and the pervasive medium for engagement, health, and healthcare delivery. This will require a significant cultural shift within HDOs, but it is essential for the future of the healthcare industry.

In Colombia, the healthcare sector is highly problematic and the origin of the most significant citizen complaints. However, digital transformation is somehow changing the healthcare sector in Colombia. The government has invested in digital health initiatives, such as electronic health records and telemedicine, in order to improve quality, efficiency, and access to care. These investments have the potential to make a significant impact on healthcare in Colombia. In the strictly technological aspect, there are still no guidelines for implementing architectures that integrate health information systems and technological platforms, and many local solutions have emerged over time (Bernal-Acevedo and Forero-Camacho, 2011). In many cases, IT solutions are not integrated with or incompatible with government systems, causing serious problems for information exchange. As a result, isolated systems support business processes, leading to duplicated efforts, low data quality, and low coverage. This lack of integration, resources, and standardization results in poor patient care and low-quality health services.

Furthermore, health data constitute a fundamental input for the management of information and knowledge aimed at strengthening regional and state-wide decision-making. The information generated would make it possible to anticipate health trends and take proactive and deliberate actions to face any challenges or threats to the population (Castrillón *et al*., 2012). However, this use of data, and their subsequent transformation into information and knowledge requires an adequate alignment of the business units and ITs used in the health sector. In light of the above, the research question that guides our case study is as follows:

What is the maturity level of the strategic alignment between the business and IT units within a healthcare entity in Villavicencio, Colombia?

To answer this question, and as an initial exploratory step, this study aims to diagnose the maturity of the strategic alignment within an entity in the health sector using Luftman's strategic alignment maturity (SAM) model. Diagnosis is an essential first step to plan IT-business alignment for the region's health sector and subsequently change the business strategy and adopt advanced digital technologies that support it (Zeebroeck *et al*., 2023).

We analyzed the technological alignment factors in the dimensions given by Luftman's SAM model. Specifically, we conducted a case study to diagnose the maturity of alignment between the IT and business units in an institution from Villavicencio's health sector (Colombia). Additionally, we used this study to assess the instrument used as a cross-sectional aspect. Based on this experience, we provide recommendations for using the model and point out its possible limitations and benefits with regard to the studied problem.

The remainder of this paper explores the following aspects. Firstly, it presents the conceptual foundations, background, and description of Luftman's SAM model. Next, it describes the most relevant characteristics of the healthcare entity under analysis. The paper then explains the implementation of Luftman's SAM instrument within this entity. Subsequently, it summarizes the results according to the dimensions defined by Luftman. A comparison between our results and those of previously reported healthcare and Latin American organizations follows. Finally, the paper discusses the possible limitations of this work and the application of the instrument and concludes with key takeaways and proposals for future work.

Strategic alignment and Luftman's SAM

This section presents the conceptual foundations of business-IT alignment (BITA) and the existing strategic models, emphasizing the features of the SAM model.

Conceptual foundations

BITA studies the process and results of strategic alignment between IT units and the business strategy. A detailed definition by Luftman *et al*. (2017, p. 2) states that BITA "addresses both how IT is in harmony with the business, and how the business should, or could be in harmony with IT. Alignment evolves into a relationship where the function of IT and other business functions adapt their strategies together." Other authors who have studied the benefits of using ITs in companies also call it *leverage/linkage* (Henderson and Venkatraman, 1993), *enablement* (Chan *et al*., 1997), *transformation* (Luftman, 2000), and *business processes optimization* (Sabherwal and Chan, 2001). Other studies focus on the relationship between the IT and business domains and refer to it as *fit* (Benbya and McKelvey, 2006), *linkage* (Reich and Benbasat, 2000), *matching* (Chan *et al*., 1997), *bridging* (Zee and Jong, 1999), *fusion* (Smaczny, 2001), and *harmonization* (Chan, 2008).

Achieving BITA involves translating communication and strategy at executive levels (Sabherwal and Chan, 2001). Today, a successful BITA focuses on organizations' tactical and operations levels and management activities in order to achieve cohesive business and IT operations (Tarafdar and Qrunfleh, 2009). Therefore, it requires strong support from senior managers, proper prioritization, trusting relationships, and effective communication between different levels (Avison *et al*., 2004). Other authors suggest that enterprise architecture (EA) has the potential to improve and support strategic alignment. Bhattacharya (2017, 2018) proposes a model that relates the Open Group Architecture Framework's constructs (TOGAF) with a technique for modeling strategic alignment that also enriches the original SAM model (Henderson and Venkatraman, 1993). Achieving BITA also leverages digital transformation initiatives, which requires rapid changes in IT infrastructure and strategic direction (Zeebroeck *et al*., 2023).

In management sciences, several efforts turn to studying and operationalizing alignment (Al-Surmi *et al*., 2020; Chi *et al*., 2020; Duhamel *et al*., 2021; Ghonim *et al*., 2020; Hart and Burke, 2020; Strategic Direction, 2019; Miyamoto, 2019; Patterson, 2020; Sabherwal *et al*., 2019; Sholihah *et al*., 2019). For example, Bernat and Karabag (2019) analyze how organizations can overcome shortcomings in knowledge and technology by systematically monitoring their environment and building an alignment strategy. All this, by identifying technological challenges and focusing and managing a research and development (R&D) portfolio that, in the end, optimizes technology upgrading and maximizes results.

Luftman's SAM model

Luftman presented the SAM model in his 1996 book titled *Competing in the information age: Strategic alignment in practice*(J. N. Luftman, 1996). This model defines six dimensions to categorize the degree of alignment of an organization's business and IT units. Each dimension is evaluated using a five-level capability scale (Table 1).

Likewise, Luftman defines six dimensions with 41 attributes for assessing BITA (Luftman, 2000), wherein evaluations are carried out via a five-point Likert scale questionnaire. The dimensions and their attributes are presented in Figure 1.

Other authors report using Luftman's instrument, for instance, to assess and identify the impact of IT governance implementations on BITA (de Haes and van Grembergen, 2009). Furthermore, Latin America (Gutiérrez *et al*., 2008, 2009; Torres-Moreno and Aponte-Melo, 2021) and Russia (Lisienkova, 2017) report assessments of several industrial sectors' alignment that employ data collection from individuals in SMEs.

Description of the analyzed healthcare entity

The Department of Meta (Colombia) has 98 1726 inhabitants in 29 municipalities, covering a large part of the Orinoquía region. Its capital is Villavicencio, which has more than 450 000 inhabitants (DANE, n.d.) and the region's most specialized healthcare service levels. Villavicencio has more than 397 entities that provide and promote health services. Moreover, the Departmental Hospital and 42 clinics serve 104 372 people, corresponding to 23% of the citizens (Secretaría de Salud del Meta, n.d.).

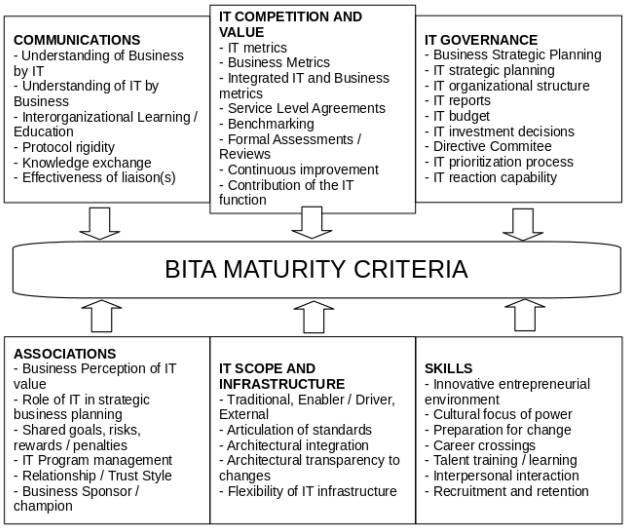


Figure 1. Alignment maturity dimensions and criteria
Source: Adapted from (J. Luftman, 2000)

The selected healthcare entity has provided both basic and specialized services to the citizens of Villavicencio and the Orinoquía region since the 1970s. However, according to the established national classifications, it does provide services at the most advanced level. Its services include intensive care units (ICU), hospitalization, surgery, outpatient consultation, diagnosis, therapeutic support, emergencies, rehabilitation, medical and surgical treatments, and highly qualified specialists. For the sake of confidentiality, this hospital is hereinafter referred to as *the healthcare entity*.

Table 1. Description of maturity levels

Level	Description
1 – Initial	The organization has the lowest maturity level regarding alignment.
2 – Committed processes	The organization has an initial commitment to strategic alignment between the IT and business units.
3 – Established and focused processes	The organization exhibits strategic alignment focused on the IT areas and prioritizes some organizational objectives.
4 – Improved/ managed processes	The organization shows a good management regarding strategic alignment with IT. The administration stresses the concept of *IT* as an element of value for the organization.
5 – Optimized processes	The organization has achieved a high alignment between ITs and business strategies.

Source: (J. Luftman, 2000)

Organizational structure

The healthcare entity has more than 700 employees, including administrative and functional areas. The entity's workforce can be grouped into four areas: administrative, healthcare, IT, and support staff (Table 2). There is a general management unit in its organizational structure, on which two management units and three directorates depend. The two management areas are called *Medical Management* and *Administrative/Financial Management*, while the three directorates are *Human Resources*, *Operations*, and *Technology*. The Technology Directorate provides services to all of the entity's supporting processes, which are both managerial and operational in nature.

Table 2. Amount of personnel per area

Area	People	%
Administrative	19	3
Healthcare	541	73
IT	7	1
Support staff	172	23

Source: Authors

Diagnosis of the business-IT maturity level

This study is a descriptive research with a mixed approach (Yin, 2010). It is based on the questionnaire defined by Luftman (Luftman *et al.*, 2017). Considering that the instrument is designed for business and IT areas, a discretionary sample of 31 people from the administrative, healthcare, and IT areas was interviewed and surveyed.

This section clarifies the processes carried out to apply Luftman's instrument: data collection, information processing, and results interpretation.

Applying Luftman's SAM instrument

One of the authors interviewed the selected staff, collecting qualitative information to understand the organization and the personnel's perspectives on each evaluated factor. The interviewer clarified the study's objectives in each meeting and characterized the interviewee (role, responsibilities, and experience). Then, the questionnaire was applied. The interviews lasted approximately one hour, and they were carried out between December 2019 and April 2020.

Data collection

The data collection process was backed by the healthcare entity's Quality Office and was conducted as follows. First, we identified key personnel in each organizational unit in order to interview them. They were classified into three categories according to their participation and responsibilities in strategic management within their units: i) **strategic**: people who make strategic decisions; ii) **functional**: employees responsible for the operation and monitoring of strategic decisions within the unit; iii) **operational**: staff who are in charge of executing day-to-day operations and are generally are not involved in making strategic decisions. Second, we contacted the selected staff to explain the study's goals and procedure. Third, we carried out five pilot tests in the healthcare entity with people from each administrative, IT, and support area. These tests allowed fine-tuning the

procedure and vocabulary used, making the necessary adjustments to clarify the questions, and improving the tools employed (presentation, form, and interview). Lastly, we scheduled and conducted the interviews in order to apply the Luftman instrument and collect information. Table 3 shows the categorization of the interviewees.

This study's data come from the interviews and Luftman questionnaire responses. In addition, 27 out of 31 interviews were recorded, transcribed, and examined in order to obtain relevant information for each dimension of the Luftman instrument. The data collected were synthesized in tables and graphs, and basic descriptive statistical methods were used to analyze them.

Table 3. Distribution of the interviewed personnel

Process	Strategic	Functional	Operative
General management	1		
User management		1	
Quality management	1	4	
Purchases and storage management	1		
Infrastructure management		2	
Admissions		1	
Financial management	1		
Human resources		1	
Hospitalization		2	
Legal		1	
Service delivery planning	1		
Radiology and imaging		1	
Ambulatory services		2	
ICU Adults		1	
Surgery		2	
UCIA (intensive care unit for adults)		1	
UCIN (intensive care unit for neonates)		1	
Emergencies		1	
Statistics		1	
IT Directorate	1		3
Total	**6**	**22**	**3**

Source: Authors

Results interpretation

This section summarized the results obtained according to the six dimensions defined by Luftman(J. Luftman *et al.*, 2017). The average response to each question is tabulated for each dimension, and the general average is also presented. In all cases, the scores range from one (the lowest grade) to five (the highest grade). These quantitative data are briefly analyzed and complemented with a synthesis of the interviewees' perceptions.

Communication

Table 4. Communication

Question	1	2	3	4	5	6
Average	2,68	2,48	2,77	2,10	2,35	1,61
Dimension average: 2,33						

Source: Authors

First of all, the scores indicate that one area's knowledge of the other is limited (Table 4. Communication). The data collected show that the current flow and effectiveness of the communications between the IT and business areas do not allow them to fully understand each other's processes, functions, strategies, plans, IT environments, risks, and priorities. They exchange information primarily when the internal needs of the entity require it. Instead of having permanent channels to promote the transfer of information and collaboration, the manager of each area works individually, guided by their own needs and plans.

Moreover, the effectiveness of the knowledge exchange is low. They have semi-structured guidelines, and there is no complete understanding of them. There is no cohesion at the strategic level between the IT and business areas; their communication is mainly transactional, where priorities are associated with operational information.

The organization is increasingly aware of the benefits of technology to all its processes. For example, the management of information systems has helped the institution fulfill the national regulations regarding the periodic report of the variables requested by the government institutions for surveillance and control purposes. Furthermore, it is worth noting that the structuring and exchange of knowledge between the IT and business areas are underway.

Furthermore, it is worth mentioning that the entity uses standard methods to offer organizational learning sessions, as defined in its internal communication manual. We can conclude that the organization has a hierarchical and bureaucratic organizational structure, and that the communication style is mainly unidirectional: from the strategic levels to the other dependencies.

IT competence and value

Table 5. IT competence and value

Question	1	2	3	4	5	6	7	8
Average	1,40	1,87	2,00	2,00	2,03	2,13	1,94	4,00
Dimension Average: 2,17								

Source: Authors

The results in this regard (Table 5) reveal that the healthcare entity does not use metrics for measuring the value added by IT to the business strategy. The institution uses measures to assess technical factors such as the system's availability

and response times. However, they only perform a cost-benefit evaluation of IT investment regarding administrative and managerial matters when the need arises. There is no systematic feedback between the IT and business areas, which is corroborated by the interviewees' opinions.

Service-level agreements are oriented towards operational issues. In that sense, any established processes should measure and track user satisfaction, and the information collected should be the basis for implementing improvement actions.

Benchmarking practices are beginning to be used. During this study, the organization was migrating to a new information system that offers more and better functionalities and encompasses more business processes. This new system was selected through a rigorous benchmarking process that involved all interested areas.

It was found that 77,4% of the participants believe that IT has made a solid contribution to achieving the organization's strategic objectives. This suggests that IT supports their business goals despite lacking formal metrics.

IT governance

The evidence shows that business planning is carried out with minimal participation from the IT area (Table 6. IT governance). Moreover, there is no specific IT budget, and, like any other institution's IT department, the area determines its needs and submits them to the financial area. After that, the financial area makes decisions regarding financial approval, considering all requests from all areas. Due to this competition for resources, IT projects are sometimes not funded.

Table 6. IT governance

Question	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Average	2,29	2,33	1,48	3,03	2,10	1,90	3,10
Dimension Average: 2,32							

Source: Authors

Despite this, the participants know that IT is an enabler of business strategy. In other words, the organization recognizes that IT investment decisions are strongly connected to improving business effectiveness. Furthermore, it is well known that IT cannot appropriately respond to all of the organization's requests, as it is such a small area. Thus, the perception that IT does not respond appropriately to business requirements is justified.

In this dimension, it is interesting that the assessment of the seven factors was dispersed. The lower-graded factors indicate that the healthcare entity must improve the IT budget criteria and prioritize IT projects. The need to strengthen the IT department is evident; it requires a stable budget that allows them to carry out more technology projects. A requirements engineering process should be implemented, so that each department can determine the best way to enhance business processes through IT.

Partnership/association

Table 7. Partnership/association

Question	**1**	**2**	**3**	**4**	**5**	**6**
Average	3,39	2,39	1,67	2,14	3,00	2,32
Dimension Average: 2,48						

Source: Authors

As already mentioned, the IT area is perceived as a fundamental enabler of business operations. All interviewees consider that information systems are easing decision-making by providing accurate and timely data. Although the IT area is still used more to enable business processes than to drive its strategy, they already regard the association between the areas as a long-term relationship.

According to the responses, the entity manages relationships on an *ad hoc* basis, without formal processes between areas or a high-level promoter from the business area who drives initiatives (Table 7. Partnership/association).

A fact that stands out is that 45,16% of the respondents perceive the IT area as a fundamental driver for the entity's activities. Moreover, 51,61% consider IT to be responsible for most risks and state that it receives little reward for successful initiatives.

There are divergent opinions regarding the type of relationship between IT and the business. For the functional areas, the relationship with IT is transactional, *i.e.*, they mainly request and receive technical support. In contrast, the relationship is viewed as a partnership with regard to administrative roles and IT.

Scope and infrastructure of IT

For the majority of the interviewees, the technological infrastructure is driven by business strategy requirements. For some, it is emerging as a resource that can enable a rapid response to changes in the market (Table 8. Scope and infrastructure of IT). However, from the IT area's perspective, this is achieved at a very high cost and with a meager reward. Thus, for them, there is a clear effort-reward imbalance.

Table 8. Scope and infrastructure of IT

Question	**1**	**2**	**3**	**4**	**5**
Average	2,39	2,21	2,17	1,23	2,55
Dimension Average: 2,11					

Source: Authors

The healthcare entity mostly has transaction-oriented information systems and business process enablers. In this case, standards are applied at the functional area level for all patient care management.

Furthermore, changes in the entity systems are generally disruptive, affecting the functional units where the change requirements arise. 77,42% of the interviewees indicate that changes cause inconveniences and affect their activities. Furthermore, 35,48% indicate that the business strategy's requirements determine the organization's technological infrastructure.

This dimension exhibited the lowest score, with a maturity degree of 2. This result reveals a low level regarding infrastructure flexibility and the ability to incorporate and manage new technologies.

IT skills

Table 9. IT skills

Question	1	2	3	4	5	6	7
Average	2,35	2,19	2,97	1,82	1,97	3,06	2,38
Dimension Average: 2,39							

Source: Authors

For the analyzed entity, entrepreneurship is moderately encouraged by functional units, and essential technology decisions are regularly made by senior management (Table 9. IT skills).

As an institution regulated by health protocols, it is well prepared for changes at the functional level. IT is currently making a software change to manage its missional patient care and transversal processes. Likewise, we identified a well-defined process for managing resistance to change through training programs, awareness campaigns, information dissemination initiatives, and feedback.

Staff exchange between the IT and business areas seldom occurs. Most of the personnel interviewed express that genuine trust has been reached between the two areas. Hiring is focused on the applicant's technical expertise. The institution has training plans, but they depend on the needs of the functional units.

Findings and recommendations

According to our analysis, there is an evident effort-reward imbalance in the IT area. Currently, IT is the area that assumes all the risks associated with IT-based decisions and receives a low reward for its work with all of the entity's units. Risks must be shared with the business area as a necessary attitude for establishing trust and partnership between them. As stated by Maciá Pérez *et al*. (2021), the senior management level must be involved in IT strategy and plans to improve alignment and enable the digital transformation of business processes.

Similarly, we suggest starting an internal program that addresses the lack of knowledge about the value of IT for the entire organization. This should make IT plans, strategies, and actions known and stress the importance of funding IT initiatives. Understanding that technology infrastructure and the strategic direction of the organization must be synchronized is pivotal for addressing new business opportunities (Zeebroeck *et al*., 2023).

It would be helpful for managers to acquire a deeper understanding of the processes, roles, and responsibilities of other areas. This will enrich knowledge and facilitate the construction of more cohesive strategic plans, allowing to develop projects that meet the requirements of the business, as well as for the continuous evaluation, control, and evolution of the organization's enterprise architecture.

The interviews uncover the lack of effective communication, showing that there is currently no full integration of IT into business areas. Each area is only qualified to perform its tasks and executes them according to what is established. Most interviewees have a comprehensive understanding of organizational strategies, plans, and goals but are unaware of IT environments and their management.

The priorities regarding interaction between the IT and business areas are determined more by periodic needs than by precise requirements that allow the institution to achieve a managed business architecture. In addition, there is no consolidation of risks that allows generating a response and contingency plans to transversal processes; only those of the IT area are established. Therefore, IT must lead a holistic study of the risks since they have more knowledge about their consequences and the strategies for mitigating them.

Likewise, our analysis indicates that the healthcare entity does not use metrics to assess the contribution of IT in reaching its strategic goals. It is quite noticeable that there is little participation by the IT and business areas in planning (IT governance dimension). This state of affairs leads to a basic structuring that seeks to meet the functional needs rather than strategic ones.

Moreover, it is necessary to strengthen mutual trust between the areas in order to consolidate the role of IT in the organization, as well as to formalize processes so that IT initiatives are better supported by business personnel.

The scope and strength of the IT area are limited due to the unit's small size, both in staff and equipment. This lack of resources reduces its capacity to provide flexibility and support for the infrastructure. Improving IT size, resources, training, and skills should aim towards ensuring its suitability to carry out current activities and overcome future challenges.

Despite the low maturity level and the limitations regarding personnel and resources in the IT area, it is essential to recognize that the IT department supports information

technologies that allow the institution to fulfill the goals of the healthcare sector and timely report to the regional surveillance entities.

It is worth mentioning that the healthcare entity has taken an essential step by changing the technology used in order to control transversal and core mission processes. Gradually, the other areas are giving more importance to the role of IT. Thus, the entity is evolving into a state where the IT area defines business strategies and plans. The role of IT will transition from a mere service provider to an enabler of the organization's strategic plans.

The Ambulatory and Hospital Health accreditation process that the entity has undertaken allows it to design and improve processes that will be fundamental to reach the next degree of maturity. Some of them are related to information management, as the Ministry of Health seeks to integrate all healthcare and administrative areas. This integration will also bolster all technical resources and improve decision-making at all levels, in addition to technology management (Ministerio de Salud, 2018).

Limitations of the case study

While the COVID-19 pandemic undoubtedly presented challenges, we were fortunate to complete the interviews for this study. Now, we can turn our attention to understanding the impact it had on both the interview process and the institution itself.Second, based on our experience with the Luftman instrument, we believe that its application must be preceded by a characterization of the entity, allowing the interviewers to become familiar with core aspects of the institution, such as its organizational culture, structure, mission and vision, values, and principles. Third, although the instrument's application reveals the alignment gaps between the IT and business areas, the design of an action plan that ensures alignment is not evident and direct. Finally, we consider it essential to carry out the interviews in person, in addition to guiding the respondents, as they may not understand the terminology of the questionnaire or misinterpret its questions.

Conclusions and future work

In this paper, we applied the strategic alignment maturity (SAM) model to diagnose the maturity of strategic alignment in a healthcare entity. Our findings indicate that the entity's IT area is little involved in the organization's strategic planning. This suggests that more integration between the IT and business areas is needed to develop joint strategic plans and programs.

Gartner reports (Gartner, n.d.-a) that healthcare providers are planning to invest in AI/ML, distributed cloud, responsible AI, multi-experience development platforms, 5G, and other technologies for 2025. They are also investing in cybersecurity/information security, application modernization, business intelligence/data analytics, cloud platforms, and enterprise resource planning. This is in line with our findings, which indicate that the healthcare entity is investing in new technologies but is not yet fully aligned with its strategic goals.

The pace of digital transformation in healthcare provider organizations is accelerating, and CIOs and technology executives are anticipating increased investments in IT to achieve their organizations' strategic objectives. However, they are also facing challenges in establishing a shared enterprise digital vision.

Here are some specific challenges that CIOs and technology executives are facing:

- *Reconciling new technologies with existing enterprise strategies*. New technologies like AI/ML offer transformational clinical and business benefits, but they can be difficult to integrate with existing enterprise systems.

- *Achieving consensus and developing a clear action plan*. It can be difficult to get everyone on the same page regarding the organization's digital transformation strategy.

Despite these challenges, the future of healthcare is digital. CIOs and technology executives who can successfully navigate these challenges will be well-positioned to lead their organizations into the digital age.

In addition to the above, we believe that the Luftman instrument remains a relevant tool for IT governance and digital transformation initiatives within organizations. It can be used to diagnose the maturity of strategic alignment, to develop a roadmap for improvement, and to complement initiatives such as digital maturity models (DMM) (Nebati *et al*., 2023) or readiness assessments (Silva *et al*., 2022).

We also believe that it is important to develop methods and actions aimed at creating and maintaining a close and harmonic relationship between business and IT strategies. This will be essential for healthcare organizations to succeed in the digital age.

We hope that our findings will be useful to healthcare organizations seeking to improve their strategic alignment and digital transformation efforts.

CRediт author statement

All authors have participated in a) conception and design or data analysis and interpretation; (b) drafting the article or revising it critically for important intellectual content; and (c) approving of the final version of the manuscript.

References

Al-Surmi, A., Cao, G., and Duan, Y. (2020). The impact of aligning business, IT, and marketing strategies on firm performance. *Industrial Marketing Management*, *84*, 39-49. https://doi.org/10.1016/j.indmarman.2019.04.002

Avison, D., Jones, J., Powell, P., and Wilson, D. (2004). Using and validating the strategic alignment model. *The Journal of Strategic Information Systems*, *13*(3), 223-246. https://doi.org/10.1016/j.jsis.2004.08.002

Baker, J., and Singh, H. (2019). The roots of misalignment: Insights on strategy implementation from a system dynamics perspective. *Journal of Strategic Information Systems*, *28*(4), 101576. https://doi.org/10.1016/j.jsis.2019.101576

Benbya, D. H., and McKelvey, B. (2006). Using coevolutionary and complexity theories to improve IS alignment: A multi-level approach. *Journal of Information Technology, 21*(4), 2000080. https://doi.org/10.1057/palgrave.jit.2000080

Berberat, S., and Baudet, C. (2019). *Assessing a business software application using strategic IT alignment factors: A new way for IS evaluation?* [Conference presentation]. ICEB 2019, Newcastle, UK. https://aisel.aisnet.org/iceb2019/19

Bernal-Acevedo, O., and Forero-Camacho, J. C. (2011). Sistemas de información en el sector salud en Colombia. *Gerencia y Políticas de Salud*, *10*(21), 21. https://doi.org/10.11144/Javeriana.rgsp10-21.siss

Bernat, S., and Karabag, S. F. (2019). Strategic alignment of technology: Organising for technology upgrading in emerging economy firms. *Technological Forecasting and Social Change*, *145*, 295-306. https://doi.org/10.1016/j.techfore.2018.05.009

Bhattacharya, P. (2017). Modelling strategic alignment of business and IT through enterprise architecture: Augmenting Archimate with BMM. *Procedia Computer Science*, *121*, 80-88. https://doi.org/10.1016/j.procs.2017.11.012

Bhattacharya, P. (2018). Aligning enterprise systems capabilities with business strategy: An extension of the strategic alignment model (SAM) using enterprise architecture. *Procedia Computer Science*, *138*, 655-662. https://doi.org/10.1016/j.procs.2018.10.087

Canhoto, A. I., Quinton, S., Pera, R., Molinillo, S., and Simkin, L. (2021). Digital strategy aligning in SMEs: A dynamic capabilities perspective. *The Journal of Strategic Information Systems*, *30*(3), 101682. https://doi.org/10.1016/j.jsis.2021.101682

Castrillón, H. Y., González, C., and López, D. M. (2012). Modelo arquitectónico para interoperabilidad entre instituciones prestadoras de salud en Colombia. *Revista Ingeniería Biomédica*, *6*(12), 3-13, https://dialnet.unirioja.es/servlet/articulo?codigo=4226425

Chan, Y. E. (2008). Why haven't we mastered alignment? The importance of the informal organization structure. *MIS Quarterly Executive*, *1*(2), 2.

Chan, Y. E., Huff, S. L., Barclay, D. W., and Copeland, D. G. (1997). Business strategic orientation, information systems strategic orientation, and strategic alignment. *Information Systems Research*, *8*(2), 125-150. https://doi.org/10.1287/isre.8.2.125

Chan, Y. E., and Reich, B. H. (2007). IT alignment: What have we learned? *Journal of Information Technology*, *22*(4), 297-315. https://doi.org/10.1057/palgrave.jit.2000109

Chi, M., Huang, R., and George, J. F. (2020). Collaboration in demand-driven supply chain: Based on a perspective of governance and IT-business strategic alignment. *International Journal of Information Management*, *52*, 102062. https://doi.org/10.1016/j.ijinfomgt.2019.102062

Departamento Administrativo Nacional de Estadística (DANE) (n.d.). *Censo nacional de población y vivienda 2018*. https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/censo-nacional-de-poblacion-y-vivenda-2018/informacion-tecnica

Darii, H., Laval, J., Botta-Genoulaz, V., and Goepp, V. (2020). *Measurement of the Business/IT Alignment of Information Systems* [Conference presentation]. International Conference on Information Systems, Logistics, & Supply Chain, Austin, TX, USA. https://publis.icube.unistra.fr/docs/14252/ILS2020_final_HD_JL_VBG_VG.pdf

de Haes, S., and van Grembergen, W. (2009). An exploratory study into IT governance implementations and its impact on business/IT alignment. *Information Systems Management*, *26*(2), 123-137. https://doi.org/10.1080/10580530902794786

Duhamel, F., Gutiérrez-Martínez, I., Picazo-Vela, S., and Luna-Reyes, L. F. (2021). Strategic alignment, process improvements and public value in public-private IT outsourcing in Mexico. *International Journal of Public Sector Management*, *34*(5), 489-507. https://doi.org/10.1108/IJPSM-07-2020-0183

Gartner (n.d.-a). *Infographic: top priorities, technologies and challenges for healthcare providers in 2023*. https://www.gartner.com/document/4020356?ref=solrResearch&refval=406134615&

Gartner (n.d.-b). *It's time for healthcare delivery organizations to adopt a digital-first strategy*. https://www.gartner.com/document/3988796?ref=solrResearch&refval=406134858&

Ghonim, M. A., Khashaba, N. M., Al-Najaar, H. M., and Khashan, M. A. (2020). Strategic alignment and its impact on decision effectiveness: A comprehensive model. *International Journal of Emerging Markets*, *17*(1), 198-218. https://doi.org/10.1108/IJOEM-04-2020-0364

Gutiérrez, A., Mylonadis, C., Orozco, J., and Serrano, A. (2008). *Business-IS alignment: Assessment process to align IT projects with business strategy* [Conference presentation]. AMCIS 2008. http://aisel.aisnet.org/amcis2008/290

Gutiérrez, A., Orozco, J., and Serrano, A. (2009). Factors affecting IT and business alignment: A comparative study in SMEs and large organisations. *Journal of Enterprise Information Management*, *22*(1/2), 197.

Haghighi Rad, F., and Rowzan, S. M. (2018). Designing a hybrid system dynamic model for analyzing the impact of strategic alignment on project portfolio selection. *Simulation Modelling Practice and Theory*, *89*, 175-194. https://doi.org/10.1016/j.simpat.2018.10.001

Hart, M., and Burke, J. (2020). An exploratory study on the Devops IT alignment model. *Interdisciplinary Journal of Information, Knowledge & Management*, *15*, 127-154. https://doi.org/10.28945/4595

Henderson, J. C., and Venkatraman, N. (1993). Strategic alignment: Leveraging information technology for transforming organizations. *IBM Systems Journal*, *1*, 4.

Issa, A., Hatiboglu, B., Bildstein, A., and Bauernhansl, T. (2018). Industrie 4.0 roadmap: Framework for digital transformation based on the concepts of capability maturity and alignment. *Procedia CIRP*, *72*, 973-978. https://doi.org/10.1016/j.procir.2018.03.151

Luftman, J. (2000). Assessing business-IT alignment maturity. *Communications of the Association for Information Systems*, *4*(1), 14. http://aisel.aisnet.org/cais/vol4/iss1/14

Luftman, J., Lyytinen, K., and ben Zvi, T. (2017). Enhancing the measurement of information technology (IT) business alignment and its influence on company performance. *Journal of Information Technology*, *32*(1), 26-46. https://doi.org/10.1057/jit.2015.23

Luftman, J. (Ed.). (1996). *Competing in the Information Age: Strategic alignment in practice* (1st Edition). Oxford University Press.

Ministerio de Salud (2018). *Manual de acreditación salud ambulatoria y hospitalaria*. https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/VS/PSA/manual-acreditacion-salud-ambulatorio.pdf

Maciá Pérez, F., Berna Martinez, J. V., and Lorenzo Fonseca, I. (2021). Strategic IT alignment projects. Towards good governance. *Computer Standards & Interfaces*, *76*, 103514. https://doi.org/10.1016/j.csi.2021.103514

Miyamoto, M. (2019). *IT-business alignments among different divisions of Japanese corporations* [Conference presentation]. 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Okinawa, Japan. https://doi.org/10.1109/ICAIIC.2019.8669032

Nebati, E. E., Ayvaz, B., and Kusakci, A. O. (2023). Digital transformation in the defense industry: A maturity model combining SF-AHP and SF-TODIM approaches. *Applied Soft Computing*, *132*, 109896. https://doi.org/10.1016/j.asoc.2022.109896

Patterson, M. (2020). A structured approach to strategic alignment between business and information technology objectives. *South African Journal of Business Management*, *51*(1). https://doi.org/10.4102/sajbm.v51i1.365

Reich, B. H., and Benbasat, I. (2000). Factors that influence the social dimension of alignment between business and information technology objectives. *MIS Quarterly*, *24*(1), 81-113.

Sabherwal, R., and Chan, Y. E. (2001). Alignment between business and IS strategies: A study of prospectors, analyzers, and defenders. *Information Systems Research*, *12*(1), 11-33. https://doi.org/10.1287/isre.12.1.11.9714

Sabherwal, R., Sabherwal, S., Havakhor, T., and Steelman, Z. R. (2019). How does strategic alignment affect firm performance? The roles of information technology investment and environmental uncertainty. *MIS Quarterly, 43*(2), 453-474. https://doi.org/10.25300/MISQ/2019/13626

Secretaría de Salud del Meta (n.d.). *Análisis de situación de salud VIllavicencio 2019*. http://www.villavicencio.gov.co/Documents/AN%C3%81LISIS%20DE%20SITUACI%C3%93N%20DE%20SALUD%20VILLAVICENCIO%20A%C3%91O%202019.pdf

Sholihah, M., Maezono, T., Mitake, Y., and Shimomura, Y. (2019). PSS strategic alignment: Linking service transition strategy with PSS business model. *Sustainability*, *11*(22), 6245. https://doi.org/10.3390/su11226245

Silva, R. P., Saraiva, C., and Mamede, H. S. (2022). Assessment of organizational readiness for digital transformation in SMEs. *Procedia Computer Science*, *204*, 362-369. https://doi.org/10.1016/j.procs.2022.08.044

Strategic Direction (2019). Manufacturing capabilities and performance improvement: How strategic alignment can deliver a lasting competitive edge. *Strategic Direction*, *35*(5), 25-27. https://doi.org/10.1108/SD-02-2019-0031

Smaczny, T. (2001). Is an alignment between business and information technology the appropriate paradigm to manage IT in today's organisations? *Management Decision*, *39*(10), 797-802. https://doi.org/10.1108/EUM0000000006521

Tarafdar, M., and Qrunfleh, S. (2009). IT-Business Alignment: A Two-Level Analysis. *Information Systems Management*, *26*(4), 338-349. https://doi.org/10.1080/10580530903245705

Torres-Moreno, M. E., and Aponte-Melo, J. H. (2021). Assessing business-IT alignment maturity at a Colombian university. *Journal of Cases on Information Technology* (*JCIT*), *23*(4), 1-22. https://doi.org/10.4018/JCIT.20211001.oa8

Lisienkova, T. S. (2017). Evaluation of Business and it strategic alignment maturity in Russian companies. *Вестник Южно-Уральского Государственного Университета. Серия: Компьютерные Технологии, Управление, Радиоэлектроника*, *17*(2), 2. https://cyberleninka.ru/article/n/evaluation-of-business-and-it-strategic-alignment-maturity-in-russian-companies-1

Yin, R. K. (2010). *Qualitative research from start to finish* (1st ed.). The Guilford Press.

Zee, J. T. M. V. D., and Jong, B. D. (1999). Alignment is not enough: Integrating business and information technology management with the balanced business scorecard. *Journal of Management Information Systems*, *16*(2), 137-158. https://doi.org/10.1080/07421222.1999.11518249

Zeebroeck, N. V., Kretschmer, T., and Bughin, J. (2023). Digital "is" strategy: The role of digital technology adoption in strategy renewal. *IEEE Transactions on Engineering Management*, *70*(9), 3183-3197. https://doi.org/10.1109/TEM.2021.3079347

Zhu, X., Ge, S., and Wang, N. (2021). Digital transformation: A systematic literature review. *Computers & Industrial Engineering*, *162*, 107774. https://doi.org/10.1016/j.cie.2021.107774

Research Article/Engineering Education

http://doi.org/10.15446/ing.investig.109389

An actionable learning path-based model to predict and describe academic dropout

Un modelo accionable basado en el camino de aprendizaje para predecir y describir la deserción académica

Cristian Olivares-Rodríguez[1], Pedro Manuel Moreno-Marcos[2], Eliana Scheihing Garcia[3], Pedro J. Muñoz-Merino[4], and Carlos Delgado-Kloos[5]

ABSTRACT

The prediction and explainability of student dropout in degree programs is an important issue, as it impacts students, families, and institutions. Nevertheless, the main efforts in this regard have focused on predictive power, even though explainability is more relevant to decision-makers. The objectives of this work were to propose a novel explainability model to predict dropout, to analyze its descriptive power to provide explanations regarding key configurations in academic trajectories, and to compare the model against other well-known approaches in the literature, including the analysis of the key factors in student dropout. To this effect, academic data from a Computer Science Engineering program was used, as well as three models: (i) a traditional model based on overall indicators of student performance, (ii) a normalized model with overall indicators separated by semester, and (iii) a novel configuration model, which considered the students' performance in specific sets of courses. The results showed that the configuration model, despite not being the most powerful, could provide accurate early predictions, as well as actionable information through the discovery of critical configurations, which could be considered by program directors could consider when counseling students and designing curricula. Furthermore, it was found that the average grade and rate of passed courses were the most relevant variables in the literature-reported models, and that they could characterize configurations. Finally, it is noteworthy that the development of this new method can be very useful for making predictions, and that it can provide new insights when analyzing curricula and and making better counseling and innovation decisions.

Keywords: academic trajectory, student model, dropout, explainability, curriculum analysis

RESUMEN

La predicción y explicabilidad de la deserción estudiantil en programas académicos es un asunto importante, pues impacta a estudiantes, familias e instituciones. Sin embargo, los principales esfuerzos en este sentido se han centrado en el poder predictivo, aunque la explicabilidad es más relevante para los tomadores de decisiones. Los objetivos de este trabajo fueron proponer un modelo novedoso de explicabilidad para predecir la deserción, analizar su poder descriptivo para proporcionar explicaciones sobre configuraciones clave en trayectorias académicas y comparar el modelo con otros enfoques bien conocidos en la literatura, incluyendo el análisis de los factores clave en la deserción estudiantil. Para ello, se utilizaron datos académicos de un programa de Ingeniería en Informática, así como tres modelos: (i) un modelo tradicional basado en indicadores generales de rendimiento estudiantil, (ii) un modelo normalizado con indicadores generales separados por semestre y (iii) un modelo de configuración novedoso que considera el rendimiento de los estudiantes en conjuntos específicos de cursos. Los resultados mostraron que el modelo de configuración, a pesar de no ser el más poderoso, podría proporcionar predicciones tempranas precisas, así como información accionable a través del descubrimiento de configuraciones críticas, las cuales podrían ser consideradas por los directores de programa al asesorar a los estudiantes y diseñar planes de estudio. Además, se encontró que la nota promedio y la tasa de cursos aprobados fueron las variables más relevantes en los modelos reportados en la literatura, y que estas podrían caracterizar configuraciones. Finalmente, es notable que el desarrollo de este nuevo método puede ser muy útil para hacer predicciones y que puede proporcionar nuevas perspectivas al analizar planes de estudio y al tomar mejores decisiones de asesoramiento e innovación.

Palabras clave: trayectoria académica, modelo de estudiantes, deserción, explicabilidad, análisis curricular

Received: June 7th 2023
Accepted: September 30th 2023

Introduction

Year by year, university decision-makers counsel thousands of students who are not able to manage the contents of their courses, with some of them even dropping out of courses and/or the whole program. Statistics report that about 30-35% of the students drop out of a degree (Paura and Arhipova, 2014; Vidal *et al.*, 2022).

Particularly in Chile, many students drop out, and many of those who do not spend a lot of time finishing their degree, especially their first two years (commonly known as *bachillerato*) (Donoso *et al.*, 2013). There are many reasons for dropping out, such as prior academic preparedness (Smith and Naylor, 2001; Carvajal *et al.*, 2018), poor academic results, lack of funding, and loss of interest

[1] Facultad de Ingeniería, Universidad Alberto Hurtado and Member of the Centro de Estudios en Ciencia, Tecnología y Sociedad, CECTS, Chile, colivares@uahurtado.cl
[2] Universidad Carlos III de Madrid, Spain, pemoreno@it.uc3m.es
[3] Universidad Austral de Chile, Chile, escheihi@uach.cl
[4] Universidad Carlos III de Madrid, Spain, pedmume@it.uc3m.es
[5] Universidad Carlos III de Madrid, Spain, cdk@it.uc3m.es

(Breier, 2010; Li and Carroll, 2020). Therefore, decision-makers require an early understanding { not just a numerical prediction { of why students fail during the program in order to enhance their learning in the long term. They also need clear guidelines to act in these situations.

To understand students' academic behavior, data mining techniques can be used (Palacios *et al.*, 2021; Chung and Lee, 2019). The development of predictive models can provide significant benefits to different stakeholders. Students improve in self-reflection regarding their learning, which facilitates their decision-making. Teachers focus on the students at risk and try to adapt the course and methodology. Finally, university decision-makers can redesign curricula to balance workload, improve personalized counseling, and potentially enhance the learning process.

Some works have attempted to analyze dropout as an academic behavior. Bottcher *et al.* (2020) found that the number of credits passed in the first semester and the frequent changes made to bachelor's degree programs are significant indicators. Hutt *et al.* (2018) analyzed other variables such as personal factors, work experience, and academic tests. However, one of the less studied factors that might influence dropout is the set of courses that a student takes, hereafter called *configurations* (Hutt *et al.*, 2018). If predictions are made with regard to these configurations, a lot of benefits can be reaped. Moreover, when managers decide whether a learner can take a course, they can use dropout information to drive their counseling, leading students into an enhanced learning environment.

The analysis of students in the field of education can be carried out at the course level (*e.g.*, predicting student dropout from a course) or at the program level (*e.g.*, predicting student dropout from a degree). At the course level, Jiang and Li (2017) predicted dropout by using ensemble learning methods. Meanwhile, Moreno-Marcos *et al.* (2019) predicted student success in an admission test based on interactions in an edX-based blended program. There are also many contributions to prediction in online programs (Gardner and Brooks, 2018; Moreno-Marcos *et al.*, 2018; Kang and Wang, 2018; Jin, 2021; Mubarak *et al.*, 2021), as well as in blended contexts, such as the in-session model proposed by Rzepka *et al.* (2022).

Regarding the program level, there have also been plenty of relevant works. Yu *et al.* (2021) analyzed how important it is to include sensitive attributes (*e.g.*, gender, underrepresented minorities, *etc.*) in student modeling. They concluded that these attributes can only provide a marginal improvement and do not significantly impact predictive power. Dekker *et al.* (2009) predicted dropout after the first semester using several algorithms, concluding that academic data offer higher predictive power than pre-university features. Furthermore, Delen (2011) found that, while the most important variables are academic in nature, financial variables are also relevant predictors. Similarly, Quadri and Kalyankar (2010) found that low-income levels have a strong influence on dropout, while demographics are irrelevant factors.

Lázaro Alvarez *et al.* (2020) found that accuracy significantly improves after the first semester, suggesting that early predictions can be made. Berens *et al.* (2018) also analyzed this issue, reporting that accuracy could reach 79-85% after the first semester and 90-95% after the fourth one, which also suggests the possibility of early prediction. In another study (Gašević *et al.*, 2016), different values were obtained for different degrees, indicating the importance of course context and generalizability analysis. Meanwhile, Wagner *et al.* (2020) presented a preliminary study and reported good cross-program models, and Panagiotakopoulos *et al.* (2021) provide an early prediction method for massive open online courses (MOOC). Nevertheless, further research is needed in this direction, as is described in Moreno-Marcos *et al.* (2019).

This work proposes an innovative method to analyze dropout as academic behavior via variables related each set of courses taken simultaneously by a student. The aim is for this model to offer information about curricula, *i.e.*, which configurations can be taken by a student so that dropout can be reduced. This is an aspect that has not been considered in previous works. This paper contributes by analyzing how critical configurations can be obtained and how they can be used by decision-makers to analyze student behavior. Finally, this paper addresses relevant topics in the literature, such as temporal analysis to discover the moment at which early predictions are accurate enough to be used for causing a positive effect on students' learning.

This paper aims to analyze different models for predicting dropout, including a novel proposal based on configurations. To this effect, several specific objectives have been defined:

- Objective 1: To propose a prediction model based on course configurations
- Objective 2: To compare the capabilities of different models for predicting dropout
- Objective 3: To analyze the suitability of the configurations-based model to detect a set of courses that are likely or unlikely to produce or prevent dropout
- Objective 4: To analyze the key factors that affect dropout in different models

Methodology

Dropout models

This section describes the models used to analyze dropout. First, a novel model is proposed, aming to describe student trajectories while considering the set of courses (configurations) taken by students throughout their degree (objective 1). Additionally, we outline two traditional models based on student performance variables for comparison purposes.

Configuration model

Decision-makers analyze student behavior by exploring their academic trajectories, in order to support their advice and innovations. This work proposes a novel dropout analysis model based on the temporal course configurations that students take simultaneously, rather than considering opaque student behavior in integrated variables, as in traditional models.

The model is operationalized through some mathematical expressions. A degree is implemented via a program

curriculum (P), which is composed of a set of semester-based courses (α). Every student draws their academic trajectory (t) with these sets of courses, also defined as *configurations* (c). For example, a student i takes a first configuration c_1^i during their first semester. Afterwards, he takes a second configuration c_2^i, and so on. The resulting trajectory (t^i) is presented in Equation 1, which contains the ordered sequence of configurations. It is noteworthy that, when a student drops out, it implies a trajectory that does not lead to program completion, which must be analyzed nonetheless.

$$t^i = c_1^i \Rightarrow c_2^i \Rightarrow \cdots \Rightarrow c_p^i \tag{1}$$

Every academic program P has a finite set of configurations c, and each student in a program can select only one configuration per semester. Thus, it is possible to model a program through a trajectory matrix (Mt), which relates students with configurations, as shown in Equation 2.

$$M_t^p = \begin{bmatrix} W_{1,1} \cdots W_{1,N} \\ \cdots \\ W_{M,1} \cdots W_{M,N} \end{bmatrix} \tag{2}$$

In this matrix, each column represents every possible configuration c, and each row represents a student i who is enrolled in the program P. The content of each cell represents a measurement based on several weights ($w_(i, j)$), indicating the performance of the student i in the configuration c_j. The value of this measurement is 0 when the student does not select a configuration. Each row is a vector that describes the performance of a student throughout their academic trajectory, considering all configurations. The measurement w, which captures most of a student's behavior, is presented in the subsection called *Variables and techniques*.

Traditional model

This model considers a set of global variables $[X1, X2, \cdots, Xn]$, which are obtained by integrating data over the entire analysis period. Regardless of the length of the student's trajectory, the number and nature of the variables will be the same, even though they will be computed with more or less data. If there are N independent variables, one dependent variable, and M students, the matrix with the training set will always have $(N + 1) * M$ elements. An example of this matrix is presented in Equation 3, where $X_{i,j}$ represents the variable j of student i, and Y_i denotes their dependent variable (dropout).

$$Trad = \begin{bmatrix} X_{1,1} \cdots X_{1,N} Y_1 \\ \cdots \\ X_{M,1} \cdots X_{M,N} Y_M \end{bmatrix} \tag{3}$$

This is regarded as a traditional model because most contributions use a similar approach when developing predictive models. Although this approach can be effective in many scenarios, one of its drawbacks is that academic trajectories become opaque.

Normalized model

This model aims to enhance traditional approaches by considering the evolution of a student throughout the degree. To this effect, this model calculates the independent variables for each semester. If there are N variables and the student has taken six semesters, there will be $6 * N$ independent variables. As some students may have taken more semesters than others, the variables of the academic terms that the student has not yet taken are set as 0, in order to make the matrix consistent.

If there are N independent variables, one dependent variable, S possible semesters, and M students, the number of elements in the matrix will be $(N * S + 1) * M$. This number increases with the number of semesters, providing temporal information about a student's performance/behavior. An example of the matrix used for the training set is presented below. Independent variables are named $X_{i,j,k}$, where j indicates the number of the variable (from 0 to N), k the semester (from 0 to S), and i the number of the student (from 0 to M). The dependent variables are named Y_k, as they only depend on the student, as shown in Equation 4.

$$Norm = \begin{bmatrix} X_{1,1,1} \cdots X_{1,2,1} \cdots X_{1,N,S} Y_1 \\ \cdots \\ X_{M,1,1} \cdots X_{M,2,1} \cdots X_{M,N,S} Y_M \end{bmatrix} \tag{4}$$

By following this approach, it is possible to observe how a student evolves throughout a degree program, but their decisions regarding the configurations taken are still opaque and differ from one student to another. For example, student A in semester 6 might have taken a different configuration from student B in the same semester. In this case, the model is said to be *normalized*, since the variables for each semester are computed independently of the courses taken; they are normalized to the number of semesters.

Dataset

This study has been conducted using academic data from one degree program (Computer Science Engineering) at Universidad Austral de Chile (UACh), in the context of the Erasmus+ LALA project. The data collection period goes from 2011 to 2017. This period defines a complete cohort of students attending the same academic program (without changes) and sets of courses. The dataset is composed of records of 479 students, with 50.73% of student dropout. The mean number of semesters per student is 5.54 (SD: 3.76) and the mean number of courses per semester is 5,77 (SD: 2.78). At the end of every semester, each student must decide which set of courses they will take during the next term while considering several program constraints and their trajectory. Trajectories become more and more diverse as students advance through the program, as more decisions regarding courses become necessary.

The data were obtained only from the SIS (Student Information System). Only academic data were considered (*e.g.*, grades in different courses). This fact, despite being a limitation, can showcase the potential of dropout analysis without using many different data types (*e.g.*, without interaction logs of learning platforms). If models can work accurately enough with SIS data, institutional adoption will be easier. In this case, the data are in the form of a table with students' grades for different courses.

Table 1. Example of the data collected from UACh

ID	Course	Year	Semester	Concept	Grade
ID1	Course A	2011	1	TAKEN	5.2
ID1	Course B	2011	1	TAKEN	2.5
ID1	Course C	2011	2	CANCELED	0.0
ID2	Course A	2012	1	VALIDATED	6.1

Source: Authors

As shown in Table 1, there is one column for the student (ID), another for the course name (course), and two columns to identify the period during which the course was taken (year and semester, which can only be 1 or 2). The concept of a course indicates whether the student took the course (taken) or if it was canceled (taken and not finished, withdrawn), validated (taken in another degree program), recognized (taken at another institution), or retrieved from another curriculum (when the student has changed their curriculum). An enrolled course comprises all these situations. The last column indicates the student grade from 0 to 7, with 4,0 the passing score.

Dropout setup

To develop the predictive model, it was necessary to define how to determine dropout. It is worth mentioning that a ground truth was not available to train the models. As only data from 2011 to 2017 were available, a first criterion was defined to filter out students who started long before 2011 or just before the end of the data collection period.

A degree is composed of 11 semesters (including the final project). However, the first four semesters (*bachillerato*) only contain basic training courses (*e.g.*, Algebra, Programming, Communication Skills, *etc.*). In general, according to historical data, students who finish the *bachillerato* have a high probability of finishing the degree. To increase the number of cohorts and obtain sufficient data to train the model, we considered that a student has not dropped out when they have a non-*bachillerato* configuration, *i.e.*, a set of courses that do not contain any course from the first two years. This means that the students have finished the first two years in their program; otherwise, they will report at least one course from these years in their configuration. In contrast, if a student only has *bachillerato* configurations (a set of courses with at least one belonging to the first two years), it means that they have not finished the first four semesters. In some cases, these students were discarded. Note that they might not have the four *bachillerato* configurations because they might have enrolled in the program recently.

Several filters to classify and discard students were considered. First, the students who enrolled before 2011 were removed because their initial configurations were missing, as this was the start of the analyzed period. Afterwards, the students with at least one non-*bachillerato* configuration were labeled as *no dropout* since, as explained above, those who finish *bachillerato* are very likely to finish the degree. Finally, for students who did not have non-*bachillerato* configurations, it was necessary to distinguish those who were currently attending the *bachillerato* from those who had dropped out. In this vein, students were considered dropouts if they were not enrolled in the last year of the data collection period (2017), since they had started the *bachillerato* but had not finished it; otherwise, they would have been enrolled in the program and would later take non-*bachillerato* configurations. Those who were enrolled in 2017 were discarded, as they were attending the *bachillerato* at that moment and there was no dropout information. Figure 1 shows the flowchart of the process for determining dropout with the above-mentioned criteria. A sample of N = 270 students was obtained (after discarding students) for the analysis.

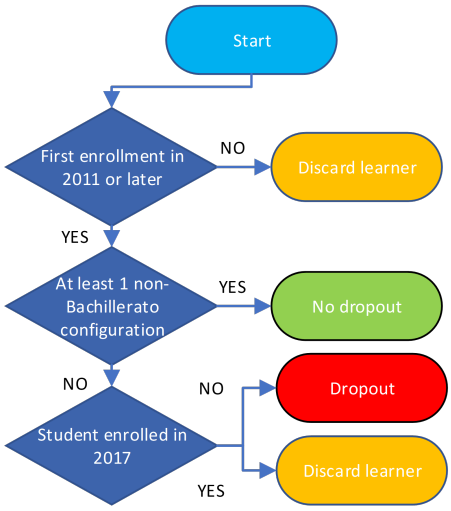


Figure 1. Flowchart with the criteria to flag students as dropouts
Source: Authors

Variables and techniques

To carry out the analyses, it was important to define the set of variables that could be obtained from the previously described dataset. Table 2 presents the full list of variables considered in this study. Some variables were directly taken from the indicators used in the LALA project (Muñoz-Merino *et al.*, 2020). Some variables were used in all models, although others were only tested in the configuration model, which required further tuning given the need to define a weight function.

Table 2. Variables used for the traditional (T), normalized (N), and configurations (C) models

Variable	Model	Description
avg_grade	T, N, C	It indicates the average grade of the enrolled courses.
r_passed	T, N, C	It indicates the relationship between passed and enrolled courses.
r_taken	T, N	It indicates the relationship between the courses taken and enrolled.
r_takreg	C	It indicates the relationship between taken and recognized courses and those enrolled.
r_cancel	T, N	It indicates the relationship between canceled and enrolled courses.
repeat	T, N	It indicates the relationship between the number of courses a student takes and the total number of attempts. If a student takes three courses for the first time and one for the second, the value is 4/(1+1+1+2)=0,8.

Source: Authors

The T and N models use these variables directly. Meanwhile, the C model aggregates such variables via a performance function w, assessing every student's behavior in every particular configuration. To compute this function,we used the formula presented in Equation 5.

$$w^{i,j} = (avg\_grade + r\_takreg) * r\_passed \quad (5)$$

In this work, the *caret* library of the R open-source software was used to develop the predictive models. Particularly, the following algorithms were used: i) random forests (RF), ii) the generalized linear model (GLM), iii) support vector machines (SVMs), iv) decision trees (DT), and (5) single-hidden-layer neural networks (NNs). The hyperparameters of these models were discovered during the training phase. In addition, ten-fold cross-validation was used to validate the results (five-fold cross-validation was also used to compare the results, although no significant changes were observed, so only those related to 10-fold cross-validation are presented), and the area under the curve (AUC) was the metric computed to evaluate the predictive power of the models, as has been done in previous works (Pelánek, 2015; Jeni *et al.*, 2013).

Results

First, this section presents an analysis of dropout behavior, comparing the three models and including a temporal analysis (related to objective 2). Next, we delve into the configuration model to determine whether it provides descriptive value regarding the academic trajectories and configurations that are more or less likely to enhance the learning context (related to objective 3).

Predictive power

We analyzed how early it was possible to predict dropout and how the aforementioned predictive models behaved with small amounts of data. The dropout models by only using data related to the *bachillerato* (four semesters). The data context was named *xS* (*e.g.*, 1S, 2S, *etc.*), where x indicates the number of semesters considered for developing the model.

Table 3. Temporal analysis of dropout prediction (results expressed in AUC)

Alg	Traditional				Normalized				Configurations			
SEM.	1S	2S	3S	4S	1S	2S	3S	4S	1S	2S	3S	4S
LR	**0.92**	0.92	**0.93**	**0.95**	**0.92**	**0.94**	0.95	0.96	0.73	0.69	0.75	0.54
DT	0.78	0.85	0.84	0.89	0.78	0.79	0.90	0.86	0.77	0.77	0.77	0.78
RF	0.90	**0.93**	0.92	0.94	0.90	0.92	0.94	**0.97**	0.77	0.77	0.80	0.81
SVM	0.88	0.90	0.91	0.94	0.88	0.91	0.95	**0.97**	**0.85**	**0.86**	**0.88**	**0.82**
NN	0.91	0.92	**0.93**	**0.95**	**0.92**	**0.94**	**0.96**	**0.97**	0.79	0.71	0.78	0.81

Source: Authors

Table 3 shows the strength of the proposed model, which performs well (AUC≥0.8) for every semester and is useful to analyze critical configurations. This is not possible in the other models. The N and T models provide high predictive power but no actionable information. Particularly, the N model can yield the most accurate results. This implies that separating the variables by semester can provide further information, despite increasing dimensionality, as performance can vary over time. This was checked with the correlation between variables in different semesters. In the case of the average grade, the correlations were between 0.56 and 0.81 (with the latter being the highest correlation between the same variable in different semesters), which shows that there is a high correlation, but that further information can still be added.

Apart from that, a surprising result is that the performance of the C model deteriorates after the third semester, while the other two models improve, which is the usual and expected behavior when adding new data. A possible reason for this is that the number of configurations to be considered in the model increases as new semesters are incorporated. This increases the number of variables, and the model becomes sparser, losing its predictive power as reported in Kohavi and John (1997) for similar situations. This implies that the C model behaves better in the first stages, when the number of configurations is manageable. The positive aspect is that these stages are the most relevant, as most dropouts occur early. Nevertheless, the C model can be used to gather actionable information about critical configurations to help decision-makers and validate the predictions of other algorithms.

In terms of anticipation, the results show that the predictive power of all models is very good since the first semester, which is consistent with Lázaro Alvarez *et al.* (2020). This means that it is possible to obtain accurate early predictions with the first results of the students, which is useful to anticipate performance and support academic decisions.

As for the algorithms, NN was the most consistent for both the T and N models, although SVM stood out in the C model, as happened in Tekin (2014). Nevertheless, the differences were not great. There was a predominance of SVMs in the C model, perhaps due to their capability to handle problems with high dimensionality (Bersimis and Varlamis, 2019).

In summary, the proposed models can provide accurate predictions from the beginning, so it is possible to anticipate student behavior with regard to the issue under study. Nevertheless, the accuracy of the N model stands out, and the C model provides promising results and new insights, since it does not only analyze overall performance, but also that in specific sets of courses.

Critical configurations

It is particularly relevant to provide actionable information about students, and the C model was designed to track both temporal relations and performance in course configurations at any given moment. Thus, this model aids in determining the most critical course configurations.

The importance of the configurations was computed through the Mean Decrease Gini, which is commonly used to evaluate the importance of features (Louppe *et al.*, 2013). This metric was computed with data from one to four semesters. The three most relevant configurations (among 2127 possible configurations) for each semester are shown in Table 4.

In light of the above, decision-makers must pay attention to the most relevant configurations and, more precisely, to the sets of courses appearing at the top in more than one of the rankings provided by the classifiers. Particularly, configurations 29, 30, 31, 46, and 97 are the most critical.

Table 4. Most critical configurations

SEM	Most relevant	2nd most relevant	3rd most relevant
1S	S29	S97	S46
2S	S29	S30	S97
3S	S29	S30	S31
4S	S30	S29	S31

Source: Authors

These configurations have the following in common: (i) they contain initial programming courses (five out of five) or initial mathematics courses, such as Algebra or Calculus (four out of five); and (ii) they have a large number of courses per semester (six or more). The negative results of the students taking these configurations implies that programming and mathematics courses are becoming harder to pass. This is consistent with other works, such as that of Figueiredo and García-Peñalvo (2021), who highlighted the difficulty of programming courses. Moreover, this could imply that students who take many courses can have a very high workload and may have difficulties in dealing with it, as was also reported in other works (Radovanović *et al.*, 2021). To prevent this, decision-makers (*e.g.*, program directors) should not allow students to take configurations with such a high workload, in addition to preventing them from taking specific courses in the same semester, as these configurations may lead to dropout. In any case, it is important to note that decision-makers should interpret this actionable information and consider possible special cases or circumstances to support their decisions.

Key factors affecting dropout

The models can be further analyzed to discover what features are more relevant or actionable for program directors. The importance of the variables was initially computed for the T and N models using the same methodology presented in the previous section. The results shown in Figures 2 and 3 indicate that the most important variables for the T model are the average grade, as reported by Delen (2011), followed by the ratio of passed courses. These two variables are very clear indicators of performance, since students who fail repeatedly or have low grades are more likely to drop out. It is also interesting that the repeat variable becomes relevant after the second semester. In the first two semesters, it is not important, as students do not repeat courses, but once they start to do so, this variable takes on great significance.

In the N, similar variables are relevant, as the average grade and the ratio of passed courses stand out. Nevertheless, additional patterns can be found when considering individual variables. For example, the average grades of the first semesters became very relevant over time. The importance of the average grade of the first semester does not significantly differ from that of the average grade of the second semester, also showing high values in the third semester. In contrast, variables from previous semesters become less relevant in the fourth semester. Like the average grade, the ratio of passed courses shows a similar behavior, although it is surprisingly the most relevant variable in the fourth semester. The repeat variable shows a similar behavior to the T model, except in the fourth semester.

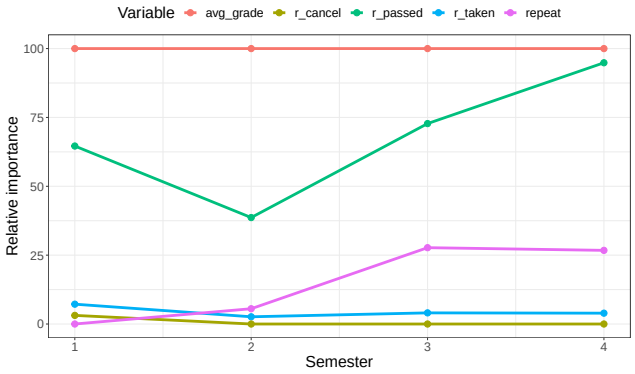


Figure 2. Evolution of the importance of variables in the T model
Source: Authors

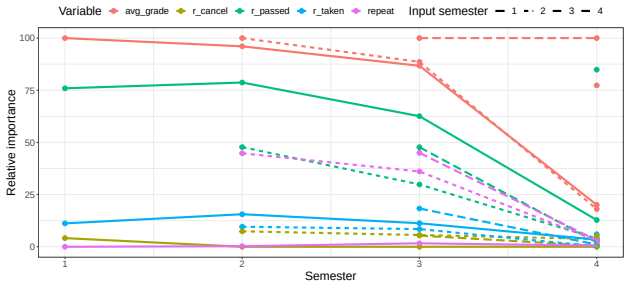


Figure 3. Evolution of the importance of variables in the N model
Source: Authors

Furthermore, it is interesting to see how the values of these relevant features are related to critical configurations. To analyze this issue, the top three features in the aforementioned models were computed at the configuration level, averaging the results of all the students who took each configuration. In this case, as knowing the dropout level was not necessary, all 695 students in the dataset were considered (we removed the filters to flag dropout presented in Figure 1). The results for the top 20 critical configurations of the first four semesters (there are 32 in total) are shown in Figure 4.

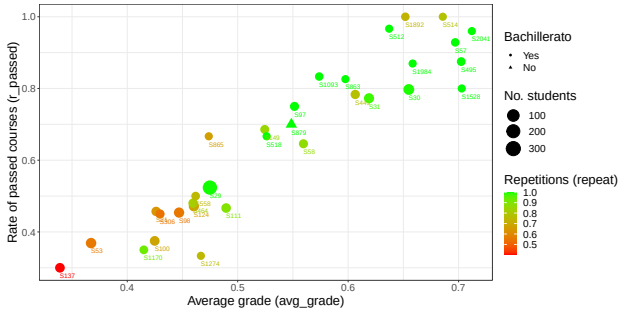


Figure 4. Characterization of the top critical configurations in the C model
Source: Authors

This Figure indicates that not all critical configurations are negative, but actionable. There are some configurations where most of the students obtain low grades and fail most of the courses (*e.g.*, S137) while there are others with very good results. This is an important finding, since program directors can not only detect configurations where students

may struggle, but also those where it might be easier to succeed.

Moreover, S29 is the most frequent configuration (320 students took it), and, although it does not contain repeated courses (repeat=1 means that all courses are taken for the first time, as shown in Table 2), many students get poor results. This actionable information serves to provide insights to program directors regarding the workload or the possible reasons behind taking this configuration.

Additionally, a hierarchical clustering analysis was conducted to discover whether there were any patterns or distinctions between the configurations. This was done by using the variables avg_grade, r_passed, r_taken, r_cancel, and repeat at the configuration level. The optimal number of clusters was computed via Silhouette analysis. Figure 5 shows the clusters obtained.

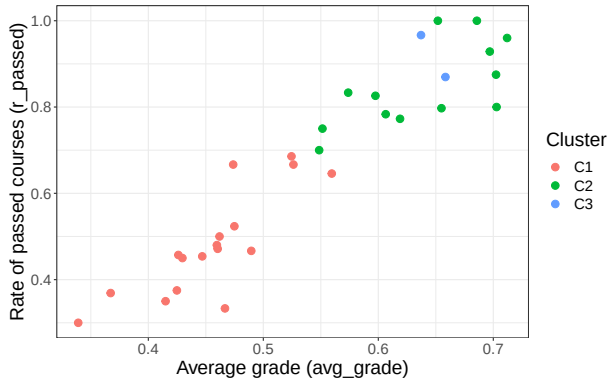


Figure 5. Hierarchical clustering of the top critical configurations
Source: Authors

Two clusters are relevant: the red one and the green one. The red cluster shows configurations with poor results and low average grades and rates of passed courses. Meanwhile, in the green cluster, students usually obtain good results and high average grades and rates of passed courses. The repeat variable is also usually high (students do not repeat courses). Therefore, red configurations should be avoided, and green ones may be encouraged by program directors. Finally, there is a special cluster (the blue one) with two configurations where all students validate all courses, which is not relevant in terms of student performance. An analysis of the the 17 red configurations reveals interesting facts. In six configurations (29, 149, 306, 464, 518, and 1170), all courses belong to the same semester in the program, but the average grade in most of the courses is below 4, implying that these configurations include difficult courses. In addition, there are seven configurations (53, 58, 98, 100, 111, 137, and 1274) where all courses belong to the same semester except one. In all those cases except one, most of the courses show an average grade below 4, implying that they are difficult configurations and the students may not perform as expected. Finally, there are four configurations (21, 124, 558, and 865) where students have courses from different semesters. In those cases, it is likely that student performance also affects the configuration. Further analysis should be conducted in this regard.

Finally, considering all the results and the actionable information provided by the C model, the flow of actions in an institution could be as follows:

1. Making predictions using the different models (as all of them are quite accurate).
2. Identifying the most critical configurations with the C model.
3. Computing the main features and identifying the cluster they belong to.
4. Program directors could review the predictions and support the students' decisions. To this effect, a dashboard could be implemented to provide this information, as in Guerra *et al.* (2019).

If these steps are taken, both descriptive and predictive information based on configurations could be used to provide better counseling to students, which may in turn result in better graduation rates.

Conclusions

In this article, a novel dropout analysis model was proposed as a method to support decision-makers during student counseling and/or curriculum innovations. Two models were compared against our proposal, upon the basis of students' performance in each specific set of courses. All models exhibited a high predictive power, although the normalized model stood out. Nevertheless, the configuration model (the proposed approach) implies a reasonable trade-off between prediction power and actionable information about critical configurations. The main findings are summarized below:

- The configuration model has a strong predictive power, particularly in the first semesters, so it can be used for early dropout prediction. However, it may face issues in later stages due to the diversity of configurations, as reported in other works (Kohavi and John, 1997), which implies too many features).
- Predictive power is very high for all studied models since the first semester, as was previously reported by Lázaro Alvarez *et al.* (2020) and by Berens *et al.* (2018) to a lesser extent. This means that it is possible to obtain accurate early predictions once the first results of the students are available.
- The configuration model aids in identifying critical configurations that program directors can consider in their counseling sessions. Program directors may warn students about negative configurations and recommend courses, providing another perspective, different from other traditional recommender systems, as is the case of Denley (2014)).
- An analysis of critical configurations shows that initial programming and mathematics courses may cause difficulties for students and configurations with more courses than expected (heavy workload). This is consistent with other works highlighting the difficulties of programming courses (Figueiredo and García-Peñalvo, 2021) or the students' issues with heavy workloads (Radovanović *et al.*, 2021).
- The most relevant variables in the models are the average grade of the courses, which is consistent with

Delen (2011); the rate of passed courses; and, to a lesser extent, a variable indicating if the student is repeating many courses. In addition, relevant variables are usually also relevant even in subsequent semesters (*e.g.*, the average grade in semester 1 was highly important in the first three semesters).

- The most relevant variables enabled a cluster analysis to identify positive and negative configurations.

In light of the above, the proposed model provides actionable information for decision-makers to enhance the learning context. If a dashboard with all this information is given, directors could easily identify students at risk and analyze their trajectories to better guide them in their learning path, which may in turn help to reduce dropouts. In addition, directors can further reflect on how the program is designed, so that the configurations for each semester are better balanced and the workload can be better managed by students.

One limitation of this proposal is that, as completing a degree takes several years and information from only the last seven cohorts (years) is available, data on completion are not available for some students. This means that the dropout definition (which is not unique, unlike a numerical grade, for example) had to be adapted, likely affecting the results. Furthermore, only data for one degree program were used, and more variability would be needed to compare different contexts and analyze the generalizability of the models. Nevertheless, these models can be easily adapted to other programs, although they may require training.

In future work, the addition of new variables, such as the student's financial situation, mental health, change of interest, personal reasons, and specific grades in critical courses, as well as the general features of the configurations (*e.g.*, popularity/frequency), may be relevant to the understanding of dropout and the improvement of the models. Further analyses involving interpretable machine learning tools such as those presented in Nagy and Molontay (2023) could be performed in order to better understand the models. Moreover, putting predictions in a dashboard and evaluating how programs use them would be a relevant contribution. In this vein, a scheme could be developed to ensure that directors get proper feedback and provide adequate counsel to students. This scheme would also allow students to provide feedback if they eventually drop out, and it would help to detect special situations that are useful for enhancing the models and identifying critical configurations. With regard to these configurations, more research is needed in order to obtain a better understanding and to determine whether poor results are related to course difficulty or student performance.

Acknowledgements

This work was partially funded by the LALA project (grant 586120-EPP-1-2017-1-ES-EPPKA2-CBHE-JP), by FEDER/Ministerio de Ciencia, Innovación y Universidades – Agencia Estatal de Investigación through project H2O Learn (grant PID2020-112584RB-C31), and by the Spanish Ministry of Science, Innovation, and Universities under an FPU fellowship (FPU016/00526). The LALA project has been funded with support from the European Commission. This publication only reflects the views of the authors, and the Commission and the Agency, as well as other funding entities, cannot be held responsible for any use of the information contained therein.

CRediT author statement

Cristian Olivares-RodrIguez: conceptualization, formal analysis, investigation, methodology, software, supervision, and writing (original draft). *Pedro Manuel Moreno-Marcos*: conceptualization, data curation, formal analysis, investigation, methodology, software, validation, visualization, and writing (original draft). *Pedro J. Muñoz-Merino*: conceptualization, formal analysis, funding acquisition, investigation, supervision, and writing (review and editing). *Eliana Scheihing Garcia*: conceptualization, investigation, methodology, and writing (review and editing). *Carlos Delgado Kloos* provided critical feedback.

Conflicts of interest

The authors declare no conflict of interest.

References

Berens, J., Schneider, K., Gortz, S., Oster, S., and Burghoff, J. (2018). *Early detection of students at risk–predicting student dropouts using administrative student data and machine learning methods*. CESifo Working Paper. Retrieved from https://ssrn.com/abstract=3275433

Bersimis, F. G., and Varlamis, I. (2019). Use of health-related indices and classification methods in medical data. In N. Dey (Ed.), *Classification techniques for medical image analysis and computer aided diagnosis* (pp. 31–66). Elsevier. https://doi.org/10.1016/B978-0-12-818004-4.00002-9.

Bottcher, A., Thurner, V., and Hafner, T. (2020, April 27-30). *Applying data analysis to identify early indicators for potential risk of dropout in cs students*. [Conference paper]. 2020 IEEE Global Engineering Education Conference, Porto, Portugal. https://doi.org/10.1109/EDUCON45650.2020.9125378.

Breier, M. (2010). *Student retention and graduate destination: Higher education and labour market access and success*. HSRC Press Cape Town.

Carvajal, C. M., González, J. A., and Sarzoza, S. J. (2018). Variables sociodemográficas y académicas explicativas de la deserción de estudiantes en la facultad de ciencias naturales de la universidad de playa ancha (chile). *Formación universitaria*, *11*(2), 3–12. http://dx.doi.org/10.4067/S0718-50062018000200003.

Chung, J. Y., and Lee, S. (2019). Dropout early warning systems for high school students using machine learning. *Children and Youth Services Review*, *96*, 346–353. https://doi.org/10.1016/j.childyouth.2018.11.030.

Dekker, G. W., Pechenizkiy, M., and Vleeshouwers, J. M. (2009, July 1-3). *Predicting students drop out: A case study*. [Conference paper]. 2nd International Conference on Educational Data Mining, Córdoba, Spain. https://www.educationaldatamining.org/EDM2009/uploads/proceedings/dekker.pdf.

Delen, D. (2011). Predicting Student Attrition with data mining methods. *Journal of College Student Retention: Research, Theory & Practice*, *13*(1), 17–35. https://doi.org/10.2190/CS.13.1.b.

Denley, T. (2014). How predictive analytics and choice architecture can improve student success. *Research & Practice in Assessment*, *9*, 61–69.

Donoso, S., Donoso, G., and Frites, C. (2013). La experiencia chilena de retención de estudiantes en la universidad. *Revista Ciencia y Cultura*, *17*(30), 141–171.

Figueiredo, J., and García-Peñalvo, F. (2021, October 26-29). *A tool help for introductory programming courses.* [Conference paper]. 9th International Conference on Technological Ecosystems for Enhancing Multiculturality, Barcelona, Spain. https://doi.org/10.1145/3486011.3486413.

Gardner, J., and Brooks, C. (2018). Student success prediction in moocs. *User Modeling and User-Adapted Interaction*, *28*, 127–203. https://doi.org/10.1007/s11257-018-9203-z.

Gašević, D., Dawson, S., Rogers, T., and Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, *28*, 68–84. https://doi.org/10.1016/j.iheduc.2015.10.002.

Guerra, J., Scheihing, E., Henríquez, V., Olivares-Rodríguez, C., and Chevreux, H. (2019, September 16-19). *TrAC: Visualizing students academic trajectories.* [Conference paper]. 14th European Conference on Technology Enhanced Learning, Delft, The Netherlands. https://doi.org/10.1007/978-3-030-29736-7_84.

Hutt, S., Gardener, M., Kamentz, D., Duckworth, A. L., and D'Mello, S. K. (2018, March 7-9). *Prospectively predicting 4-year college graduation from student applications.* [Conference paper]. 8th International Conference on Learning Analytics and Knowledge, Sydney, New South Wales, Australia. https://doi.org/10.1145/3170358.3170395.

Jeni, L. A., Cohn, J. F., and De La Torre, F. (2013, September 2-5). *Facing imbalanced data–recommendations for the use of performance metrics.* [Conference paper]. 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, Geneva, Switzerland. https://doi.org/10.1109/ACII.2013.47.

Jiang, F., and Li, W. (2017). Who will be the next to drop out? anticipating dropouts in moocs with multi-view features. *International Journal of Performability Engineering*, *13*(2), 201–210. https://doi.org/10.23940/ijpe.17.2.p201.mag.

Jin, C. (2021). Dropout prediction model in MOOC based on clickstream data and student sample weight. *Soft Computing*, *25*, 8971–8988. https://doi.org/10.1007/s00500-021-05795-1.

Kang, K., and Wang, S. (2018, March 23-25). *Analyze and predict student dropout from online programs.* [Conference paper]. 2nd International Conference on Compute and Data Analysis, DeKalb, Illinois, USA. https://doi.org/10.1145/3193077.3193090.

Kohavi, R., and John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, *97*(1-2), 273–324. https://doi.org/10.1016/S0004-3702(97)00043-X.

Lázaro Alvarez, N., Callejas, Z., and Griol, D. (2020). Predicting computer engineering students' dropout in cuban higher education with pre-enrollment and early performance data. *JOTSE: Journal of Technology and Science Education*, *10*(2), 241–258. https://doi.org/10.3926/jotse.922.

Li, I. W., and Carroll, D. R. (2020). Factors influencing dropout and academic performance: an australian higher education equity perspective. *Journal of Higher Education Policy and Management*, *42*(1), 14–30. https://doi.org/10.1080/1360080X.2019.1649993.

Louppe, G., Wehenkel, L., Sutera, A., and Geurts, P. (2013, December 5-10). *Understanding variable importances in forests of randomized trees* (Vol. 1). [Conference paper]. 26th International Conference on Neural Information Processing Systems, Lake Tahoe, Nevada, USA.

Moreno-Marcos, P. M., Alario-Hoyos, C., Muñoz-Merino, P. J., and Kloos, C. D. (2018). Prediction in moocs: A review and future research directions. *IEEE transactions on Learning Technologies*, *12*(3), 384–401. https://doi.org/10.1109/TLT.2018.2856808.

Moreno-Marcos, P. M., De Laet, T., Muñoz-Merino, P. J., Van Soom, C., Broos, T., Verbert, K., and Delgado Kloos, C. (2019). Generalizing predictive models of admission test success based on online interactions. *Sustainability*, *11*(18), 4940. https://doi.org/10.3390/su11184940.

Mubarak, A. A., Cao, H., and Hezam, I. M. (2021). Deep analytic model for student dropout prediction in massive open online courses. *Computers & Electrical Engineering*, *93*, 107271. https://doi.org/10.1016/j.compeleceng.2021.107271.

Muñoz-Merino, P. J., Kloos, C. D., Tsai, Y.-S., Gasevic, D., Verbert, K., Pérez-Sanagustín, M., ... Scheihing, E. (2020, September 14-15). *An overview of the LALA project.* [Conference paper]. Workshop on Adoption, Adaptation and Pilots of Learning Analytics in Under-represented Regions co-located with the 15th European Conference on Technology Enhanced Learning 2020, Online. https://ceur-ws.org/Vol-2704/invited1.pdf.

Nagy, M., and Molontay, R. (2023). Interpretable dropout prediction: Towards XAI-based personalized intervention. *International Journal of Artificial Intelligence in Education*, 1–27. https://doi.org/10.1007/s40593-023-00331-8.

Palacios, C. A., Reyes-Suárez, J. A., Bearzotti, L. A., Leiva, V., and Marchant, C. (2021). Knowledge discovery for higher education student retention based on data mining: Machine learning algorithms and case study in chile. *Entropy*, *23*(4), 485. https://doi.org/10.3390/e23040485.

Panagiotakopoulos, T., Kotsiantis, S., Kostopoulos, G., Iatrellis, O., and Kameas, A. (2021). Early dropout prediction in MOOCs through supervised learning and hyperparameter optimization. *Electronics*, *10*(14), 1701. https://doi.org/10.3390/electronics10141701.

Paura, L., and Arhipova, I. (2014). Cause analysis of students' dropout rate in higher education study program. *Procedia-Social and Behavioral Sciences*, *109*, 1282–1286. https://doi.org/10.1016/j.sbspro.2013.12.625.

Pelánek, R. (2015). Metrics for Evaluation of Student Models. *Journal of Educational Data Mining*, *7*(2), 1–19.

Quadri, M., and Kalyankar, D. N. (2010). Drop out feature of student data for academic performance using decision tree techniques. *Global Journal of Computer Science and Technology*, *10*(2), 2–5.

Radovanović, S., Delibašić, B., and Suknović, M. (2021). Predicting dropout in online learning environments. *Computer Science and Information Systems*, *18*(3), 957–978. https://doi.org/10.2298/CSIS200920053R.

Rzepka, N., Simbeck, K., Muller, H.-G., and Pinkwart, N. (2022, April 22-24). *Keep it up: In-session dropout prediction to support blended classroom scenarios.* [Conference paper]. 14th International Conference on Computer Supported Education, Online. https://doi.org/10.5220/0010969000003182.

Smith, J. P., and Naylor, R. A. (2001). Dropping out of university:

A statistical analysis of the probability of withdrawal for UK university students. *Journal of the Royal Statistical Society: Series A* (*Statistics in Society*), *164*(2), 389–405. https://doi.org/10.1111/1467-985X.00209.

Tekin, A. (2014). Early Prediction of students' grade point averages at graduation: A data mining approach. *Eurasian Journal of Educational Research*, *54*, 207–226.

Vidal, J., Gilar-Corbi, R., Pozo-Rico, T., Castejón, J.-L., and Sánchez-Almeida, T. (2022). Predictors of university attrition: Looking for an equitable and sustainable higher education. *Sustainability*, *14*(17), 10994. https://doi.org/10.3390/su141710994.

Wagner, K., Merceron, A., and Sauer, P. (2020, March 23-27). *Accuracy of a cross-program model for dropout prediction in higher education*. [Conference paper]. 10th International Learning Analytics & Knowledge Conference, Frankfurt, Germany.

Yu, R., Lee, H., and Kizilcec, R. F. (2021, June 22-25). *Should college dropout prediction models include protected attributes?* [Conference paper]. 8th ACM Conference on Learning @ Scale, Virtual Event, Germany. https://doi.org/10.1145/3430895.3460139.