SCIENTIA ET TECHNICA

ISSN 0122-1701 - ISSN-E 2344-7214

La revista Scientia et Technica tiene inscripción vigente en la Sección de Registro de Propiedad Intelectual y Publicaciones del Ministerio de Gobierno de la República de Colombia, mediante Resolución 131 de marzo de 1995, Publicación aceptada en el Índice Nacional de Publicaciones Seriadas de Minciencia, Publindex en categoría B hasta el 31 de diciembre de 2021, Publicación asociada a la Red de Revistas Científicas de América Latina y El Caribe, España y Portugal (Redalyc), Vinculados a DIALNET, Centro de Información Tecnológica CIT., EBSCO-Academic Search Complete.

VOL 30 N° 1

ENERO - MARZO DE 2025

Universidad Tecnológica de Pereira

Revista Scientia et Technica

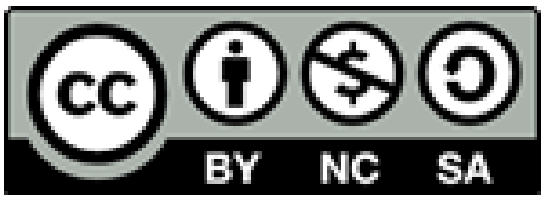
Vol. 30, Núm.01 (enero-marzo) de 2025

https://revistas.utp.edu.co/index.php/revistaciencia/index

Contacto: scientia@utp.edu.co

DOI: https://doi.org/10.22517/issn.2344-7214

ISSN Electrónico (2344-7214)



Pares evaluadores en este número

Jorge Ariel Guapacha Martinez
Raul Andres Becerra Arciniegas
Herbey Andres Milquez Sanabria
Aura Paola Andrea Ayala
Rafael Humberto Merlano Porto
Ruben Dario Hernandez Beleño
Favian Enrique Arenas Aparicio
Miguel Antonio Caro Candelazo
Christian Camilo Gomez Mosquera
Juan Pablo Gonzalez Galvis
Oscar Dario Guarin Villamizar
Javier Ernesto Holguin Gonzalez
Jenny Lucia Huertas Delgado

Contenido

Scientia et Technica Año XXIX, Vol. 30, No. 01, enero-marzo de 2025. Universidad Tecnológica de Pereira. ISSN 0122-1701 y ISSN-e: 2344-7214
DOI: 10.22517/23447214.25820

Editorial

La Importancia de la Indexación Internacional para las Revistas Científicas Colombianas

En el contexto académico colombiano, la visibilidad, el impacto y la legitimidad de las revistas científicas son elementos esenciales para fortalecer la producción nacional en ciencias básicas e ingeniería y, de esta manera, potenciar la visibilidad de la ciencia colombiana. En este escenario, la indexación en sistemas internacionales como Web of Science (WoS), Scopus y el Directory of Open Access Journals (DOAJ) representa una de las estrategias fundamentales de posicionamiento [1]–[3].

Por su lado, Web of Science, gestionado por Clarivate Analytics, aplica criterios editoriales rigurosos y selecciona solo publicaciones con alto impacto. La inclusión en esta facilita que a los artículos se brinde mayor prestigio, afectando de manera positiva a los autores y a las revistas [1]. Es de destacar que el Journal Impact Factor (JIF), derivado de WoS, constituye una referencia global para medir la influencia académica [4].

Alternativamente, Scopus ofrece una cobertura más amplia que WoS y a la vez proporciona métricas como CiteScore y SCImago Journal Rank (SJR), las cuales son fundamentales para visibilizar revistas emergentes [2], [5]. Para muchas publicaciones colombianas que aún no alcanzan el umbral de selectividad de WoS, Scopus representa un primer paso en el camino de la citación internacional, a la vez que aporta credibilidad [5].

Finalmente, DOAJ cumple una función complementaria al validar la calidad editorial de las revistas de acceso abierto, puesto que esta es su condición inicial. Para su inclusión, se debe cumplir con políticas de ciencia abierta, garantizar transparencia editorial y un amplio alcance de visibilidad [3].

Las revistas colombianas funcionan principalmente bajo el modelo de acceso abierto, por lo que DOAJ resulta ser un paso más en el cumplimiento de los requisitos de visibilidad latinoamericana, sin dejar de tener un alcance mundial [6].

Si bien cada índice presenta fortalezas y limitaciones, su importancia varía según el objetivo editorial: WoS ofrece reputación y selectividad; Scopus, cobertura y métricas alternativas; y DOAJ, globalización de la ciencia y legitimidad ética [1]–[3], [6]. Por lo anterior, lograr que las revistas científicas estén indexadas en estos sistemas de manera conjunta puede maximizar significativamente su impacto [5].

No es un hecho menor que a la fecha la revista Scientia et Technica (SeT) está indexada en diversos sistemas y bases de datos reconocidos que amplifican su visibilidad y prestigio en la comunidad académica [15], dentro de los cuales destacan: la red de revistas científicas de acceso abierto (Redalyc) que promueve la difusión del conocimiento en Iberoamérica [7]; el sistema regional que reúne información sobre revistas académicas de América Latina, el Caribe, España y Portugal (Latindex) [8]; la Matriz de Información para el Análisis de Revistas, que proporciona datos clave para la identificación y evaluación de publicaciones científicas (MIAR) [9]; el motor de búsqueda académico que facilita el acceso a literatura revisada por pares (Google Scholar) [10]; un portal bibliográfico que recopila contenidos científicos hispanos (Dialnet) [11]; y la base de datos líder que ofrece acceso a contenido académico multidisciplinario (EBSCO) [12].

La inclusión en estos índices refleja el compromiso de Scientia et Technica con la calidad editorial y la difusión del conocimiento científico, asegurando una mayor accesibilidad y reconocimiento de los trabajos publicados. En este contexto, la revista sigue buscando estar también en un más amplio conjunto de indicadores, a pesar de que algunos sean cerrados y otros abiertos, lo cual se puede lograr a través de un arduo trabajo del equipo editorial [15]. Actualmente, se está trabajando para conseguir de manera oportuna la vinculación adicional en WoS, Scopus y DOAJ, por lo que los requerimientos de calidad se reflejan en la rigurosidad de las evaluaciones, la producción editorial y la selectividad de los artículos, principalmente sobre aquellos que son resultado de investigación.

La Universidad Tecnológica de Pereira, a través de la Vicerrectoría de Investigaciones, Innovación y Extensión, ha prestado especial atención a estos aspectos, por lo que, adicionalmente, se está buscando que Scientia et Technica pertenezca al Comité de Ética de COPE (Committee on Publication Ethics) [13]. Todos los colaboradores de la revista han recibido la capacitación necesaria para alcanzar los estándares exigidos en el logro de estos objetivos, por lo que se espera prontamente tener las respectivas vinculaciones en estos sistemas y contar con el aval del comité de ética de COPE de manera oficial.

Se sabe que no es un proceso simple, razón por la cual se ha realizado de manera cautelosa, lo que ha exigido ajustes permanentes, desde aspectos de forma en el portal de la revista en su OJS (Open Journal Systems), hasta asuntos más profundos como las políticas editoriales y la rigurosidad en la publicación y en la ética de evaluación y publicación [14].

El comité editorial y científico de la revista Scientia et Technica, en cabeza del editor, valora el interés de los autores por someter sus trabajos a la revista y el de los evaluadores que lo hacen de

manera desinteresada y recíproca por mejorar los aspectos de calidad que poco a poco se evidencian en el crecimiento de la ciencia con un alcance internacional [15].

Referencias:

[1] Clarivate Analytics, "Web of Science Journal Evaluation Process and Selection Criteria," Clarivate, 2022. [Online]. Available: https://clarivate.com/webofsciencegroup/journal-evaluation-process-and-selection-criteria/
[2] Elsevier, "Content Policy and Selection," Scopus Content Coverage Guide, 2023. [Online]. Available: https://www.elsevier.com/solutions/scopus/how-scopus-works/content
[3] DOAJ, "DOAJ Best Practice Guide," Directory of Open Access Journals, 2023. [Online]. Available: https://doaj.org/publishers
[4] Clarivate Analytics, "Understanding the Journal Impact Factor," Web of Science Group, 2022. [Online]. Available: https://clarivate.com/blog/understanding-journal-impact-factor/
[5] SCImago, "SCImago Journal Rank (SJR) Indicator," SCImago Journal & Country Rank, 2023. [Online]. Available: https://www.scimagojr.com/aboutus.php
[6] J. Björk and M. Solomon, "Open Access — Are the Benefits Worth the Costs?," Learned Publishing, vol. 28, no. 1, pp. 23–25, Jan. 2015. doi: 10.1087/20150106
[7] "Acerca de Redalyc," Redalyc, [En línea]. Disponible: https://www.redalyc.org/. [Último acceso: 31 de marzo de 2025].
[8] "Scientia et Technica," Latindex, [En línea]. Disponible: https://www.latindex.org/latindex/ficha/17004. [Último acceso: 31 de marzo de 2025].
[9] "MIAR: Matriz de Información para el Análisis de Revistas," MIAR, [En línea]. Disponible: https://miar.ub.edu/. [Último acceso: 31 de marzo de 2025].
[10] "Acerca de Google Scholar," Google Scholar, [En línea]. Disponible: https://scholar.google.com/intl/es/scholar/about.html. [Último acceso: 31 de marzo de 2025].
[11] "¿Qué es Dialnet?," Dialnet, [En línea]. Disponible: https://dialnet.unirioja.es/. [Último acceso: 31 de marzo de 2025].
[12] "EBSCO Information Services," EBSCO, [En línea]. Disponible: https://www.ebsco.com/es. [Último acceso: 31 de marzo de 2025].
[13] Committee on Publication Ethics (COPE), "COPE Principles of Transparency and Best Practice in Scholarly Publishing," 2022. [Online]. Available: https://publicationethics.org/resources/guidelines
[14] Public Knowledge Project, "OJS 3 Guide for Editors," PKP Documentation, 2023. [Online]. Available: https://docs.pkp.sfu.ca/learning-ojs/en/
[15] "Scientia et Technica," Revista Scientia et Technica, Universidad Tecnológica de Pereira, [En línea]. Disponible: https://revistas.utp.edu.co/index.php/revistaciencia. [Último acceso: 31 de marzo de 2025].

Autores

PhD. Jimmy Alexander Cortes Osorio
Docente Titular Departamento de Física Editor
jefe -Revista Scientia et Technica
Grupo Investigación Robótica Aplicada
Línea: Computer Vision and Machine Learning
Investigador Senior Reconocido por MINCIENCIAS
Universidad Tecnológica de Pereira
ORCID: https://orcid.org/0000-0002-0413-807X

Evaluation of polymeric coatings applied to a natural fique fiber mesh of a water harvesting fog catcher system

Evaluación de recubrimientos poliméricos para malla de fibra natural de fique en sistema atrapa nieblas para la captación de agua.

S. Gómez-Suarez; E. Córdoba-Tuta
DOI: https://doi.org/10.22517/23447214.25756
Scientific and technological research paper

***Abstract*—One of society's priority needs is acquiring fresh water due to high contamination levels and limited access in areas where this valuable resource is scarce. Non-traditional methods of water acquisition, such as fog catcher systems, are increasingly relevant because of their low cost and versatility. These systems use collection meshes to condense fog microdroplets. The water then undergoes filtration, adsorption, and disinfection processes to ensure its potability. Unfortunately, the materials commonly used in fog catcher meshes are synthetic, making them resistant to degradation. Consequently, natural fibers present a viable alternative for their replacement. However, the hydrophobicity of natural fibers is low, which results in limited water capture. This necessitates the development of new solutions, such as coatings, to enhance water capture efficiency. This article presents an evaluation of various polymeric coatings applied to natural fique fiber meshes installed in fog catchers, focusing on the impact of these coatings on water capture efficiency. Additionally, a mechanical and morphological characterization of the coated meshes was performed to assess their mechanical properties and adhesion. Mechanical characterization was conducted using tensile testing, which revealed improved properties in the epoxy-coated fique mesh system. Morphological analysis, using scanning electron microscopy, showed better adhesion between the epoxy and polyester resins and the natural fiber. Water capture tests conducted both in the field and in the laboratory demonstrated that the fique-epoxy coating is the most effective, increasing water uptake by 124.4% compared to uncoated fique fiber.**

***Index Terms*— Coatings; Composite; Fique; Fog catchers; Natural fiber; Water capture.**

***Resumen*—Una de las necesidades prioritarias de la sociedad es la adquisición de agua dulce debido a los altos niveles de contaminación y al acceso limitado en zonas donde este valioso recurso escasea. Los métodos no tradicionales de obtención de agua, como los sistemas atrapanieblas, son cada vez más relevantes por su bajo coste y versatilidad. Estos sistemas usan mallas colectoras para condensar microgotas de niebla. Luego, el agua pasa por filtración, adsorción y desinfección para asegurar su potabilidad. Lamentablemente, los materiales utilizados habitualmente en las mallas de los atrapanieblas son sintéticos, lo que los hace resistentes a la degradación. En consecuencia, las fibras naturales presentan una alternativa viable para su sustitución. Sin embargo, la hidrofobicidad de las fibras naturales es baja, lo que se traduce en una captura de agua limitada. Esto hace necesario el desarrollo de nuevas soluciones, como los recubrimientos, para mejorar la eficacia de la captura de agua. Este artículo presenta una evaluación de varios recubrimientos poliméricos aplicados a mallas de fibra de fique natural instaladas en atrapanieblas, centrándose en el impacto de estos recubrimientos en la eficiencia de captura de agua. Además, se realizó una caracterización mecánica y morfológica de las mallas recubiertas para evaluar sus propiedades mecánicas y su adherencia. La caracterización mecánica se llevó a cabo mediante ensayos de tracción, que revelaron una mejora de las propiedades en el sistema de malla de fique recubierto de epoxi. El análisis morfológico, mediante microscopía electrónica de barrido, mostró una mejor adherencia entre las resinas epoxi y de poliéster y la fibra natural. Las pruebas de captación de agua realizadas tanto en el campo como en el laboratorio demostraron que el recubrimiento de fique-epoxi es el más eficaz, ya que aumenta la captación de agua en un 124,4% en comparación con la fibra de fique sin recubrir.**

***Palabras claves*— Atrapanieblas; Captación de agua; Compuestos; Fibra natural; Fique; Recubrimientos.**

I. INTRODUCTION

One of the current challenges facing humanity is the availability of clean freshwater, due to the contamination of various water sources as a result of economic expansion, industrial development, and climate change. This contamination affects the accessibility of water for essential uses such as drinking, hygiene, and food security, directly impacting public health [1].

Of various water sources as a result of economic expansion, industrial development, and climate change.

This contamination affects the accessibility of water for

This manuscript was submitted on January 22, 2025. Accepted on March 15, 2025. And published on March 31, 2025. This research work was funded by the Universidad Pontificia Bolivariana

S. Gómez S. Author is with the Mechanical Engineering Department, Universidad Pontificia Bolivariana, Km 7 autopista via Piedecuesta, Floridablanca, Santander, Colombia, (e-mail: sergio.gomezs@upb.edu.co)

E. Cordoba T. Author is with the Mechanical Engineering Department, Universidad Pontificia Bolivariana Km 7 autopista via Piedecuesta, Floridablanca, Santander, Colombia, (e-mail: edwin.cordoba@upb.edu.co)

essential uses such as drinking, hygiene, and food security, directly impacting public health [1].

In the world, fresh water represents 3% of the existing total, making it difficult to obtain since 70% of it is present in the form of glaciers and snow in high mountains [2]. In this context, new ways of obtaining fresh water that are practical and at low cost have been explored.

Among the emerging non-traditional methods of water acquisition, the most prominent are rainwater harvesting, groundwater collection, desalination, and atmospheric water recovery, with the latter being the most significant due to its low cost and versatility [3].

The system commonly used to capture water from the air is known as fog catcher, which provides an accessible water supply at high altitudes [4]. Its operation consists of the condensation of atmospheric water vapor that is present in the air, concentrating it into drops of liquid water, known as dew [5].

Fog catchers are generally made of polypropylene or polyethylene meshes known as Raschel meshes. This material is used due to its high hydrophobicity, this being one of the most important parameters in the efficiency of this type of system [6], together with its high commercial availability, low price and good mechanical behavior [7].

The latest research in this type of devices is focused on the development of bio-inspired designs for increased water capture and the development of new materials [8]. Unfortunately, polypropylene and polyethylene are materials of synthetic origin and therefore remain inert to degradation, which leads to their accumulation creating serious environmental problems [9].

A material that can replace the synthetic fibers that make up the meshes of conventional collectors are natural fibers, since they have high mechanical resistance, high rigidity, low density and high availability at a low cost. In addition to the above, production requires little energy, is carried out with low emission of toxic fumes and with less abrasive impact on the processing equipment [10] [11].

However, the problem with natural fibers to be applied in fog catcher systems is that they have a hydrophilic nature, absorbing the water present in the air without allowing it to be used, and are further degraded by microorganisms and sunlight [12].

In order to increase the efficiency of fog catchers, research has focused on the use of surface coatings that increase the hydrophobicity of the systems, mainly applied to fibers made of synthetic materials [13] [14]. Additionally, research is limited to comparisons of different synthetic textile meshes and fog catchers in the laboratory or in the field [15]. However, there is not enough information on studies where an application in fog-catching systems with natural fibers and coatings is performed.

Fique fiber is extracted from the leaves of the plant of the same name. It is a resistant and versatile material that is grown mainly in Colombia, Ecuador and Mexico. It belongs to the agavaceae family, this fiber has a length of 1.5 to 2 meters and is known for its hardness. Traditionally, it is used to make ropes and sacks. In addition, fique is ideal for packing coffee for export due to its ability to preserve freshness. In a more innovative approach, yarns and textile bases have been developed from fique fiber, offering a sustainable alternative to traditional fibers and contributing to the economic development of rural communities in Latin America [16].

It is for this reason that in this article the evaluation of different polymeric coatings applied on a natural fiber mesh of fique installed in a fog catcher was carried out, knowing the influence of the use of these coatings on the level of water capture in the laboratory and in the field. Additionally, a mechanical characterization was carried out to know the resistance to external forces that can be caused by the environment and a morphological characterization was also carried out to know the adhesion of the coating with the natural fiber. Although natural fibres such as fique represent a sustainable alternative to synthetic fabrics due to their renewable and biodegradable nature, they require the application of a polymer coating to improve their hydrophobic properties. Although this coating, which is of synthetic origin, reduces their degradability, the controlled integration of this material on a natural and biodegradable fibre allows a significant reduction in the amount of polymer used compared to conventional plastic fabrics, thus reducing their environmental impact.

II. MATERIALS AND METHODS

A. Materials

As a natural material for the meshes, fique fiber was used in a weaving configuration, which presented a 0/90 braid configuration, with a double fiber in the warp and a double fiber in the weft. This fiber was extracted from packaging used to transport coffee. Fig. 1 shows the fiber used.

Fig. 1. Fique fiber used

Four different types of commercial polymeric coatings were used in order to identify which of these allowed the greatest amount of water capture when the fique fiber mesh was coated

and used in the fog catchers. The general characteristics of each of the coatings are mentioned below.

- Epoxy resin: The percentage of resin-catalyst applied by volume was 1:1 respectively. The resin was acquired from the company Ingequimicas of Bucaramanga, Santander, Colombia.
- Asphalt emulsion (EUCD): The resin was applied diluted in a ratio of 1 to 3 by volume of water. Curing was carried out for 48 hours at room temperature. The resin applied was Sika brand asphalt emulsion.
- Impercryl: This resin is formulated with styrene acrylic resins, plasticizers and ceramic particles. The application was direct with one coat. The resin brand used was P-7 of Poliescol in white color.
- Polyester: The polyester resin applied was pre-accelerated, catalyzed at 3% by weight with methyl ethyl ketone peroxide (MEK peroxide). The resin was purchased from the company Ingequimicas of Bucaramanga, Santander, Colombia, reference 856.

B. Characterization of meshes with coatings

Four fique-coating meshes were evaluated, presenting the configuration shown in Table I, and an additional configuration of uncoated fique mesh was also analyzed, as well as an additional poly-shade mesh (Raschel), which is the one usually used in this type of fog-catching devices.

TABLE I.
CONFIGURATION OF FIQUE-COATING SYSTEMS

Material	Description
Material 1	Fique mesh coated by epoxy resin
Material 2	Fique mesh coated by asphalt emulsion
Material 3	Fique mesh coated by impercryl
Material 4	Fique mesh coated with Polyester
Material 5	Uncoated fique mesh
Material 6	Polyshade mesh (Raschel)

The different coatings were applied on the fique fabrics by means of conventional brushes, superficially covering the natural fibers with the synthetic resins. All the resins were cured for 72 hours at room temperature, and the product was applied in a single layer on the fique fiber mesh.

To quantify the amount of polymer deposited on the natural fibers, a mass analysis of the applied coatings was performed. For this purpose, three representative fragments of each mesh, measuring 8 cm × 8 cm, were weighed before and after the polymer application using a Highland HCB602H precision digital balance. This procedure enabled the determination of the average mass of polymer added per sample.

Mechanical characterization was carried out by means of the tension test following the ASTM D4595-17 standard, this standard was developed for geotextiles, however, as expressed by Dios Rivera *et al*. [17], it is the most appropriate standard to characterize this type of materials since there is no specific standard for the evaluation of textiles in fog-catching systems.

Although the primary function of the fique fiber meshes in fog water harvesting systems does not require high mechanical strength, the tensile properties of both coated and uncoated meshes were evaluated in this study to assess their structural integrity under environmental conditions. Natural fibers are exposed to wind loads, handling, and long-term weathering, which can cause mechanical degradation over time. Therefore, understanding the effect of polymeric coatings on the tensile behavior provides relevant information about the durability and stability of the meshes during prolonged outdoor exposure, even if mechanical performance is not a critical design requirement.

The tests were carried out on a 10 KN MTS model C43.104 universal machine. The results obtained present the tensile stress (maximum force per unit width) on deformation applied to 5 specimens of each system with a geometry of 40 mm wide by 200 mm long, as established in the standards. The results presented are the average of those obtained on the five test specimens. The strain rate was defined as 11% /min with an ambient temperature of 21.2 +/- 1.2°C and a humidity of 63.2 +/- 1.2 %. It is important to note that the specimens were tested in their dry state, meaning they had only absorbed the moisture naturally present in the laboratory environment. Fig. 2 shows the mechanical characterization of material 2.

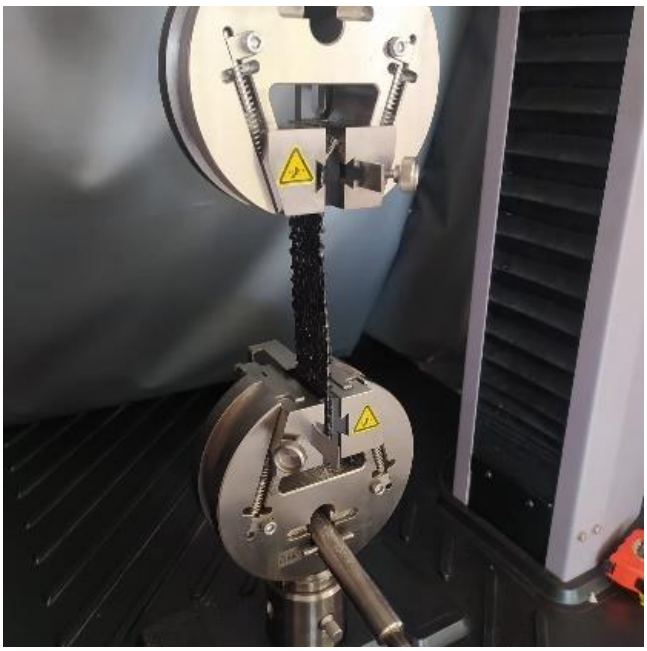


Fig. 2. Mechanical characterization

Additionally, a morphological characterization of the mesh materials was performed using scanning electron microscopy. Samples smaller than 3 cm² (the size limit required by the equipment) were coated with gold to enhance conductivity, utilizing a Cressington Model 108 Auto/SE sputtering system. Fig. 3 presents the samples and their coating.

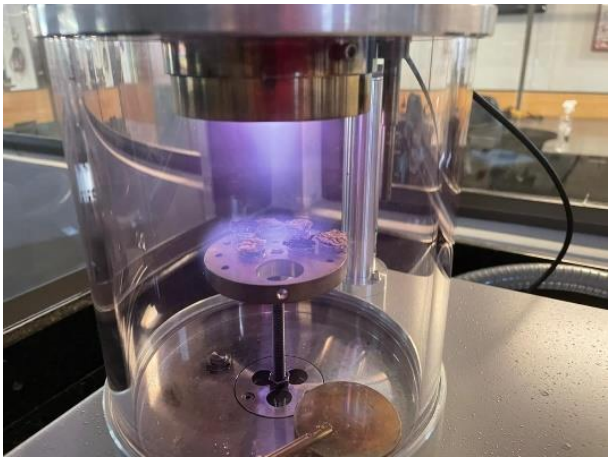


Fig. 3. Meshes coated with gold

The study was carried out with a Tescan scanning electron microscope model MIRA 3 FEG-SEM with a secondary electron detector model A65c SED. Images were taken at 200X, 500X and 1000X magnifications.

The water capture test of the coated mesh materials was performed in the laboratory, using the setup designed by Rajam et al. [13] and Wang [14] as a reference. In this procedure, fog was generated using an ultrasonic humidifier, submerged in 3 liters of water stored in a polymeric cubicle. Additionally, a fan operating at 3600 revolutions per minute was used to increase the outflow velocity of the fog. A plastic tube was installed in the cubicle to direct the flow over the different test samples of each system. Fig. 4 shows the schematic used in the laboratory test.

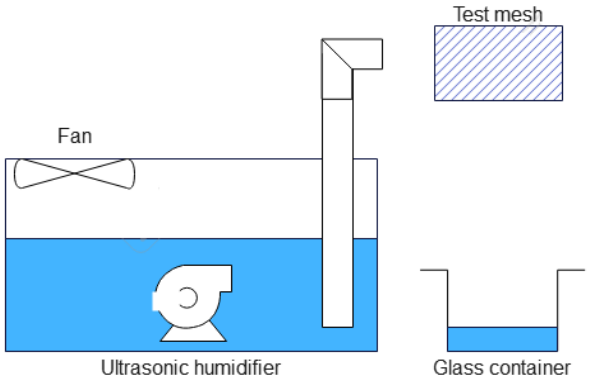


Fig 4. Schematic of the laboratory test

A glass container was installed under the samples of each of the systems to collect the drained water. The different materials were installed 5 cm from the pipe outlet. Each sample was exposed to the fog generated by the device for 60 minutes, the amount of water collected was weighed with a digital balance to obtain the results. The samples evaluated had a geometry of 12 cm by 12 cm.

C. Construction and evaluation of fog catchers

Six flat type fog catchers were constructed using different mesh materials. The meshes used for each of the fog catchers were 1 m wide by 1.5 m high. The lateral supports were made of 2.5 m high bamboo wood pillars embedded in the ground. At the bottom of the mesh, 1-meter-long PVC gutters with a diameter of 110 mm were installed, through which the captured fluid precipitated towards the storage tank by gravity due to the inclination of the gutter of approximately 30°. Each fog catchers configuration was installed one after the other in order to keep the water collection conditions as repeatable as possible.

The application of the coatings for each of the meshes of the systems that contained fique fiber was carried out manually with a brush and on both sides (front and back), replicating what was done with the materials. Fig. 5 shows the application process.

Fig. 5. Coating application process

The mesh anchoring system to the supports was made using stainless steel cables and tensioners with a diameter of 5.16 mm. Fig. 6 shows the fog catchers that were manufactured and installed.

Fig. 6. Fog catchers manufactured and installed

The water collection with the fog catcher systems was carried out in the municipality of Floridablanca, Santander, Colombia (7.03835, -73.07218). The region has a tropical climate. The climatological parameters for the two-month period during which the tests were conducted are shown in Table II.

TABLE II.
CLIMATOLOGICAL PARAMETERS

Parameters	Month 1	Month 2
Maximum temperature (°C)	28.88	27.77
Average temperature (°C)	23.57	23.21
Minimum temperature (°C)	18	20
Precipitation (mm)	180	195
Average Humidity (%)	87%	86%
Wind Speed (mph)	4.82	4.24

It should be noted that the climatological data were obtained from a weather monitoring station located very close to where the fog-catching systems were installed. Fig. 7 shows the monitoring station.

Fig. 7. Monitoring station

The water captured by the fog catchers was stored in plastic containers, recording the volume of water captured, in milliliters, every 3 days. A total of 10 measurements were obtained for each fog catchers, however 2 measurements were discarded as they were affected by rainfall precipitation in the area. The results show the average of the 8 endorsed measurements.

D. Statistical analysis

In order to define if the differences obtained between the results of the mechanical characterization of the different systems and water capture in the six types of fog catchers were statistically significant, an analysis of variance (ANOVA) was used, in which if the P value is less than the significance level (defined as 0.05), it is concluded that at least one mean of the mechanical properties and water capture of the systems is different. Additionally, Scheffe's post hoc tests were performed to perform multiple comparisons of the means and to recognize which of them was different.

III. RESULTS

A. Characterization of meshes

Table III shows an increase in the mass of the fique meshes after the application of polymeric coatings, which varies according to the type of resin used. This increase reflects the amount of polymeric material adhered to the natural fibers.

TABLE III.
QUANTIFICATION OF POLYMER MASS DEPOSITED ON FIQUE FIBER MESHES.

Material	Initial Mass (g)	Final Mass (g)	Coating Mass (g)	Coating Mass (%)
Material 1	3.22 ± 0.8	4.15 ± 1.1	0.93	28.9
Material 2	3.14 ±1.1	3.68 ± 0.9	0.54	17.2
Material 3	2.91 ± 0.7	3.62 ± 0.8	0.71	24.4
Material 4	3.23 ±0.5	4.24 ± 0.4	1.01	31.3

The results show an increase in the mass of all fique meshes after the application of the polymeric coatings, suggesting the formation of a layer on the natural fibers. Among the materials evaluated, Material 4 exhibited the highest mass gain. This behavior is attributed to the high solid content of the polyester resin, which promotes greater deposition and retention of the coating on the fibers.

In contrast, Material 2 recorded the lowest mass increase, due to the high water content of the asphalt emulsion used. During the curing process, this water evaporates, leaving a lower amount of solids adhered to the fiber.

Meanwhile, Material 1 (coated with epoxy resin) and Material 3 (coated with Impercryl) showed intermediate increases. The epoxy resin, like the polyester resin, has a considerable solid content, although slightly lower, which explains its behavior. In the case of Impercryl, an acrylic-based coating, its lower mass gain is also related to its high water content, similar to the asphalt emulsion, which upon evaporation reduces the amount of solid material retained on the fibers.

However, it is important to note that the observed mass increase is mainly associated with the amount of material deposited and does not necessarily reflect the uniformity or quality of the coating adhesion.

Table IV shows the results obtained from the tensile mechanical tests performed on the meshed materials coated with the different polymers.

TABLE IV.
TENSILE TEST RESULTS

Material	Tensile Strength (N/m)	Tensile Modulus (N/m)
Material 1	35550 ± 6098	603697 ± 34325
Material 2	9275 ± 674	143093 ± 47275
Material 3	13792 ± 1102	190142 ± 88387
Material 4	31308 ± 6972	498374 ± 49683
Material 5	8375 ± 3340	120412 ± 32605
Material 6	2658 ± 253	10416 ± 1116

All systems showed an improvement in tensile stress when the coatings were used, compared to the uncoated fique mesh system. The best performance was obtained by the epoxy resin-coated fique mesh system, which showed an increase of 324.4% compared to the uncoated fique fabric. This was followed by the polyester resin coated system with an increase of 273.8%.

The fique fiber mesh systems exhibited a higher tensile strength than the poly-shade mesh, this is due to the fact that although natural fibers usually have lower mechanical properties than polymeric fibers [18], there were more natural fibers that were supporting the load due to the configuration and manufacture of the fique fabric used. According to Singh *et al*. [19] the higher the fiber volume fraction, the better the mechanical behavior. Fig. 8 shows the configuration and quantity of the fibers of the fique fabric and the poly-shade mesh. The results obtained for the polyshade mesh are in the order of those reported by Rivera *et al*. [20], the results of fique could not be compared since there are no studies where they were characterized with the standard used in the study.

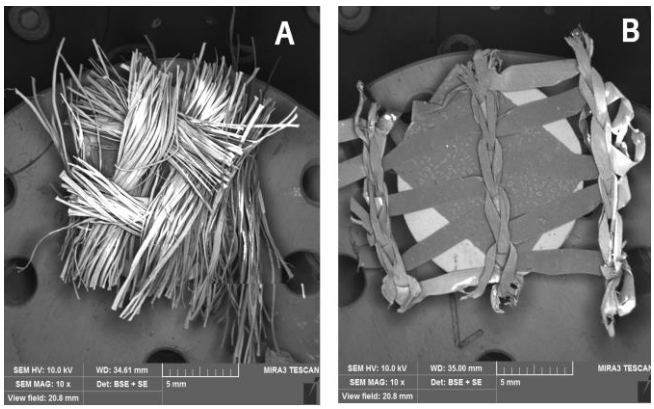


Fig. 8. Woven configuration. A) Fique fiber, B) polyshade

The coatings used made all the meshed systems stiffer compared to the uncoated fique fabric. The highest tensile modulus was presented by the fique mesh coated with epoxy resin being 401.36% higher than the uncoated fique fiber fabric and the lowest was presented by the fabric coated with asphalt emulsion, however, being also higher than the uncoated fabric by 18.83%.

Table V shows the ANOVA tests where the significant differences between the means of the different mechanical properties of the meshed systems were evaluated.

TABLE V.
ANOVA TEST FOR MECHANICAL PROPERTIES

Property	Source	Sum of squares	Mean square	FO	P
Tensile Strength	Residual systems	2.70 E+9 1.97 E+8	5.39 E+9 1.64 E+7	32.8	<0.001
Tensile Modules	Residual systems	6.36 E+11 4.51 E+10	1.27 E+11 3.76 E +9	33.8	<0.001

As can be seen, a P value of less than 0.05 is obtained for both tensile stress and modulus of elasticity, indicating that at least one of the coated meshed systems had a statistically different mean from the others.

Post hoc tests show that the system coated with epoxy resin and the one coated with polyester have no statistically significant differences between them in their tensile stress and modulus of elasticity; however, there are differences with the other mesh systems. Additionally, the asphalt emulsion, impercryl, uncoated fique fabric and polyshade systems have no statistically significant differences between them in their tensile modulus and stress.

Scanning electron microscopy analysis of the different fique mesh systems with epoxy resin, asphalt emulsion, Impercryl, and polyester resin coatings are shown in Fig. 9, 10, 11 and 12, respectively.

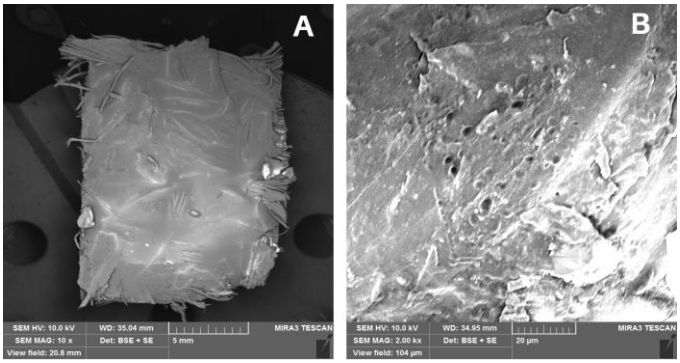


Fig. 9. Electron microscopy of epoxy resin coating. A) 10X, B) 2000 X

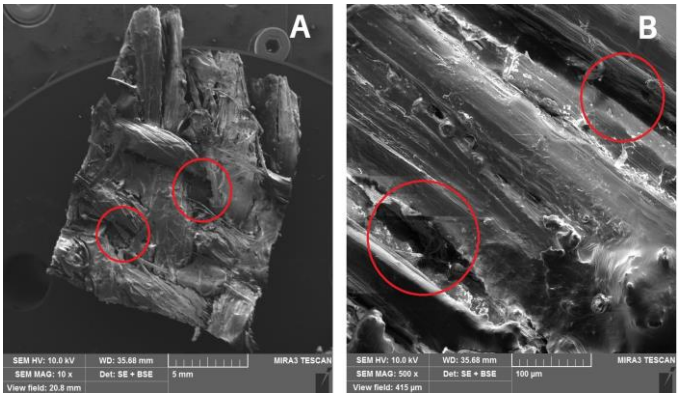


Fig. 10. Electron microscopy of the asphaltic emulsion coating. A)10 X, B)500 X

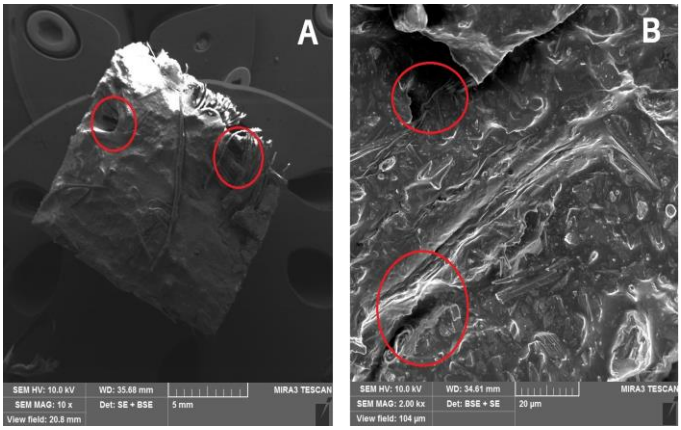


Fig. 11. Electron microscopy of the impercryl coating. A) 10 X, B) 2000 X

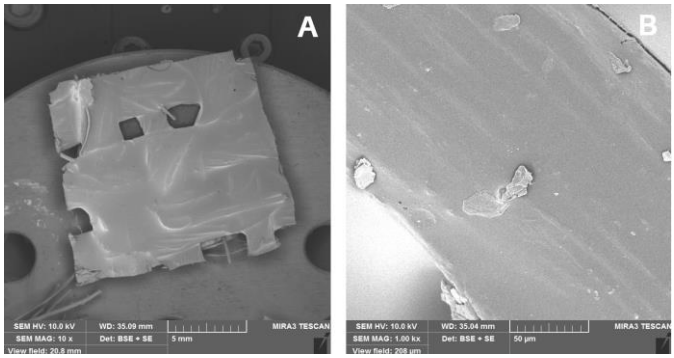


Fig. 12. Electron microscopy of the polyester resin coating A)10 X, B) 1000 X

A low compatibility between the fique fiber with the asphalt emulsion and with the impercryl coating is evidenced, presenting cracks and spaces where the resin did not have good adherence with the fibrous surface. This can be attributed to the nature of their functional groups, which do not form strong interactions with the hydroxyl groups present in the fique fiber. Additionally, the mass analysis confirmed that materials with lower adhesion also exhibited lower mass, suggesting that the amount of deposited polymeric material influences crack formation. This indicates that both chemical compatibility and material quantity affect adhesion performance. However, with respect to the epoxy and polyester resin, the morphological analysis infers that there was a better adhesion with the fique, since it presents a wide, smooth and compact surface. This improved adhesion is likely due to the reactive epoxide groups in the epoxy resin, which form strong covalent bonds with the hydroxyl groups of the natural fiber, and to the crosslinking reaction of polyester resin catalyzed with MEK peroxide, which enhances rigidity and adhesion. It should be noted that the polymer resin-natural fiber adhesion is not so good due to the hydrophilic characteristics of the fibers and the hydrophobic nature of the matrix together with the impregnation process used, which was manual [21].

The SEM micrographs revealed that, after the application of the polymeric coatings, the natural roughness and porous structure of the fique fibers were partially covered by the synthetic layer. This morphological change suggests a reduction in the effective surface area available for water droplet condensation. Although the coatings improved the hydrophobic behavior of the meshes, the smoother surface generated by the polymer layer may limit the anchoring points for droplet formation, potentially affecting the fog water harvesting efficiency.

The water capture tests of the coated mesh materials carried out in the laboratory showed the results presented in Fig. 13.

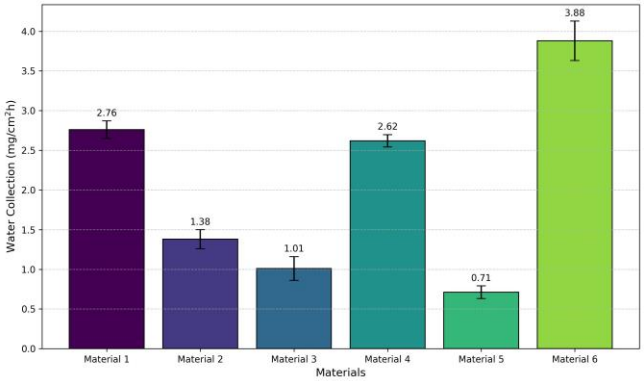


Fig. 13. Water capture testing of coated mesh materials in the laboratory.

The highest water capture was obtained in the Polyshade mesh system (Rachel), being 3.88 ± 0.258 mg/cm2h. This result was not only due to the characteristics of the Polyshade mesh material but, additionally, to the geometry and size of the mesh, since these are factors that directly affect the capture efficiency as mentioned by Vásquez et al. [22], these parameters being different in the fique fiber mesh used with the other coatings.

The use of polymeric coatings applied to the natural mesh increased the amount of water capture, with a higher percentage in the mesh with epoxy resin, where the increase obtained was 289.08% with respect to the uncoated natural fiber, followed by the mesh with polyester resin at 268.66%. The lowest growth was presented with the impercryl at 41.54%.

When the ANOVA test was applied to evaluate significant differences in water capture in the laboratory, the values shown in Table VI were obtained.

TABLE VI.
ANOVA TEST FOR WATER CAPTURE IN THE LABORATORY

Property	Source	Sum of squares	Mean square	FO	P
Water capture in lab	Residuals systems	30.013	6.0026	209	<0.001
		0.517	0.0287		

As can be seen, a P value of less than 0.05 is obtained, so that at least one of the coated mesh systems presents a statistically different mean from the others in the amount of water capture.

Post hoc tests show that the poly-shade mesh had a statistically significant mean water capture difference versus the others. Additionally, the water capture of the mesh coated with epoxy resin and polyester did not show significant differences between the mesh coated with asphalt emulsion, Impercryl and the uncoated fique mesh, and there were no statistically significant differences.

B. Evaluation of fog catchers

The average obtained from the 8 measurements of water capture by the fog catchers installed in the field is shown in Fig. 14.

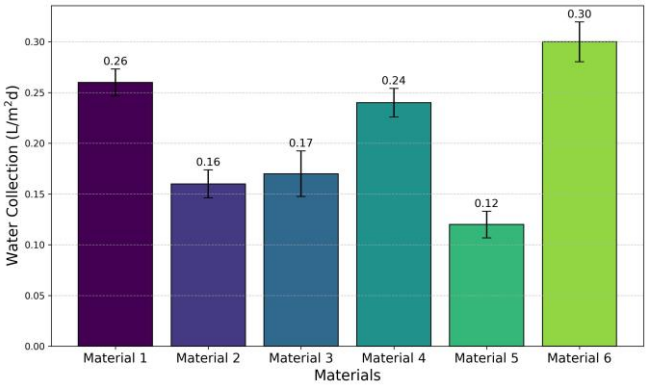


Fig.14. Water capture testing of coated mesh materials in the field.

The highest water capture was obtained in the fog-catcher with the poly-shade mesh, followed by the fog-catcher with fique mesh coated with epoxy resin, which had an increase of 124.4% in water capture with respect to the fog-catcher with the uncoated fique mesh.

The coatings on the fog catchers, with polyester resin, impercryl, asphalt emulsion increased water capture by 105.9%, 41.5%, and 37.2%, respectively, compared to the uncoated fique fiber.

The increase in water capture, as reported by Rajaram et al. [23], is due to the fact that increasing the hydrophobicity of the mesh increases the fog collection efficiency. Despite the increase in water capture with the applied coatings versus natural fiber, the amount collected by the fog catchers is below the common values where these systems are installed which are 0.8 to 10 L/m2d [24].

When the ANOVA test was applied to evaluate significant differences in the field water capture, the values shown in Table VII were obtained.

TABLE VII.
ANOVA TEST WATER CAPTURE IN THE FIELD

Property	Source	Sum of squares	Mean square	FO	P
Water capture in field	Residuals systems	0.250	0.050	150	<0.001
		0.0179	3.32 E-4		

The post hoc tests reflected the same behavior as the laboratory tests, where the polyshade mesh showed a statistically significant difference in water capture versus the others; between the mesh coated with epoxy resin and polyester there are no significant differences and additionally between the meshes coated with asphalt emulsion, Impercryl and the uncoated fique mesh there are no statistically significant differences.

IV. CONCLUSIONS

The mechanical, morphological, and water capture properties of four polymer-coated fique fiber meshes (epoxy, polyester, Impercyrl, and asphalt emulsion) were evaluated and compared to an uncoated fique mesh and an additional poly-shade mesh (Raschel).

An increase in the mechanical properties was evidenced in all the polymer-coated fique meshes, compared to the natural fique mesh system. The best performance was obtained in the epoxy resin coated fique mesh system.

The morphological analysis with scanning electron microscopy allowed observing low adhesion between the fique fiber with the asphalt emulsion and with the impercryl coating. However, with respect to the epoxy and polyester resin, a medium adhesion was observed.

The use of all the polymeric coatings applied to the natural fiber mesh increased the amount of water capture in the laboratory, with the highest increase when using the epoxy resin, compared to the uncoated natural fiber.

Fog catchers were manufactured and installed in the field with fique fiber coated with the polymers, which showed an increase in water capture compared to uncoated natural fique fiber. The results, as in the laboratory, showed that the increase was greater with the epoxy coating.

As the mechanical properties of the materials in this study could be affected by water absorption, further research is needed to evaluate these materials under operational humidity conditions. This will provide more relevant information for practical reuse.

It is important to acknowledge that the application of polymeric coatings not only modifies the hydrophobic behavior of the natural fibers but also alters their surface morphology and increases their mass. This study did not quantify the specific surface area reduction caused by the coatings, which could have an impact on the condensation and water harvesting efficiency. Future studies should address this limitation by evaluating the relationship between surface area loss, added polymer mass, and water collection performance in order to optimize the coating application.

REFERENCES

[1] J. A. Pascual, M. F. Naranjo, R. Payano, y Ojilve Ramon Medrano Perez, «Tecnología para la recolección de agua de niebla», 2011, doi: 10.13140/RG.2.1.4806.7048.

[2] H. Yue, Q. Zeng, J. Huang, Z. Guo, y W. Liu, «Fog collection behavior of bionic surface and large fog collector: A review», *Adv. Colloid Interface Sci.*, vol. 300, p. 102583, feb. 2022, doi: 10.1016/j.cis.2021.102583.

[3] M. Qadir, G. Jiménez, R. Farnum, L. Dodson, y V. Smakhtin, «Fog Water Collection: Challenges beyond Technology», *Water*, vol. 10, n.º 4, p. 372, mar. 2018, doi: 10.3390/w10040372.

[4] D. V. Carrera-Villacrés, I. C. Robalino, F. F. Rodríguez, W. R. Sandoval, D. L. Hidalgo, y T. Toulkeridis, «An Innovative Fog Catcher System Applied in the Andean Communities of Ecuador», *Trans. ASABE*, vol. 60, n.º 6, pp. 1917-1923, 2017, doi: 10.13031/trans.12368.

[5] C. A. A. Corredor, V. Buitrago, S. J. D. Ayala, T. Ambiental, K. Y. C. Almeida, y T. Ambiental, «Propuesta De Un Sistema De "Atrapa-Nieblas", Como Fuente De Agua No Convencional En La Vereda La Fuente, Municipio De Los Santos, Departamento De Santander.», p. 9, 2017.

[6] S. Korkmaz y İ. A. Kariper, «Fog harvesting against water shortage», *Environ. Chem. Lett.*, vol. 18, n.º 2, pp. 361-375, mar. 2020, doi: 10.1007/s10311-019-00950-5.

[7] Y. Cheng *et al.*, «Fog catcher brushes with environmental friendly slippery alumina micro-needle structured surface for efficient fog-harvesting», *J. Clean. Prod.*, vol. 315, p. 127862, sep. 2021, doi: 10.1016/j.jclepro.2021.127862.

[8] S. Zhang, J. Huang, Z. Chen, y Y. Lai, «Bioinspired Special Wettability Surfaces: From Fundamental Research to Water Harvesting Applications», *Small*, vol. 13, n.º 3, p. 1602992, ene. 2017, doi: 10.1002/smll.201602992.

[9] J. Arutchelvi, M. Sudhakar, A. Arkatkar, M. Doble, S. Bhaduri, y P. V. Uppara, «Biodegradation of polyethylene and polypropylene», *Indian j biotechnol*, p. 15, 2008.

[10] K. L. Pickering, M. G. A. Efendy, y T. M. Le, «A review of recent developments in natural fibre composites and their mechanical performance», *Compos. Part Appl. Sci. Manuf.*, vol. 83, pp. 98-112, abr. 2016, doi: 10.1016/j.compositesa.2015.08.038.
[11] Y. G. Thyavihalli Girijappa, S. Mavinkere Rangappa, J. Parameswaranpillai, y S. Siengchin, «Natural Fibers as Sustainable and Renewable Resource for Development of Eco-Friendly Composites: A Comprehensive Review», *Front. Mater.*, vol. 6, p. 226, sep. 2019, doi: 10.3389/fmats.2019.00226.
[12] A. Ali *et al.*, «Hydrophobic treatment of natural fibers and their composites—A review», *J. Ind. Text.*, vol. 47, n.º 8, pp. 2153-2183, may 2018, doi: 10.1177/1528083716654468.
[13] M. Rajaram, X. Heng, M. Oza, y C. Luo, «Enhancement of fog-collection efficiency of a Raschel mesh using surface coatings and local geometric changes», *Colloids Surf. Physicochem. Eng. Asp.*, vol. 508, pp. 218-229, nov. 2016, doi: 10.1016/j.colsurfa.2016.08.034.
[14] Y. Wan, J. Xu, Z. Lian, y J. Xu, «Superhydrophilic surfaces with hierarchical groove structure for efficient fog collection», *Colloids Surf. Physicochem. Eng. Asp.*, vol. 628, p. 127241, nov. 2021, doi: 10.1016/j.colsurfa.2021.127241.
[15] C. M. Regalado y A. Ritter, «The design of an optimal fog water collector: A theoretical analysis», *Atmospheric Res.*, vol. 178-179, pp. 45-54, sep. 2016, doi: 10.1016/j.atmosres.2016.03.006.
[16] S. A. Gómez-Suarez y E. Córdoba-Tuta, «Composite materials reinforced with fique fibers – a review», *Rev. UIS Ing.*, vol. 21, n.º 1, ene. 2022, doi: 10.18273/revuin.v21n1-2022013.
[17] J. de D. Rivera y D. Lopez-Garcia, «Mechanical characteristics of Raschel mesh and their application to the design of large fog collectors», *Atmospheric Res.*, vol. 151, pp. 250-258, ene. 2015, doi: 10.1016/j.atmosres.2014.06.011.
[18] M. R. Sanjay, G. R. Arpitha, L. L. Naik, K. Gopalakrishna, y B. Yogesha, «Applications of Natural Fibers and Its Composites: An Overview», *Nat. Resour.*, vol. 07, n.º 03, pp. 108-114, 2016, doi: 10.4236/nr.2016.73011.
[19] J. I. P. Singh, S. Singh, y V. Dhawan, «Influence of fiber volume fraction and curing temperature on mechanical properties of jute/PLA green composites», *Polym. Polym. Compos.*, vol. 28, n.º 4, pp. 273-284, may 2020, doi: 10.1177/0967391119872875.
[20] J. de D. Rivera y D. Lopez-Garcia, «Mechanical characteristics of Raschel mesh and their application to the design of large fog collectors», *Atmospheric Res.*, vol. 151, pp. 250-258, ene. 2015, doi: 10.1016/j.atmosres.2014.06.011.
[21] M. N. A. M. Taib y N. M. Julkapli, «4 - Dimensional stability of natural fiber-based and hybrid composites», en *Mechanical and Physical Testing of Biocomposites, Fibre-Reinforced Composites and Hybrid Composites*, M. Jawaid, M. Thariq, y N. Saba, Eds., en Woodhead Publishing Series in Composites Science and Engineering. , Woodhead Publishing, 2019, pp. 61-79. doi: https://doi.org/10.1016/B978-0-08-102292-4.00004-7.
[22] L. Vásquez-Ramírez, L. Cieza-León, y D. Cieza-León, «Efficiency of water collection for three types of mesh trappers in rural highlands of the northern highlands of Peru», n.º 3, p. 9, 2020.
[23] M. Rajaram, X. Heng, M. Oza, y C. Luo, «Enhancement of fog-collection efficiency of a Raschel mesh using surface coatings and local geometric changes», *Colloids Surf. Physicochem. Eng. Asp.*, vol. 508, pp. 218-229, nov. 2016, doi: 10.1016/j.colsurfa.2016.08.034.
[24] O. Klemm *et al.*, «Fog as a fresh-water resource: overview and perspectives.», *Ambio*, vol. 41, n.º 3, pp. 221-234, may 2012, doi: 10.1007/s13280-012-0247-8.

Sergio Andrés Gómez S. Was born in Bucaramanga, Colombia in 1988. He received the B.S in mechanical engineering and industrial engineering from Universidad Pontificia Bolivariana. Magister in industrial engineering from Universidad Pamplona, Colombia. He is currently professor and researcher in the Faculty of Engineering at Universidad Pontificia Bolivariana.
ORCID: https://orcid.org/0000-0002-6425-7062.

Edwin Córdoba Tuta. was born in Bucaramanga, Colombia in 1974. He received the B.S in mechanical engineering from Universidad Industrial de Santander. Magister in Mechanical engineering from Universidad Industrial de Santander, Colombia. He is currently professor and researcher in the Faculty of Engineering at Universidad Pontificia Bolivariana.
ORCID: https://orcid.org/0000-0001-8298-5007

Model for location of temporary shelters and routing of specialized personnel for assisting vulnerable population in case sudden natural disasters

Modelo de localización de albergues temporales y ruteo de personal especializado para la atención de población vulnerable ante un desastre natural súbito

K. V. Marín-Gómez; J. D. Morales-Corredor; C. A. Rojas-Trejos

DOI: https://doi.org/10.22517/23447214.25487
Scientific and technological research paper

***Abstract*— Natural disasters have long affected populations worldwide, resulting in significant consequences for both people and their environments. Most disasters occur suddenly, and several studies have highlighted logistical weaknesses in both the prevention and response to such events. In Colombia, flooding is the leading cause of death from hydrometeorological phenomena. Based on this context, the present research proposes a mixed-integer linear programming model for the location of temporary relief centers and the routing of specialized personnel. These decisions were addressed in two phases. In the first phase, the location of temporary shelters was determined, considering designated safe zones as candidate sites. In the second phase, route planning for specialized personnel was carried out, using a local distribution center as a reference point. The results indicate that addressing both decision-making processes and their interrelation contributes to minimizing response times for vulnerable populations.**

***Index Terms*— Aid distribution, Humanitarian logistics, Location of shelters, Mixed integer linear programming, Natural disasters.**

***Resumen*— Los desastres naturales han afectado a la población mundial, provocando consecuencias relevantes para las personas y su entorno. La mayoría de los desastres ocurren repentinamente, por lo que varios estudios han detectado deficiencias logísticas en el momento de prevenir y afrontar estos casos. En Colombia, las inundaciones son la principal causa de muertes por fenómenos hidrometeorológicos. Con base en lo anterior, la presente investigación propone un modelo de programación lineal entera mixta para la localización de albergues temporales y enrutamiento de personal especializado. Estas decisiones se abordaron en dos fases. En la primera fase se estableció la localización de albergues temporales considerando como sitios candidatos las zonas seguras destinadas para el emplazamiento de estas instalaciones. Los albergues seleccionados, permitió establecer en la segunda fase la configuración de rutas del personal especializado, considerando un centro local de distribución. Como resultado obtenido se pudo establecer que abordar ambos esquemas de decisión y su interrelación, contribuye a la minimización de los tiempos de respuesta a la población vulnerable.**

***Palabras claves*— Desastres naturales, Distribución de ayuda, Localización de albergues, Logística humanitaria, Programación Lineal Entera Mixta.**

I. INTRODUCTION

A natural disaster is defined as a phenomenon caused by nature, influenced by human activity. In 2011, there were more than 30,000 fatal victims, and 245 million people affected worldwide. This caused economic losses estimated at 386 trillion dollars, due to various disasters. In Colombia, 28 thousand events have been recorded in the last 40 years, 60% of which happened between 1990 and 2011; damage costs amounted to 623 billion dollars for the 2010-2011 winter season [1]. This shows that there has been an increase in the occurrence of disasters, which directly and indirectly affect the population and their environment.

From an engineering point of view, the most important challenge concerning humanitarian logistics is coordinating logistics activities [2] and managing resources in chaotic environments together with the uncertainty of emergency situations, which is a complex logistical task [1].

In Colombia, according to the National Risk and Disaster Management Unit, some of the most common disasters include floods, earthquakes, and landslides. These events have a negative effect on communities, and cause infrastructure damage and endanger human lives.

This manuscript was submitted on November 23, 2023. Accepted on March 04, 2025. And published on March 31, 2025.

Karen Vanessa Marín Gómez: Escuela de Ingeniería Industrial, Universidad del Valle Cali, Colombia, Postal code 760042. (email: karen.vanessa.marin@correounivalle.edu.co).

Jessica Dayan Morales Corredor: Escuela de Ingeniería Industrial, Universidad del Valle Cali, Colombia, Postal code 760042. (email: jessica.dayan.morales@correounivalle.edu.co).

Carlos Alberto Rojas Trejos: Departamento de Ingeniería de Sistemas e Industrial, Facultad de Ingeniería, Universidad Nacional de Colombia, Bogotá, Colombia, Postal code 111321.

Escuela de Ingeniería Industrial, Universidad del Valle, Cali, Colombia, Postal code 760042. (email: crojast@unal.edu.co).

Table 1 shows the distribution of these disasters that occurred in Colombia between 1906 and 2019. This also shows that floods, landslides, and earthquakes have the highest accumulated percentage of occurrences, totaling 81.77%.

TABLE I
DISTRIBUTION OF EVENTS IN COLOMBIA (1906-2019).

Type of Natural Disaster	Absolute Frequency (# events)	Relative Frequency	Accumulated Relative Frequency	Total frequency affected (# persons)	Total deaths (# persons)	Total damages (thousands of dollars)
Flooding	86	44.79%	44.79%	16.356,3	3.585	3.591,353
Landslide	43	22.40%	67.19%	76.082	3.460	102.400
Earthquake	28	14.58%	81.77%	1.460,39	3.972	2.318,666
Volcanic activity	11	5.73%	87.50%	56.964	22,826	1,000,000
Storm	9	4.69%	92.19%	140.415	49	53.050
Epidemic	6	3.13%	95.31%	121.194	672	0
Forest fire	3	1.56%	96.88%	200	31	0
Mass movement (dry)	3	1.56%	98.44%	2.411	247	0
Drougth	2	1.04%	99.48%	100.000	0	0
Insect infestation	1	0.52%	100%	0	0	104.000
Total	192	100%		18,314,58	34,842	7.169,469

Source: UNGRD

The most significant type of disaster is flooding, which occurs due to heavy or sudden rainfall, causing the rising and overflowing of rivers. These events happen abruptly and, in many cases, are difficult to control, causing material losses and affecting people's health. This may cause the loss of human lives. During a flood, the current of rivers often carries everything in its path such as animals, trees, rocks, and various debris.

Since floods are difficult to be controlled, in necessary to have a contingency plan for efficient and timely post-disaster response. To achieve this purpose, it is important to know the affected area and its limitations. Help centers and resource distribution centers should be strategically located, and trained personnel efficiently deployed so that they can respond to emergencies of this nature.

To fulfill this purpose, it is necessary to go to a branch of logistics that is responsible for providing timely aid to populations that may be affected before, during and after a disaster. This branch is called Humanitarian Logistics, which have the same fundamental principles logistics: planning, implementing and controlling the flow of information, resources or personnel from a point of origin to a destination. While logistics aims to meet customer needs, humanitarian logistics aims to alleviate the suffering of affected populations in a timely manner.

Based on this, this project aims to develop a mixed-integer linear programming model whose objective is to minimize response time for victims by optimizing the distribution of resources between relief centers and temporary shelters. Specialized personnel are also present to provide timely assistance to victims in a municipality of the Valle del Cauca department, Colombia. Thus, a diagnosis of the area will be conducted to assess the effects of natural disasters and identify the most relevant dangers, which will be the bases for formulating the mathematical model.

II. LITERATURE REVIEW

Humanitarian logistics according to [3] is the process of planning, implementing and efficiently controlling the flow of products, materials and information from individuals and donor organizations to victims, so that their survival needs are met. Humanitarian logistics emerged as a response to the increase of natural disasters worldwide, which have affected some populations. According to data from the International Strategy Program for Disaster Reduction (ISDR), in 2004, there were 305 natural disasters in the world. As these natural disasters continue to increase, and consequently the survival needs of affected populations increase as well. Populations affected by these adversities face such multiple problems as inadequate shelter, lack of food, limited access to medical care, insufficient drug supply and the need for psychological support, particularly in cases of family loss. [4].

The authors [4] identify the challenges facing humanitarian logistics today are increasing, and they are much more complex; these issues are the speed aid of delivery, the movement of affected populations in conflict areas, the influence of humanitarian teams, deficiencies in ONG`s capabilities, lack of knowledge, limited investment in technology and communication.

Humanitarian logistics is generally divided into three stages, which are aligned with the well-known traditional risk management model: pre-disaster, during the disaster, and post-disaster**.** These stages have several subcategories [5]. These are shown in Fig.1.

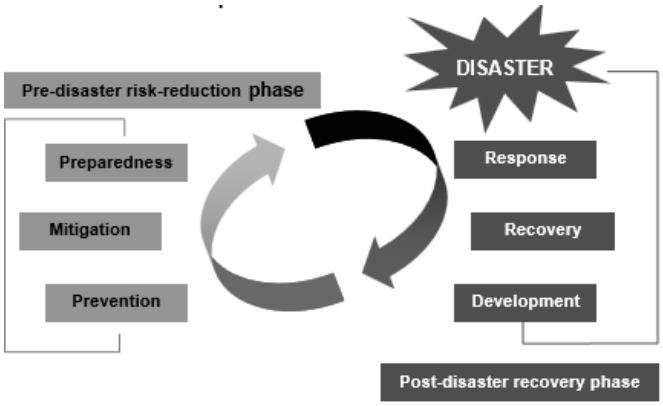


Fig. 1. Traditional Risk Management Model Source: Adapted of [5]

Authors such as [6] state that in natural disasters such as floods, the affected population is partially isolated and vulnerable. Research in Colombia highlights the importance of providing timely post-disaster support. According to IFRC (2010), cited by [7], such "organizations as the Red Cross in Colombia play an important role in developing training programs for both the population and institutions: These organizations have also established a series of protocols and guidelines to effectively manage crisis situations".

The post-disaster situation occurs in an environment

characterized by uncertainty regarding aid resources and shelters for victims. The latter is critical, since many houses are damaged or destroyed. Therefore, families are forced to seek alternative accommodation until a permanent housing solution can be found [8].

Regarding humanitarian logistics [9], this presents a qualitative analysis of inventory management strategies in the humanitarian logistics operations. They state that, each year, many natural and human-induced disasters affect thousands of people around the world. During these disasters, both government agencies and humanitarian organizations face logistical challenges; their main objective is to meet the needs of affected people and alleviate their suffering. To achieve this, an effective inventory management strategy plays a crucial role at every stage of the supply chain.

In terms of routing trained personnel and vehicles for the care of vulnerable populations, some authors have researched this topic, including:

In the emergency logistics planning, natural disaster [10] pose one of the most common problems in logistics: the Vehicle Routing Problem (VRP), which involves a set of clients (each represented as a destination node) that must be served by *m* identical vehicles located at a warehouse. Each vehicle is to return to the depot after completing its route, and its load cannot exceed its capacity at any point of the trip. Additionally, each client can be visited only once, and it is assumed that the vehicle's carrying capacity exceeds the demand of any individual client. Thus, the main objective is to minimize the total distance traveled on each route.

According to [11], specialized literature has enough information on the application of metaheuristics in route planning route. However, most research in this field have been conducted under normal conditions (e.g., standard weather conditions). The problem of route planning for repairing electrical faults can be basically modeled Capacitated Multiple Traveling Salesman Problem (CMPS)**,** due to some significant similarities with this well-known variant of the theoretical VRP. These similarities are related to the dispatching of a homogeneous fleet of vehicles (with repair technicians deployed in vulnerable areas), where each vehicle is assigned a set of nodes (affected areas) similar to MTSP (Multi-Traveling Salesman Problem). Each node is once visited by a single vehicle (or salesman).

Other authors as [12] have developed a structured plan for distributing humanitarian aid through vehicle routing in the event of a major earthquake in Lima Metropolitan and Callao. They use the Great Route method together with Linear Programming, as this minimizes the actual distance traveled and reduces transportation costs. The number of victims was also considered to optimize the allocation of resources and supplies to be used for each trip by the terrestrial vehicular fleet.

In the work of [13], a bi-level optimization model is presented for sending, receiving and distributing in-kind assistance after a natural disaster has occurred. This aims to determine the optimal configuration of shipments and the most efficient distribution method for delivering supplies affected areas through various transportation modes.

A key issue in humanitarian logistics, widely recognized by researchers, is the strategic location of temporary relief or facilities to provide timely resources and specialized assistance in affected areas.

Concerning this issue, [14] propose an approach to the problem of locating temporary relief facilities for households affected by severe natural disasters. They characterize both the demand and supply of temporary relief, identifying high-risk areas based on the type of disaster that may occur. The model's performance function seeks to minimize the weighted distances between temporary relief facilities and affected households, considering constraints such as ensuring full demand from each household type, facility capacity limitations, and the maximum number of temporary shelters that can be built.

Similarly, author [15] propose a bi-criteria model for the location of temporary relief centers. This model includes the design an evacuation plan to support and ensure the safety of the affected population in case of a flood. This includes the opening of a temporary relief and distribution centers, pre-positioning of aid package inventories, and the assignment of individuals to temporary shelters and evacuation routes.

Some of the optimization models in humanitarian logistics used for facility location, which is the focus of this research, include deterministic single-objective modes developed by [16-19], and stochastic models developed by [20-22], among others.

Based on the analyses of previous studies, it is necessary to analyze the interrelationship between decisions related to the location of distribution points and temporary shelters, and the distribution of humanitarian aid. This analysis is to consider limitations in the availability and capacity of humanitarian assistance or support units in affected areas.

Unlike the reviewed literature on humanitarian logistics this research considers the following:

- ✓ A mathematical formulation with hierarchical approach that allows integrating decisions on the location of temporary relief centers and the routing of specialized personnel.
- ✓ The allocation of temporary shelters and assignment of homeless individuals to the temporary shelters.
- ✓ Supplies allocation according to the number of victims assisted by open temporary shelters
- ✓ The routing of specialized personnel and their assignment to open temporary shelters
- ✓ Capacity constraints in temporary shelters and limitations in specialized personnel availability.

Therefore, this research proposes a two-level mode. The first

level aims to determine the optimal location of a distribution point and temporary shelters to assist populations that have been affected by a sudden natural disaster. The second level has to do with decisions on the distribution of humanitarian aid in the last-mile by considering limited resources and capacity issues in care units.

III. METHODOLOGY

A review of the literature on location and routing models will be conducted by various authors who have addressed these topics related to natural disasters. This review will be of great help as a basis for developing mathematical models to achieve an optimal post-disaster solution Fig. 2.

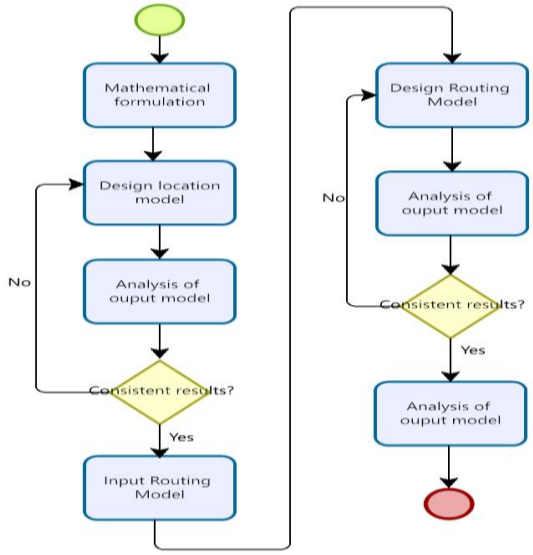


Fig. 2. Description of Methodology. Source: Authors

The formulation of mathematical models will be presented, defining input parameters, decision variables, the objective function, and constraints. These models are validated through a case study to determine whether the results are aligned with real-world contexts.

The methodology is structured in two phases. First, the location model is solved by using information provided by the case study; thus, output variables are obtained, which serve as input parameters for the routing model following a hierarchical approach. This produces output variables or model results that are to be analyzed to verify if they meet the objectives of each component. Finally, conclusions are reached based on these findings.

A. Model of location for temporary relief centers

This location model is a mathematical model formulated as a mixed-integer linear programming (MILP), which has been designed to determine the best location of a distribution center and different temporary relief centers. This done to minimize travel time for delivering supplies from the distribution center to temporary shelters, ensuring timely assistance to vulnerable populations.

Assumptions.

- The municipality in this study case lacks specialized and adequate infrastructure for establishing temporary relief centers during a flood disaster. Therefore, some educational facilities in both urban and rural areas are used instead.
- The areas affected and requiring evacuation due to such disasters are known with certainty.
- The victims are not relocated to the homes of relatives, friends or neighbors.
- Victims arrive at the assigned temporary relief.
- Supplies include temporary kits (including a pillow, mat and blanket), which is required by each victim.
- It is assumed that each victim uses only one kit.
- The kits are donated and managed by national and departmental institutions.
- The kits are pre-positioned at the Distribution Center.
- Travel times and number of victims are considered deterministic.

Main Sets

- CNS: $set\ of\ secure\ nodes\ in\ the\ region$
- CD: $subset\ safe\ nodes\ CNS\ secure\ zones\ for\ the$
- $installation\ of\ the\ distribution\ center\ indexed\ by\ i$
- CA: $subset\ of\ secure\ nodes\ CNS\ secure\ zones\ for$
- $the\ installation\ of\ temporary\ relief\ indexed\ by\ j$
- CV: $set\ of\ flooded\ areas\ of\ the\ region\ where\ they\ are$
- $the\ victims\ indexed\ by\ k$

Parameters

- $TIEMPCV_{kj}$: $tranfer\ time\ from\ the\ regions\ where\ victims\ are\ located\ in\ zone\ k\ to\ the\ temporary\ relief\ center\ in\ zone\ j$
- $DAMN_k$: $DVictims\ in\ the\ area\ k\ who\ require\ assistance$
- C_j: $capacity\ of\ the\ temporary\ relief\ center\ in\ zone\ j$ $Number\ of\ victims\ who\ can\ be\ assisted$
- U: $number\ of\ arbitrary\ supplies$
- $CONSD$: $consumption\ of\ supplies\ (kits) required\ per\ victim$
- N_i: $capacity\ of\ the\ distribution\ center\ (supplies)$

Decision Variables

- M_i $Binary$: 1 if the distribution center is located in zone i; 0 otherwise.
- W_j $Binary$: 1 if the temporary relief center is located in zone j; 0 otherwise
- G_{ij} $Binary$: 1 if the distribution center serves the temporary relief center j; 0 otherwise
- D_{kj} : number of victims in area k who are served

in the temporary relief center in zone j

- Y_{kj} *Binary*: 1 if the victims of the area k

are assigned to the temporary relief center in zone j; 0 otherwise

- Q_{ij}: quantity of supplies sent from the distribution center in zone j

Objective Function

Minimize travel time from the distribution center to the temporary relief centers so that affected victims can be assisted [minutes].

$$\sum_{\forall i,j} TIEMPCD_{ij} * G_{ij} + \sum_{k,j} TIEMPCV_{k,j} * Y_{kj}$$

Constrains

$$M_i + W_j \leq 1 \quad \forall i \in CD, j \in CA, i = j \quad (1)$$

$$\sum_{i \in CD} M_i = 1 \quad (2)$$

$$\sum_{i \in CD} G_{ij} = W_j \quad \forall j \in CA \quad (3)$$

$$\sum_{k \in CV} D_{kj} \leq \sum_{k \in CV} DAMN_k * W_j \quad \forall j \in CA \quad (4)$$

$$D_{kj} = Y_{kj} * DAMN_k \quad \forall j \in CA, k \in CV \quad (5)$$

$$\sum_{j \in CA} Y_{kj} = 1 \quad \forall k \in CV \quad (6)$$

$$\sum_{k \in CV} D_{kj} \geq 0.1 * C_j * W_j \quad \forall j \in CA \quad (7)$$

$$\sum_{k \in CV} DAMN_k * Y_{kj} \leq C_j * W_j \quad \forall j \in CA \quad (8)$$

$$Q_{ij} \leq G_{ij} * U \quad \forall i \in CD, j \in CA \quad (9)$$

$$\sum_{j \in CA} Q_{ij} \leq N_i * M_i \quad \forall i \in CD \quad (10)$$

$$\sum_{i \in CD} Q_{ij} \leq CONSD * C_j * W_j \quad \forall j \in CA \quad (11)$$

$$\sum_{j \in CA} D_{kj} \geq DAMN_k \quad \forall k \in CV \quad (12)$$

$$\sum_{j \in CA} W_j \geq \frac{\sum_{K \in CV} DAMN_k}{\sum_{j \in CA} C_j} \quad (13)$$

$$\sum_{i \in CD} Q_{ij} \geq CONSD * \sum_{k \in CV} D_{kj} \quad \forall j \in CA \quad (14)$$

$$M_i \in \{0,1\} \quad \forall i \in CD \quad (15)$$

$$W_j \in \{0,1\} \quad \forall j \in CA \quad (16)$$

$$G_{ij} \in \{0,1\} \quad \forall i \in CD, j \in CA \quad (17)$$

$$Y_{kj} \in \{0,1\} \quad \forall k \in CV, j \in CA \quad (18)$$

$$D_{kj} \geq 0 \quad \forall k \in CV, j \in CA \quad (19)$$

$$Q_{ij} \geq 0 \quad \forall i \in CD, j \in CA \quad (20)$$

Where (1) is a zone that can only be used to locate a distribution center or temporary relief center; (2) Ensures the opening of a single distribution center i; (3) The distribution center can serve a temporary relief center j only if the temporary relief center is open; (4) Ensures that victims from each zone k are accommodated in a temporary relief center only if it is open; (5) The temporary relief center can serve the victims of a zone only if it has been designated to do so; (6) Victims from each affected area k must be assigned to a single temporary relief center j; (7) A temporary relief is opened only if the number of victims to be served exceeds 10% of its capacity; (8) Each open temporary relief center must have sufficient capacity to accommodate the victims from an affected area k; (9) supplies are sent from a distribution center to the temporary relief center only if the distribution center serves the temporary relief center; (10) The amount of supplies sent from a distribution center to a shelter should not exceed the distribution center's capacity; (11) The quantity of supplies sent from distribution centers for to temporary relief centers should not exceed the temporary relief center's capacity; (12) The total number of victims accommodated in temporary relief centers in zone j of k must be at least equal to the total number of victims in that area; (13) Ensures that the number of temporary relief centers is sufficient to serve the affected population; (14) Ensure that sufficient supplies are sent to meet the demand of the shelter; (15), (16), (17), (18), (19) and (20) Define the range of values that variables can take.

B. Specialized Personnel Routing Model

To develop the model shown below, the Capacitated Vehicle Routing Model with a homogeneous fleet was used as aa reference.

Assumptions

- Three support groups (Red Cross, Civil Defense and Fire Department) have the same skill and capacity to serve the victims of such disasters and are equipped with all necessary tools.
- Specialized people form a single support unit.
- The capacity for assistance is determined based on the number of victims and available staff.
- Restrictions on access roads are not considered.
- All specialized personnel are available at the distribution center.
- Travel times are deterministic
- Support unit transport all specialized personnel required to assist victims.

Main Sets

- CDA: Set of input vertices(distribution center and temporary relief centers indexed by i and j
- UA: Set of available support units indexed by k

Parameters

- T_{ij}: Travel time from the CDA i to the CDA j (Minutes)
- C: $Capacity\ for\ assistance\ by\ specialized\ personnel$)
- A_i: Shelter demand associated with each CDA i (Victims)
- NP: Number of specialized personel available in each support unit (Personnel)

Decision Variables

- P_{ijk} *Binary*: 1 if support unit k travels through the route from i to j; 0 otherwise
- D_{ik}: Demand at the Temporary Relief Center supplied by the support unit (Personnel)
- S_i: Additional variable that represents the unit's load capacity

of the support unit after visiting the temporary relief center i (Personnel)

Objective Function

Minimize travel time from the distribution center to all temporary relief centers.

$$\sum_{\forall i,j,k\ i\neq j} T_{ij} * P_{ijk}$$

Constraints

$$\sum_{i,k\ i\neq j} P_{ijk} \geq 1 \quad \forall j \in CDA \qquad (1)$$

$$D_{ik} = \frac{A_i}{C} * \sum_{j\in CDA} P_{ijk} \quad \forall i \in CDA, k \in UA: i \geq 1 \qquad (2)$$

$$\sum_{k\in UA: i\geq 1} D_{ik} = \frac{A_i}{C} \quad \forall i \in CDA \qquad (3)$$

$$\sum_{i\in CDA: i\geq 1} D_{ik} \leq NP \quad \forall k \in UA \qquad (4)$$

$$\sum_{i\in CDA\ i\neq n} P_{ink} - \sum_{j\in CDA\ j\neq n} P_{njk} = 0 \quad \forall n \in CDA, k \in UA \qquad (5)$$

$$S_i - S_j + NP * P_{ijk} \leq NP - \frac{A_i}{C} \quad \forall i,j \in CDA, k \in UA: i \neq j, i \neq 0, j \neq 0 \qquad (6)$$

$$\frac{A_i}{C} \leq S_i \quad \forall i \in CDA: i \neq 0 \qquad (7)$$

$$S_i \leq NP \quad \forall i \in CDA: i \neq 0 \qquad (8)$$

$$P_{ijk} \in \{0,1\} \quad \forall i,j \in CDA, k \in UA \qquad (9)$$

$$D_{ik} \geq 0 \quad \forall i \in CDA, k \in UA \qquad (10)$$

$$S_i \geq 0 \quad \forall i \in CDA \qquad (11)$$

$$S_j \geq 0 \quad \forall, j \in CD \qquad (12)$$

Where (1) ensures that the specialized personnel arrives at the temporary relief center; (2) indicates that the temporary relief center i can be served by specialized personnel if k passes *i*; (3) Ensures that the demand of each open temporary relief is fully satisfied ; (4) determines that the quantity of specialized personnel assigned to each temporary relief center does not exceed the available personnel; (5) Flow balance, which indicates that the support unit leaving the distribution center returns; (6), (7) and (8) Avoid sub-routes; (9), (10), (11) and (12) define range of values that the variables can take.

IV. CASE STUDY

The case study is framed within the municipality of Tuluá is located in southwestern Colombia, in the center of Valle del Cauca department between the Central Mountain Range and the Cauca River (CVC, 2017). It has territorial area 1,014.96 km^2. The total land area of the municipality is 910.55 km^2 of which 98.78% corresponds to the rural sector and the urban sector 1.22% (Municipality of Tulua, 2017).

According to figures from DANE, Tuluá has a population of 216,619 inhabitants, 187.121 live in the municipal capital and 29,483 in rural areas. (Tuluá Chamber of Commerce, 2016).
To validate the proposed models, the municipality of Tuluá was chosen as a reference, particularly those areas affected by the 2022 rainy season. Floods occur during the winter season due to the overflow of Tuluá River, which initially affects the following neighborhoods: Tomas Uribe (CV1), La Trinidad (CV2), Siete de Agosto (CV3), San Antonio (CV4), La Inmaculada (CV5), Villa Nueva (CV6), Casa Huertas (CV7), Portales Rio in urban area (CV8), the village of Bocas de Tulua (CV9), Tres Esquinas (CV10); Morales River affecting neighborhoods: Urbanization Villa (CV11), El Bosque (CV12) and Santa Rita (CV13). (CMGR, Consejo Municipal de Gestión del Riego de Desastres Municipio de Tuluá – disaster risk managment -, 2012). These places can be seen in fig. 3.

The proposed models provide an optimal solution that meet the needs of the affected population, safe areas were identified, where temporary relief centers could possibly be located. These safe places are: Coliseo de Ferias Manuel Victoria Rojas(CA1), Salesiano High School San Juan Bosco (CA2), Coliseo Benicio Echeverry (CA3), Stadium Doce de Octubre (CA4), Institución Educativa Aguaclara (CA5) Institución Educativa Técnica Occidente Tulua (CA6), Health Center La Independencia (CA7), Rubén Cruz Vélez (CA8), Institución Educativa Julia Restrepo (CA9), Gimnasio del Pacifico School (CA10), Institución Educativa Julio Cesar Zuluaga, Tres Esquinas Village (CA11), Institución Educativa Julio Cesar Zuluaga in the village Bocas de Tulua (CA12). These places can be seen in Figure 4. A distribution center was also identified, which has become a safe potential place: these are Coliseo de Ferias Manuel Victoria Rojas (CD1) and Salesiano High School San Juan Bosco (CD2), which can be seen in Figure 5 as well; the municipality of Tulua Secretary of government lacks suitable facilities of temporary relief centers; that is the reason why some state-owned educational institutions and some health centers are used. Although it is known that, when an emergency occurs, these centers should not be used since doing do would disrupt academic and labor activities, these have been used in this study due to the lack of alternative facilities.

They can be observed in Fig. 4; in the municipality of Tulua Secretary of government lacks suitable for the location of temporary shelter facilities, which is why we make use of official educational institutions and some health centers, although it is known that when an emergency that requires locating temporary relief should not resort to educational institutions and health centers, because this way we would be intervening in academic and labor activities, but as mentioned above the city and this work they will resort to facilities because they are not they have other available. The entities or groups equipped with specialized disaster relief personnel include the Fire Department, Civil Defense, and Red Cross.

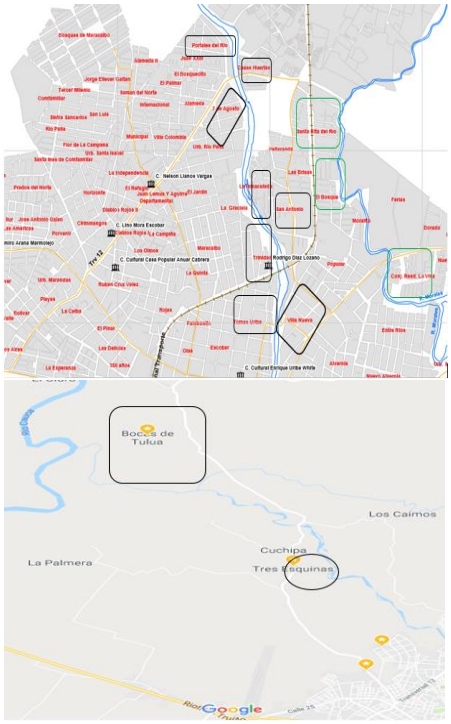


Fig. 3. Areas Flood irrigation. Source: Google Maps

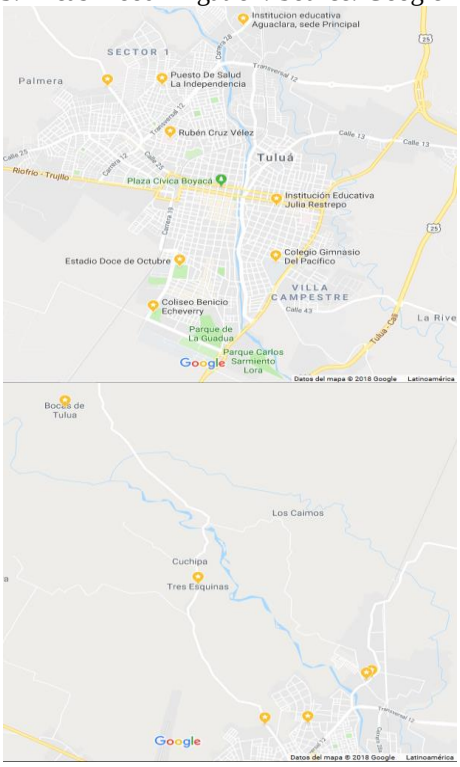


Fig. 4. Safe Zones for the location of temporary relief centers Source: Google Maps

Fig. 5. Safe areas for the location of distribution centers. Source: Google Maps

V. RESULTS

This section presents the results obtained based on the methodology developed. The results thrown by the CPLEX software in the NEOS Server platform and the mathematical programming language AMPL, indicate that the location model determined the opening of a Distribution Center No. 1 (CD1) and Temporary Relief Centers No. 5 and 11 (CA5, CA11). These temporary locations ensure a minimum travel time of 93 minutes while guaranteeing assistance for the entire vulnerable population.

Fig. 6 shows the location of the distribution center and temporary relief centers, while Figures 7 and 8 show the assignment of affected areas to each of the temporary relief center.

Fig. 6. Location of the distribution center and temporary relief Source: Google Maps

Fig. 7. Mapping affected areas to CA5.Source: Google Maps

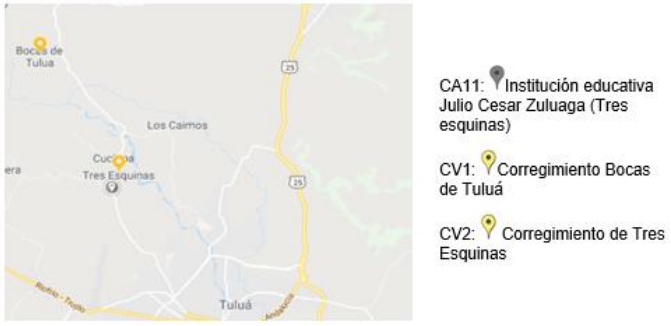


Fig. 8. Allocation of areas affected by CA11. Source: Google Maps

The number of victims and supplies assigned to each temporary relief center are shown in Table 2. Table 3 presents the results of supplies allocation from the distribution center to each temporary relief center.

TABLE II
VICTIMS OF EACH AFFECTED SITE THAT WILL BE TREATED AT EACH TEMPORARY RELIEF

CV \ AC	5	11
1	0	47
2	0	21
3	27	0
4	32	0
5	21	0
6	25	0
7	25	0
8	30	0
9	16	0
10	32	0
11	26	0
12	27	0
13	25	0

Source: Authors

TABLE III
SUPPLIES ALLOCATED FROM THE DISTRIBUTION CENTER FOR EACH TEMPORARY RELIEF

AC \ CD	1
5	286
11	68

Source: Authors

These results show that all affected areas within where the victims are in the municipal head are assisted by Institución Educativa Aguaclara Sede principal (CA5), which is also located in the city center and close to the distribution center Coliseo de Ferias Manuel Victoria Rojas (CD1). Meanwhile, affected rural areas are served by Institución Educativa Julio Cesar Zuluaga (CA11), which also receives supplies from the same distribution center (CD1).

The routing model that has been designed to minimize travel time and provide timely assistance to victims, determined that the support unit with specialized personnel leaves the Coliseo de Ferias Manuel Victoria Rojas (CD1), goes to the Institución Educativa Aguaclara temporary relief (CA5), where 96 specialized personnel are left. After that, it goes Institución Educativa Julio Cesar Zuluaga (CA11), where other 23 specialized personnel are left. Finally, it returns to point of departure (CD1). Table 4 shows the results produced by the model.

TABLE IV
RESULTS MODEL ROUTING OF SPECIALIZED PERSONNEL.

Specialized personnel (D) temporary relief (CA) open		
AC	CA5	CA11
D	96	2. 3

Source: Authors

- Travel time: 40 minutes
- Route: CD1 → CA5 → CA11 → CD1

It is observed in the base model of routing that the path that specialized staff starts at the distribution center CD1, there is going to lodge to CA5 where 80% of specialist staff meets the demand of the temporary relief because most (+ - 80%) of the victims are concentrated there, so that is necessary more specialist staff to alleviate the suffering of those affected by a disaster natural, thence he goes to meet the other temporary shelter open CA11 which has 20% of victims, generating a time minimum of 40 minutes of transfer from support unit with specialized personnel to temporary relief.

These models determine the number of temporary relief centers that are to be opened by, identify optimal locations and distribution centers with good capacity, assign victims to shelters, and also identify the routing of specialized personnel. This is aimed to seek a solution that generates the shortest possible time in assisting the affected victims. The combined results related to Location and Routing are shown in Fig. 9.

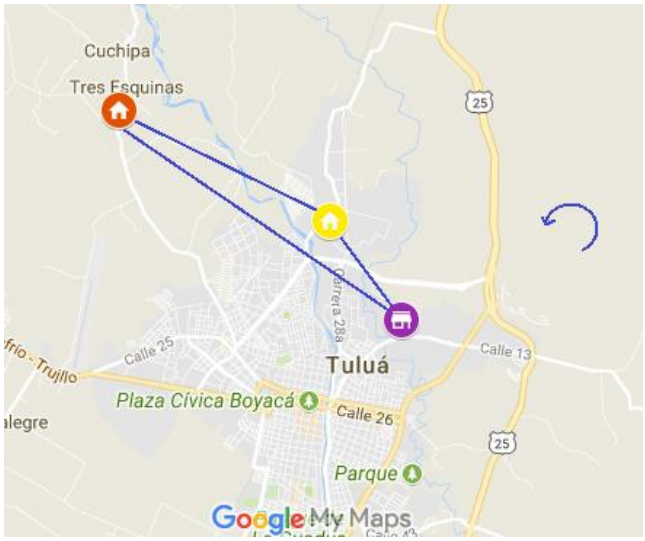


Fig. 9. Results Model Location and routing of specialized personnel.Source: Google Maps

Although the location model provides a decision-making decision solution for placing temporary relief centers to assist flood victims in the municipality of Tulua, and the routing model ensures a solution that meets the objective of minimizing travel time for the support unit and specialized personnel. Thus, further scenario analysis is required based on the variation of parameters of interest, so that the behavior of the location and routing model can be evaluated to improve response strategies and optimize decision-making in the event of a flood.

Addressing the decisions about the location of temporary shelters and the configuration of routes for specialized personnel together recognizes the connection between both decision-making processes. This integrated approach helps improve response times for affected populations compared to tackling these decisions separately, as shown in the studies conducted by [8], [16], and [17]. Treating these decisions individually may lead to suboptimal solutions [2].

VI. Conclusion

In emergency situations caused by natural disasters, the most important issue is to have a well-structured emergency response to meet victims' needs and assist them as soon as possible. The objective of this study is to minimize time from the occurrence of a disaster to the arrival of aid in affected areas.

To achieve this, a location model was proposed to determine the number and location of temporary shelters, which should be served by a single distribution center. This model also took into account allocation of victims and necessary supplies; this also accounts for the capacity of both shelters and the distribution center. A routing model was also designed to optimize the movement of specialized personnel to provide timely assistance to vulnerable population and the assignment of personnel, according to the needs of each shelter.For this purpose, the mathematical programming language AMPL was used, as well as the NEOS Server platform and the CPLEX solver to obtain and analyze quantitative results of the proposed models.

Unlike previous studies in humanitarian logistics, this research employs a mathematical hierarchical approach, which allows the integration of location decisions for distribution centers and temporary shelters with the routing of specialized personnel. This model considers issues related to victim assignment to shelters, supply assignment according to the number of victims served by open shelters, and assignment of specialized personnel, while accounting for the limited capacity of relief units.

For future research, it is necessary to improve models for temporary relief centers and routing of specialized personnel by incorporating stochastic travel and setup times and number of victims. Future studies could also include cost models related to opening and operating a temporary shelter and distribution centers. Furthermore, the impact environmental factors in the performance function are considered as well.

Acknowledgment

The study was funded by Bicentennial Doctoral Excellence Scholarship Program of the Ministry of Science, Technology and Innovation, defined in Article 45 of Law 1942 of 2018 of the General System of Royalties in Colombia. The authors of the research would like to acknowledge to SEPRO and GEDESC research groups of Universidad Nacional de Colombia and Universidad del Valle, respectively, for their valuable comments.

Acknowledgment the National Risk and Disaster Management Unit (UNGRD, by its Spanish acronym) of Tulua city, Colombia.

Declaration of interests

The authors declare no conflict of interest.

References

[1] N. Morshedlou, A. D. González, and K. Barker, "Work crew routing problem for infrastructure network restoration," Transp. Res. Part B Methodol., vol. 118, pp. 66-89, 2018.W.-K. Chen, Linear Networks and Systems. Belmont, CA, USA: Wadsworth, 1993, pp. 123-135. https://doi.org/10.1016/j.trb.2018.10.001

[2] M. S. Habib, Y. H. Lee, and M. S. Memon, "Mathematical Models in Humanitarian Supply Chain Management: A Systematic Literature Review," Math. Probl. Eng., 2016. https://doi.org/10.1155/2016/3212095

[3] Y.-H. Lin, R. Batta, P. A. Rogerson, A. Blatt, and M. Flanigan, "A logistics model for emergency supply of critical items in the aftermath of a disaster," Socioecon. Plann. Sci., vol. 45, no. 4, pp. 132-145, 2011. https://doi.org/10.1016/j.seps.2011.04.003

[4] Y. Xin Chen, P. R. Tadikamalla, J. Shang, and Y. Song, "Supply allocation: bi-level programming and differential evolution algorithm for Natural Disaster Relief," Cluster Comput., vol. 23, no. 1, pp. 203-217, 2020. https://doi.org/10.1007/s10586-017-1366-6

[5] G. Kovacs and K. Spens, "Identifying challenges in humanitarian logistics," Int. J. Phys. Distrib. Logist. Manag., vol. 39, no. 6, pp. 506-528, 2009.https://doi.org/10.1108/09600030910985848

[6] P. J. Ceballos-Parra, W. A. Sarache, and D. M. Gómez, "A bibliometric analysis of trends in humanitarian logistic [Un Análisis Bibliométrico de las Tendencias en Logística Humanitaria]," Inf. Tecnol., vol. 29, no. 1, pp. 91-104, 2018. https://doi.org/10.4067/S0718-07642018000100091

[7] M. Ahmadi, A. Seifi, and B. Tootooni, "A humanitarian logistics model for disaster relief operation considering network failure and standard relief time: A case study on San Francisco district," Transp. Res. Part E-Logistics Transp. Rev., vol. 75, pp. 145-163, 2015. https://doi.org/10.1016/j.tre.2015.01.008

[8] N. Al Theeb and C. Murray, "Vehicle routing and resource distribution in postdisaster humanitarian relief operations," Int. Trans. Oper. Res., vol. 24, no. 6, pp. 1253-1284, 2017. https://doi.org/10.1111/itor.12308

[9] A. Edrissi, M. Nourinejad, and M. J. Roorda, "Transportation network reliability in emergency response," Transp. Res. Part E Logist. Transp. Rev., 2015.https://doi.org/10.1016/j.tre.2015.05.005

[10] O. Rodríguez-Espíndola, P. Albores, and C. Brewster, "Disaster preparedness in humanitarian logistics: A collaborative approach for resource management in floods," Eur. J. Oper. Res., 2018. https://doi.org/10.1016/j.ejor.2017.01.021

[11] Clarke, G. and Wright, W.. "Scheduling of vehicles from a central depot to a number of delivery points". Operations Research, 1964, No. 12 pp. 568-581.https://doi.org/10.1287/opre.12.4.568

[12] Cornejo, C., Vargas, J., Aragon-Casa, L., Serpa-Oshiro, "Localización de almacenes y distribución de ayuda humanitaria para atención de damnificados en caso de desastre natural", 2016.

[13] Brzezicki A., "Modelos Matemáticos en Gestión de Desastres y Logística Humanitaria". Instituto de Matemáticas Universidad de Sevilla, 2017, España.

[14] Y. Xin Chen, P. R. Tadikamalla, J. Shang, and Y. Song, "Supply allocation: bi-level programming and differential evolution algorithm for Natural Disaster Relief," Cluster Comput., vol. 23, no. 1, pp. 203-217, 2020. https://doi.org/10.1007/s10586-017-1366-6

[15] Y.-H. Lin, R. Batta, P. A. Rogerson, A. Blatt, and M. Flanigan, "A logistics model for emergency supply of critical items in the aftermath of a disaster," Socioecon. Plann. Sci., vol. 45, no. 4, pp. 132-145, 2011. https://doi.org/10.1016/j.seps.2011.04.003

[16] Arroyo-López, P. E, Colón, R. E., Iniestra-Gaytán, J., "Un modelo bi-criterio para la ubicación de albergues, como parte de un plan de evacuación en caso de inundaciones", 2012.

[17] Rennemo, S. J., & Hvattum, L. M. "A three-stage stochastic facility routing model for disaster response planning". Transportation Research Part E: Logistics and Transportation Review, 2014, pp. 116-135. https://doi.org/10.1016/j.tre.2013.12.006

[18] Morales Celis, L. X., & Palacios Valencia S. A. "Localización de instalaciones y asignación de recursos mediante un algoritmo heurístico, para la atención pos-desastre, en caso de inundación o remoción de masa", 2017, Bucaramanga, Colombia.

[19] L. Özdamar and M. A. Ertem, "Models, solutions and enabling technologies in humanitarian logistics," Eur. J. Oper. Res., vol. 244, no. 1, pp. 55-65, 2015. https://doi.org/10.1016/j.ejor.2014.11.030

[20] Wei, L., & Li, W. "Decision Support for Urban Shelter Locations Based on Covering Model". Procedia Engineering, 2012, pp. 43, 59 - 64. https://doi.org/10.1016/j.proeng.2012.08.011

[21] Xu, J., & Yin, X. "Multi-criteria location model of earthquake evacuation shelters to aid". International Journal of Disaster Risk Reduction, 2016, Vol. 20, 51- 62. https://doi.org/10.1016/j.ijdrr.2016.10.009

[22] Trivedi, A., & Singh, A. "Prioritizing emergency shelter areas using hybrid multi-criteria decision approach: a case study". J. Multi-Criteria Decis. Anal, 2017, vol. 24, 133-145. https://doi.org/10.1002/mcda.1611

Karen Vanessa Marín Gómez
Industrial Engineer - Universidad del Valle, Colombia. She has worked in projects related to humanitarian logistics, logistics and supply chains at industrial and commercial level, through mathematical models approaches. The areas of knowledge are: project management, logistics, production, quality management and operations research.ORCID: https://orcid.org/0009-0003-9155-418X

Jessica Dayan Morales Corredor
Industrial Engineer - Universidad del Valle, Colombia. Experience in process optimization, continuous improvement and management of operational and technological projects in sectors such as manufacturing, logistics and pharmaceuticals. Specialized in Lean Manufacturing, Six Sigma and digital transformation through Agile Project Management, she is currently responsible for the Progreso process with a focus on projects with impact on competitiveness, expansion and business sustainability .ORCID: https://orcid.org/0009-0000-3776-1557

Carlos Alberto Rojas Trejos
Industrial Engineer and Magister in Engineering of Universidad del Valle, Colombia. PhD in Engineering - Universidad Nacional de Colombia, Bogotá, Colombia. Actually, he is professor in Operation's Research, of academic program of Industrial Engineering in Universidad del Valle, Colombia. His research interests include applied mathematical modeling, humanitarian logistics, hospital logistics and optimization using metaheuristics. ORCID: https://orcid.org/0000-0002-5440-5631

Prediction model of financial suppliers in the vehicle manufacturing industry in Pereira.

Modelo de predicción de proveedores financieros en la industria manufacturera de vehículos en Pereira

D. C. Romero –Cárdenas ;A. O. Mejia ; L. A. Serna Cardona
DOI: https://doi.org/10.22517/23447214.25692
Scientific and technological research paper

***Abstract*— This research work presents the development of a model using data mining techniques to identify financial variables in a manufacturing company of automotive vehicle bodies in Pereira. The study is structured into four key phases. The first phase focuses on data preprocessing, including characterization, normalization, and dimensionality reduction using PCA, Relief, and Correlation. The second phase applies unsupervised learning with K-means and Gaussian Mixture Models (GMM) to cluster and validate data based on a defined target variable. In the third phase, supervised classifiers such as Bayesian Classifier, Artificial Neural Networks, Support Vector Machines, and KNN are employed to predict supplier efficiency, optimizing investment and costing processes. Finally, the fourth phase integrates preprocessing and prediction into a practical form, using libraries such as Plotly and Dash for detailed visualizations, and tools like GitHub and Heroku for application development. This study highlights the importance of artificial intelligence in business decision-making, demonstrating how data science techniques and visualization tools can facilitate the interpretation and utilization of data analysis results.**

***Index Terms* — Data Mining, Dash Plotly, Machine Learning, Neural Networks, Supplier Classification.**

***Resumen*— Este trabajo de investigación presenta el desarrollo de un modelo utilizando técnicas de minería de datos para identificar variables financieras en una empresa manufacturera de carrocerías para vehículos automotores en Pereira. El estudio se estructura en cuatro fases claves. La primera fase se centra en el preprocesamiento de datos, incluyendo la caracterización, normalización, y reducción de dimensionalidad mediante PCA, Relief y Correlación. La segunda fase aplica aprendizaje no supervisado con K-means y Mezclas Gaussianas (GMM) para agrupar y validar datos según una variable objetivo definida. En la tercera fase, se emplean clasificadores supervisados como el Clasificador Bayesiano, Redes Neuronales Artificiales, Máquinas de Soporte Vectorial y KNN para predecir la eficiencia de los proveedores, optimizando los procesos de inversión y costeo. Finalmente, la cuarta fase integra el preprocesamiento y la predicción en un formulario práctico, utilizando librerías como Plotly y Dash para visualizaciones detalladas, y herramientas como GitHub y Heroku para el desarrollo de la aplicación. Este estudio destaca la importancia de la inteligencia artificial en la toma de decisiones empresariales, demostrando cómo las técnicas de ciencia de datos y las herramientas de visualización pueden facilitar la interpretación y el aprovechamiento de los resultados del análisis de datos.**

***Palabras claves*— Aprendizaje de Máquinas, Clasificación de proveedores, Dash Plotly, Minería de datos, Redes Neuronales.**

I. INTRODUCTION

In today's digital era, organizations face a constant challenge: it is no longer enough to store large volumes of data—they must also be able to manage it intelligently and convert it into actionable knowledge. This capability has become essential to maintaining competitiveness and improving operational efficiency in an increasingly demanding business environment [1]. In this context, data science has emerged as a key discipline, offering advanced tools and techniques to analyze large-scale information, identify patterns, predict behaviors, optimize processes, and discover new strategic opportunities [2].

This study stems from a specific need identified in a vehicle body manufacturing company located in Pereira, Colombia. Despite its growth in the national market and its International projection, the company lacks advanced mechanisms to objectively and automatically evaluate the efficiency of its financial suppliers. Although it uses platforms such as Siesa, Power BI, and proprietary tools, its current analysis heavily depends on Excel macros and manual processes, limiting data integration, reducing analytical quality, and delaying strategic decision-making [3].

To address this issue, we propose the design of a predictive model based on data mining and machine learning techniques, capable of classifying suppliers into two categories: type 1 (efficient) and

This article was submitted for review on September 13, 2024. Accepted on March 25, and published on March 31, 2025. This work was supported by CIAF University under the project entitled “Prediction Model of Financial Suppliers in the Vehicle Manufacturing Industry in Pereira.” The study was conducted by Diana Carolina Romero Cárdenas and Alejandro Ospina Mejía in collaboration with Master's student Luis Ariosto Serna Cardona. For correspondence: karol272@gmail.com, alejandro880411@gmail.com, moog22002@hotmail.com

type 2 (less efficient). This model is built upon key information such as product identifiers, unit of measure, inventory type, batch type (domestic or imported), product line and subline, supplier and buyer codes, purchased quantities, unit values, local tax rates, and country of origin in the case of imports. Proper identification and analysis of these variables is essential to optimizing investments, reducing operational costs, and strengthening procurement management [4].

This leads to the following research question: Is it possible to design a predictive model, based on data mining and machine learning techniques, that can classify the financial efficiency of suppliers in a manufacturing company? To address this question, a four-phase methodology is proposed: (i) data preprocessing and dimensionality reduction using PCA, correlation analysis, and Relief-F [2][3][4]; (ii) clustering using unsupervised algorithms such as K-Means [5] and Gaussian Mixture Models (GMM)[6][12]; (iii) classification using supervised algorithms such as K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Artificial Neural Networks (ANN), and Bayesian Classifiers[13][14][15]; and (iv) development of an interactive application, designed with Dash and Plotly, that enables real-time visualization of results to support decision-making [16].

The proposed approach is supported by existing scientific literature. Studies such as those by Ralambondrainy (1995) [5] and Müller & Guido (2016) [13] have demonstrated the effectiveness of hybrid models—combining clustering and classification techniques—for segmenting economic actors and predicting performance. More recently, Zhang et al. (2021) [17] applied hybrid models for supplier selection in the manufacturing industry, while Kim and Lee (2022)[18] used clustering algorithms to classify suppliers in complex industrial environments. However, these models have been developed and validated in contexts that differ significantly from the Colombian setting, presenting a valuable opportunity to adapt them to emerging economies such as those in Latin America.

This paper is structured as follows: Section II presents the methodology; Section III reports the experimental results; Section IV discusses the findings in light of previous research; and Section V concludes the study and outlines future research directions.

II. II. METHODOLOGY

1. Data Preprocessing

The first phase focused on how to preprocess 350,732 sample data points taken between 2018 and 2023, segmented into 23 columns, which has become a problem for the vehicle manufacturing company when optimizing the investment and costing process. Therefore, this section indicated the ideal way to classify the information, then preprocess it and normalize it to 20 characteristic variables.

This section focused on the recognition of categorical data, identifying variables using conventional algorithms from different methods [7].

a. Collection and Normalization

As part of the preprocessing, it was necessary to normalize the database before applying the prediction algorithms. This was done to organize the data into logical groups, placing all variables on the same scale, thus allowing fair comparisons and minimizing data dispersion [2].
For this research and based on the state of the art supporting the current document, the normalization method used was StandardScaler, which generates normalized data to be processed with unsupervised and supervised learning algorithms.[8]

TABLE I.
SAMPLE OF DATA USED IN THE EXPERIMENT

Documento	*Ítem*	*Razón Social*	*Docto. orden*	*Docto. solicitud*	*Desc. ítem*
15921	9186	403	69	21763	1223
15922	9332	403	69	21763	4197
15926	5330	89	182	21763	2802

b. Dimensionality Reduction

Once the database was obtained, the first step was to perform categorical encoding of the information by classifying the data into nominal, ordinal, and numerical variables [9]. During data preprocessing, variable selection was conducted with the goal of identifying or determining which methods or techniques to apply, as conventional machine learning algorithms do not handle categorical variables efficiently. Therefore, it is necessary to convert them to quantitative data [10].

2. Unsupervised Learning

A clustering approach was developed using unsupervised learning with K-Means and Gaussian Mixture Models, in which performance metrics were used to validate the algorithm's accuracy level according to the target variable recommended by the expert [11].

Based on the state of the art reviewed in this research, the K-Means algorithm was used with ordinal variables because using nominal variables would considerably increase computational cost. Therefore, ordinal variables were used to divide the dataset into K distinct clusters, aiming to obtain two groups: efficient and less efficient suppliers [5].

A. K-Means Algorithm

The following details how the K-Means algorithm operates:

Initialization:
Select k initial points u1, u2, ..., uk called centroids.

Cluster Assignment:
For each data point xi, assign it to the cluster Cj whose centroid uj is closest.

$$C\_j = \{ x\_i : ||\, x\_i - \mu\_j \,||\hat{}2 \leq ||\, x\_i - \mu\_l \,||\hat{}2 \,, \forall l, , 1 \leq l \leq k \} \quad (1)$$

Centroid Update:
Recalculate the centroid of each cluster as the mean of the points assigned to the cluster.

$$\mu_j = \frac{1}{\{|C_j|\} \sum_{\{x_i \epsilon C_j\}} x_i} \quad (2)$$

Convergence Criterion:

Continue iterating between steps 2 and 3 until the centroids do not change significantly, i.e., until:

$$\left|\left|\mu_j^{t+1} - \mu_j^t\right|\right|^2 < \varepsilon \quad (3)$$

Where t is the iteration number and ϵ\epsilonϵ is a small positive value defining the tolerance for convergence.

a. B. Gaussian Mixture Model (GMM)

After generating the labels with K-Means using the 20 features and in agreement with the expert, it was decided to use the labels generated by this model as the Gold standard to verify the performance of another unsupervised algorithm, the Gaussian Mixture Model (GMM). GMMs are a probabilistic approach to represent the presence of subpopulations within an unlabeled dataset. Unlike the K-Means algorithm, which assigns each data point to a single cluster, GMMs consider that data points can belong to multiple clusters with different probabilities [12].

The following details how the Gaussian Mixture Model operates:

A Gaussian mixture is defined by:

1. K: The number of Gaussian components.

2. π_k**:** The weight of the k-th Gaussian component where

$$\sum_{k=1}^{K} \pi_k = 1 \quad (4)$$

3. u_k The mean vector of the k-th component.
4. $\sum k$: The covariance matrix of the k-th component.
The Gaussian mixture model is defined as:

$$p(x) = \sum_{k=1}^{K} \pi_k \, N\,(x \,|\mu_k, \sum\nolimits_k) \quad (5)$$

Where $N\left(x \mid u_k, \sum\nolimits_k\right)$ is the probability density function of the multivariate normal distribution.

3. Supervised Learning

In this phase, supervised learning algorithms were applied through classification tasks to validate their performance. This research utilized the Bayesian Classifier, the K-Nearest Neighbors (KNN) algorithm, Artificial Neural Networks (ANN), and Support Vector Machines (SVM). It should be noted that for supervised algorithms, classification is a very important approach for recognizing categorical data. However, there are methods for this purpose that focus on Kernels, which have accuracy problems and high computational costs [13]. For this reason, an approach to identifying ordinal categorical variables was proposed using conventional classifiers such as Bayesian Classification, ANN, KNN, and SVM, which not only improved accuracy levels but also offered lower computational costs.
Subsequently, the performance of the different proposed methods was evaluated, identifying that the three standard classifiers significantly improved the accuracy of the database without any procedure, making this applicability an ideal way to group and classify categorical data [14].

The following describes the supervised KNN algorithm, the most accurate one applied in this research.

The K-Nearest Neighbors algorithm is a supervised classification method that assigns a label to a new instance based on the labels of its k nearest neighbors in the feature space [15]. The learning process of the KNN method is based on calculating the distance

of the new element to each existing one and sorting these distances from smallest to largest to select the group to which it belongs. This group is the one with the highest frequency with the smallest distances [16].

The method operates as follows:

1. The training data X = x1, x2, … xN with labels y = y1; y2, … yN (N being the number of data samples) are stored in memory.
2. For a new sample xi e RD, where D is the number of attributes, the k nearest neighbors are found using a distance d in the entire training set (k can be 1, 3, 5, 7,…).
3. A procedure is carried out to select the class of the new sample xi.The common distances d are:

4. The common distances d are:

- **Mahalanobis Distance:**

$$D_{M(x,y)} = \sqrt{\{(x - y)^T \Sigma^{\{-1\}(x-y)}\}}$$

$$donde \sum^{-1} es\ la\ matriz\ de\ covarianza\ entre\ x\ y\ y.$$

- **Euclidean Distance:**

$$||x - y||_2 = \sqrt{\{(x - y)^T (x - y)\}}$$

- **Manhattan Distance:**

$$(Mathb(x,y) = \left| (x - y)^{\top(x-y)} \right| \quad (6)$$

III. III. Dataset and Experimental Setup

A. Dataset

The dataset used in this study was obtained from a vehicle manufacturing company in the city of Pereira. Table II presents the relevant information. It is important to clarify that the information was anonymized to maintain the confidentiality of the processed data. The data used in this project consists of 350,732 sample records collected between the years 2018 and 2023.

TABLE II.
SAMPLE OF THE ORIGINAL DATA USED FOR THIS EXPERIMENT

Tipo docto	Documento	Docto. referencia	Item	Razón Social	Sucursal
EA	EA-00187457	FACT-120731	9186	ALEJANDRO S.A.S	1
EA	EA-00000949	FACT-21	71630	PRIETO LUIS	1
EA	EA-00187463	FACT-41941	5330	CAROLINA COLOMBIA	1

A comma-separated flat file containing this information was used for subsequent processing in Python using the Spyder development environment. The data underwent preprocessing, dimensionality reduction, and feature extraction relevant to the experiment. Following this, unsupervised and supervised algorithms were applied, measuring their levels of accuracy and precision to determine the most accurate algorithm. This allows the company to predict the types of suppliers it has, classified as Rating 1 (Efficient Suppliers) and Rating 2 (Less Efficient Suppliers).

A. B. Specifications and Training

The identification and selection of the most relevant data sources on supplier purchases in the financial area of a manufacturing company in Pereira were carried out using the ERP system Siesa. The information was exported to a CSV file for analysis. The Extraction, Transformation, and Loading (ETL) process ensured the proper formatting of the data, including data cleaning.

Handling of missing and duplicate values and correction of descriptive errors.
A supplier classification module was developed, validating state-of-the-art methods such as k-means and Gaussian mixtures to generate grouping labels, which were evaluated using performance metrics like accuracy, precision, recall, and F1-Score. Additionally, a user interface with a form was designed to enter relevant data, demonstrating the applicability of the results through the use of the supervised KNN algorithm and Dash Plotly, with **70%** of the data for training and **30%** for validation. The combination of these methods and tools facilitated effective classification and clear visualization of the data, providing valuable information for financial decision-making.

Future work will focus on advanced predictive analysis and machine learning techniques to optimize the identification of critical financial variables and improve operational efficiency. Deep neural networks, real-time data processing, advanced interactive dashboards, and the expansion of the model to other business areas will be explored. A feedback loop system is also proposed for continuous improvement, explainable artificial intelligence techniques to ensure transparency, and fostering interdisciplinary collaborations, thus enhancing financial management and establishing a solid foundation for the application of AI and data analysis in the organization.

IV. RESULTS AND DISCUSSIONS

A. Dimensionality Reduction Algorithms

a. PCA Algorithm

The covariance matrix was used, utilizing the Eigenvalues and Eigenvectors of the function, reducing the dimensionality of the variables by activating the PCA algorithm, as shown in Fig. 1. Once the process was completed, the variance method was used, which allowed determining which variables provided the most information to the algorithm.

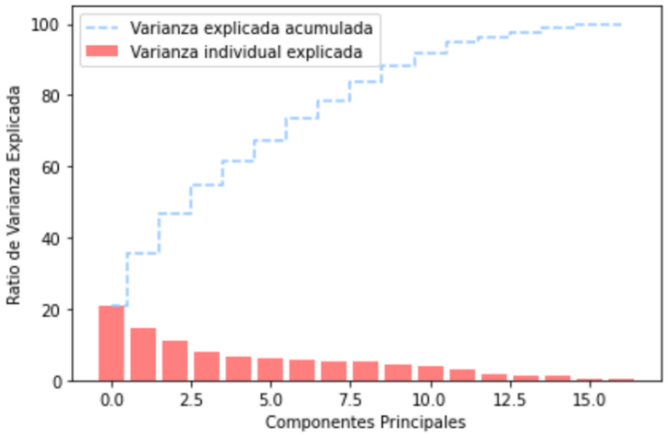


Fig. 1. Explained and Cumulative Variance

b. Correlation

As part of the dimensionality reduction process and with the objective of identifying the linear relationship and proportionality between statistical variables, correlation analysis was performed on the information obtained after applying principal component analysis in the previous step [3]. In Fig. 2, the 18 features are observed with respect to each other, where, for example, the total correlation between column 15 and column 1 can be seen, as well as between column 13 and column 3. The aim is to reduce dimensionality by removing the correlation that exists between the features.

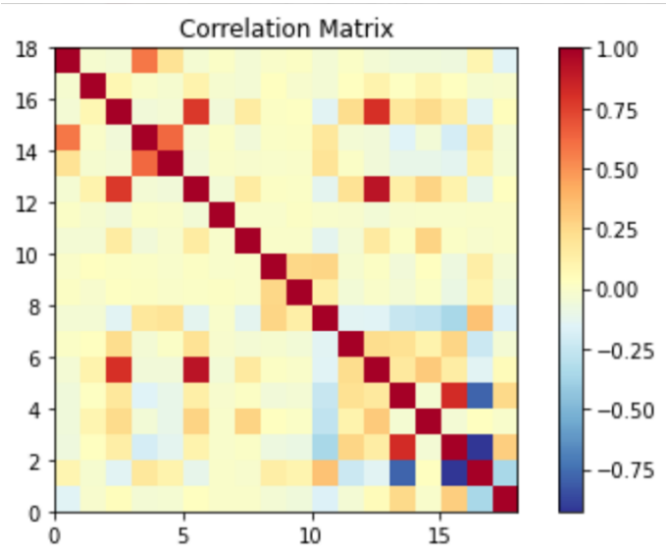


Fig. 2 Matrix with Full Correlation

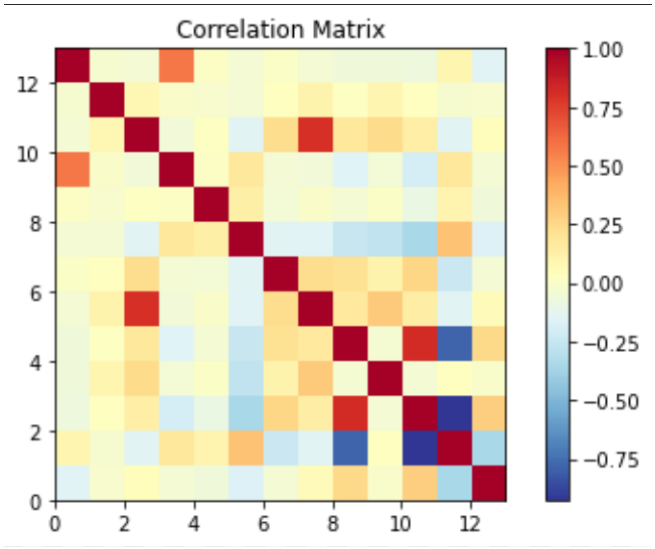


Fig. 3 Reduced Correlation Matrix with PCA

In this matrix, by eliminating 5 features, it becomes evident that the tone is lighter, indicating less correlation between the columns.

c. Relief F Algorithm

For the development of the Relief algorithm, the most recommended approach in the state of the art was used. In this section, the Relief algorithm estimated the relevance of the features based on how neighboring instances of different classes differ, setting the weight of each feature to 0, defining the number of nearest neighbors K and the number of iterations m. After executing all iterations, the features were selected based on their highest weights as the most relevant features [17].

In Fig. 4, the Eigenvalues are shown to determine which values are below the required threshold according to state-of-the-art recommendations. In this context, it is determined that the values below the threshold or negative values in the Eigenvalues are features that do not contribute value to the function.

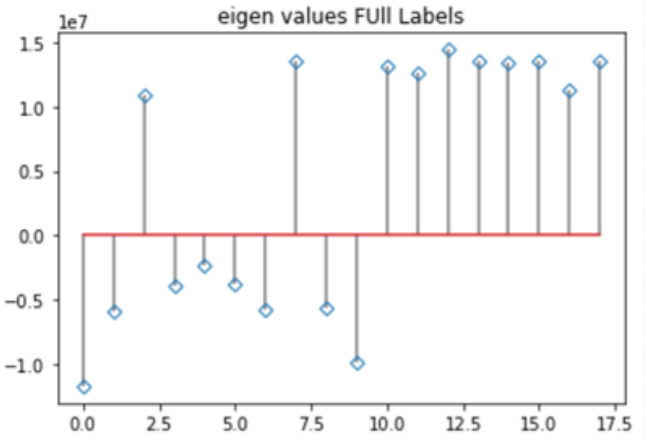


Fig. 4. Labels with Eigenvalues

After preprocessing the threshold and removing the less important features, Fig. 5 visualizes the most influential features in the database.

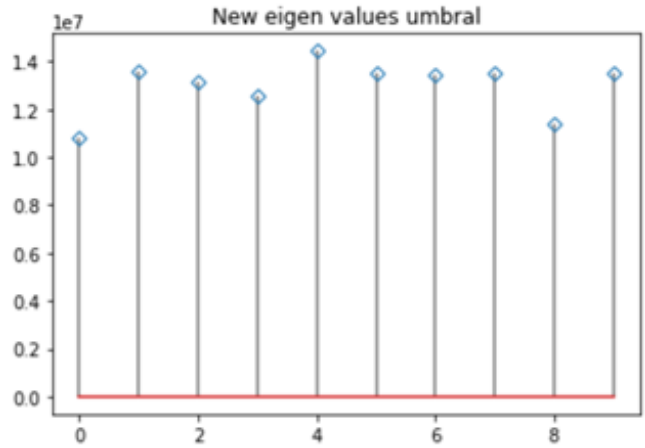


Fig. 5. Threshold with New Eigenvalues

For data preprocessing and dimensionality reduction, three techniques were used: PCA, Correlation, and Relief F, as recommended by the company expert. The results showed a reduction to 12 features with PCA, 13 features with Correlation, and 9 features with Relief F. Consequently, it was decided not to use any of the three techniques for dimensionality reduction and to use the complete database with all 20 features, as recommended by the expert.

Based on the state of the art reviewed in this research, the K-Means algorithm was used with ordinal variables because using nominal variables would considerably increase computational costs. Therefore, ordinal variables were used to divide the dataset into K distinct clusters, aiming to obtain two groups: efficient suppliers and less efficient suppliers [5].

The following presents the prediction data obtained by applying GMM to the test dataset with 20 features using the labels generated by K-Means. It is important to highlight that the labels generated by the GMM algorithm were duly reviewed by the company.

TABLE III.
PERFORMANCE OF THE GMM ALGORITHM WITH 20 FEATURES

Métricas	Valores
Matriz de Confusión	
TP (Prov. Clase 0)	43.110
FN (Prov. Clase 0)	0
FP (Prov. Clase 1)	10.582
TN (Prov. Clase 1)	297.041
Exactitud	97%
Clase 0	
Precisión	0.80
Recall	1.00
F1-Score	0.89
Clase 1	
Precisión	1.00
Recall	0.97
F1-Score	0.98

Based on this, an accuracy of **97%** was identified, indicating that it is a high-performing model. Subsequently, performance metrics were analyzed by class, revealing that class 0 had a **precision of 80%** and a **recall of 100%,** while class 1 achieved a precision of 100% and a recall of 97%. This led to the conclusion that both classes demonstrated satisfactory performance, with an **F1-score of 89%** for class 0 and **98%** for class 1.

B. Unsupervised Algorithms

a. K-Means Algorithm vs. Gaussian Mixture Model (GMM)

In test 2, the prediction data obtained by applying the GMM algorithm to the test dataset with the 13 features reduced by PCA using the labels generated by K-Means are presented. It is

important to highlight that the labels generated by the GMM algorithm were duly reviewed by the company.

TABLE IV.
PERFORMANCE OF THE GMM ALGORITHM WITH 13 FEATURES USING PCA

Métricas	Valores
Matriz de Confusión	
TP (Prov. Clase 0)	43.125
FN (Prov. Clase 0)	0
FP (Prov. Clase 1)	14.303
TN (Prov. Clase 1)	293.305
Exactitud	96%
Clase 0	
Precisión	0.75
Recall	1.00
F1-Score	0.86
Clase 1	
Precisión	1.00
Recall	0.95
F1-Score	0.98

Based on this, an accuracy of **96%** was identified, indicating that it is a high-performing model. Subsequently, performance metrics were analyzed by class, revealing that class 0 had a **precision of 75%** and a recall of 100%, while class 1 achieved a **precision of 100%** and a **recall of 95%.** This led to the conclusion that both classes demonstrated satisfactory performance, with an **F1-score of 86%** for class 0 and **98% for class 1.**

In test 3, the prediction data obtained by applying the GMM algorithm to the test dataset with the 9 features reduced by Relief F using the labels generated by K-Means are presented. It is important to highlight that the labels generated by the GMM algorithm were duly reviewed by the company.

The following presents the prediction data obtained by applying the GMM algorithm to the test dataset with the 9 features reduced by Relief F using the labels generated by K-Means. It is important to highlight that the labels generated by the GMM algorithm were duly reviewed by the company.

TABLE V.
PERFORMANCE OF THE GMM ALGORITHM WITH 9 FEATURES USING RELIEF F

Métricas	Valores
Matriz de Confusión	
TP (Prov. Clase 0)	307.181
FN (Prov. Clase 0)	441
FP (Prov. Clase 1)	33
TN (Prov. Clase 1)	43.078
Exactitud	100%
Clase 0	
Precisión	1.00
Recall	1.00
F1-Score	1.00
Clase 1	
Precisión	0.99
Recall	1.00
F1-Score	0.99

Within this framework, it was observed that the experiment of K-Means with Gaussian Mixtures, using Relief F, also achieved an accuracy of 100%. However, considering all the features, the algorithm achieved a high accuracy rate of 97%. In light of these results, the expert ultimately decided to retain all the labels generated by K-Means, recognizing their relevance in the context of the intended financial analysis.

C. Supervised Algorithms

Subsequently, the performance of SVM, ANN, Bayesian Classifier, and KNN was evaluated, where it was identified that all four standard classifiers significantly improved the accuracy of the database without any preprocessing, making this approach ideal for grouping and classifying categorical data [14].

a. SVM Algorithm

Below are the prediction results obtained by applying each supervised algorithm.

TABLE VI.
PERFORMANCE OF THE SVM CLASSIFICATION MODEL

Métricas	Valores

Matriz de Confusión	
TP (Prov. Clase 0)	8.834
FN (Prov. Clase 0)	5.301
FP (Prov. Clase 1)	26.699
TN (Prov. Clase 1)	29.313
Exactitud	54%
Clase 0	
Precisión	0.25
Recall	0.62
F1-Score	0.36
Clase 1	
Precisión	0.85
Recall	0.52
F1-Score	0.65

The confusion matrix in Table VI shows that out of 14.135 instances of class 0, 8.834 instances were correctly predicted and 5.301 were misclassified. For class 1, out of 56.012 instances, 29.313 were correctly predicted and 26.699 were misclassified, with an **accuracy of 54%.**

b. Bayesian Classification Algorithm

TABLE VII.

PERFORMANCE OF THE BAYESIAN CLASSIFICATION MODEL

Métricas	Valores
Matriz de Confusión	
TP (Prov. Clase 0)	9.270
FN (Prov. Clase 0)	4.865
FP (Prov. Clase 1)	27.503
TN (Prov. Clase 1)	28.509
Exactitud	82%
Clase 0	
Precisión	1.00
Recall	1.00
F1-Score	1.00
Clase 1	
Precisión	0.03
Recall	1.00
F1-Score	0.06

The confusion matrix in Table VII shows that out of 14.135 instances of class 0, 9.270 were correctly predicted and 4.865 were misclassified. For class 1, out of 56.012 instances, 28.509 were correctly predicted and 27.503 were misclassified, with an **accuracy** of **82%.**

c. Algoritmo ANN

TABLE VIII.

ANN MODEL PERFORMANCE

Métricas	Valores
Matriz de Confusión	
TP (Prov. Clase 0)	3.848
FN (Prov. Clase 0)	10.287
FP (Prov. Clase 1)	1.422
TN (Prov. Clase 1)	54.590
Exactitud	83%
Clase 0	
Precisión	0.73
Recall	0.27
F1-Score	0.40
Clase 1	
Precisión	0.84
Recall	0.97
F1-Score	0.90

The confusion matrix in Table VIII shows that out of 14.135 instances of class 0, 3.848 were correctly predicted and 1.422 were misclassified. For class 1, out of 56.012 instances, 54.590 were correctly predicted and 1.422 were misclassified, with an **accuracy** of **83%.**

d. KNN Algorithm

TABLE IX.

KNN MODEL PERFORMANCE

Métricas	Valores
Matriz de Confusión	
TP (Prov. Clase 0)	8.197
FN (Prov. Clase 0)	5.938
FP (Prov. Clase 1)	3.301
TN (Prov. Clase 1)	52.711
Exactitud	87%
Clase 0	
Precisión	0.71
Recall	0.58
F1-Score	0.64
Clase 1	
Precisión	0.90
Recall	0.94
F1-Score	0.92

The confusion matrix in Table IX shows that out of 14.135 instances of class 0, 8.197 were correctly predicted, and 5.938 were misclassified. For class 1, out of 56.102 instances, 52.711 were correctly predicted and 3.301 were misclassified, achieving **87% accuracy.**

Based on this, a **87% precision** was achieved, indicating that it is a high-performance model. Subsequently, performance metrics by class were analyzed, finding that both **precision** and **recall** were **71%** and **90%** for both classes, and an **F1-score** of **64% and 92%** was observed for both class 0 and class 1. This concluded that the supervised KNN algorithm is the most accurate among those studied in this research.

The results obtained in this research demonstrate the high effectiveness of the KNN model, achieving an accuracy of 87% and an F1-Score of 92% for efficient suppliers. This validates the model's ability to accurately classify categorical data in a business context.

These findings are consistent with studies such as that of Zhang et al. [17], who applied hybrid models for supplier selection in the manufacturing industry, reporting high levels of precision. Similarly, Kim and Lee [18] showed that clustering algorithms applied prior to supervised learning significantly improved segmentation performance in complex industrial environments.

Unlike those studies, which were developed in advanced organizational contexts in Asia, this research adapts the models to a Colombian setting using data extracted from a local ERP system and tailored to the conditions of a real manufacturing company in Pereira. This contextual difference explains why the KNN model proved more effective than SVM or ANN in this case, particularly due to its lower computational cost and its compatibility with visual tools such as Dash and Plotly [16].

In conclusion, this study offers a methodological contribution by demonstrating that classification models like KNN, when supported by robust feature reduction and interactive visualization processes, can yield results comparable to international research and remain applicable to emerging business environments.

V. CONCLUSIONS AND FUTURE WORK

This work presented an approach for characterizing and grouping categorical data using ordinal variables. The variable "Línea," considered key by the company expert, was used for dimensionality reduction and supplier classification. Preprocessing techniques such as Z-score scaling were applied, improving experimental results and providing structured data for supervised and unsupervised algorithms.

The clustering approach included K-Means to generate labels, showing great similarity with the target variable. The labels were compared with the Gaussian Mixture Models (GMM) algorithm, achieving **97% accuracy** with 20 features, **96%** with 13 features (PCA), and **100%** with 9 features. However, the expert preferred using the 20 features from K-Means due to their business relevance. Finally, an optimized KNN model with various kernels and features was constructed, successfully identifying and classifying suppliers with high accuracy. This model is useful for analysts and investors in financial decision-making for the vehicle manufacturing company in Pereira, demonstrating the effectiveness of the K-Means and GMM algorithms in improving data separability and reducing computational times.

Future work will focus on advanced predictive analysis and Machine Learning techniques to optimize the identification of critical financial variables and improve operational efficiency. Deep neural networks, real-time data processing, advanced interactive dashboards, and the expansion of the model to other business areas will be explored. A feedback loop system for continuous improvement, explainable artificial intelligence techniques to ensure transparency, and interdisciplinary collaborations are also proposed to enhance financial management and establish a solid foundation for the application of AI and data analysis in the organization.

The performance of both supervised and unsupervised algorithms in this study is inherently conditioned by the characteristics of the dataset, which originates from a single manufacturing company in Pereira. As such, the predictive models may not generalize directly to other organizations without conducting similar analyses to determine the most suitable algorithms for each specific context. While the current approach effectively classifies supplier types within the studied company, its applicability to other environments requires contextual adaptation and possible retraining of the models.

IV. VI. ACKNOWLEDGMENTS

V.

Thanks to the Master's program in Systems and Computing Engineering at the Universidad Tecnológica de Pereira, to the Director of the Institutional Corporation for Administration and Finance (CIAF), Luis Ariosto Serna, and to Ph.D. Julián David Echeverry.

VI. VII. REFERENCES

[1] Gonzalez Disla, R. R. (2013). Big data: El cambio en los paradigmas de la información. ResearchGate. https://www.researchgate.net/profile/Renato-GonzalezDisla/publication/311950584_BIG_DATA_El_Cambio_en_el_Paradigma_de_la_Informacion/links/58645e6208ae329d6203a9d5/BIG-DATA-El-Cambio-en-el-Paradigma-de-laInformacion.pdf

[2] Sablón, B., et al. (2019). Gestión de la información y toma de decisiones en organizaciones educativas. Revista de Ciencias Sociales, XXV(2), 120–130. https://doi.org/10.31876/rcs.v25i2.27341

[3] Amo Cubillo, A. (n.d.). Research data management. Universidad de Valladolid. Recuperado el 25 de agosto de 2024, de http://uvadoc.uva.es/handle/10324/31269

[4] Chien, C.-F., Chang, Y.-J., & Wang, W.-C. (2018). AI and big data analytics for wafer fab energy saving and chiller optimization to empower intelligent manufacturing. IEEE Xplore. https://ieeexplore.ieee.org/document/8374411

[5] Ralambondrainy, H. (1995). A conceptual version of the K-means algorithm. Pattern Recognition Letters, 16(11), 1147–1157. https://doi.org/10.1016/0167-8655(95)00075-R

[6] Liang, Y., Quan, D., Wang, F., Jia, X., Li, M., & Li, T. (2020). Financial big data analysis and early warning platform: A case study. IEEE Access, 8, 36515–36526. https://doi.org/10.1109/ACCESS.2020.2969039

[7] Romero, A. C., Sanabria, J. S. G., & Cuervo, M. C. (n.d.). Utilidad y funcionamiento de las bases de datos NoSQL. Facultad de Ingeniería,

21(33), 21–32. Recuperado el 25 de agosto de 2024, de https://www.redalyc.org/articulo.oa?id=413940772003
[8] Iturria Aguinaga, A. (n.d.). Reduction of false positives in online outlier detection over time series using ensemble learning. Recuperado el 25 de agosto de 2024, de https://hdl.handle.net/10481/82540
[9] Hasan, B. M. S., & Abdulazeez, A. M. (2021). A review of principal component analysis algorithm for dimensionality reduction. Journal of Soft Computing and Data Mining, 2(1), 1–10. https://doi.org/10.30880/jscdm.2021.02.01.003
[10] Macqueen, J. (n.d.). Some methods for classification and analysis of multivariate observations. Recuperado el 25 de agosto de 2024, de https://doi.org/
[11] Karatzoglou, A., Meyer, D., Wien, W., & Hornik, K. (n.d.). Journal of Statistical Software: Support vector machines in R. Recuperado el 25 de agosto de 2024, de http://www.jstatsoft.org/
[12] Rasmussen, C. E. (2004). Gaussian processes in machine learning. Lecture Notes in Computer Science, 3176, 63–71. https://doi.org/10.1007/978-3-540-28650-9_4
[13] Castrillón, O. D., Sarache, W., & Ruiz-Herrera, S. (2020). Prediction of academic performance using artificial intelligence techniques. Formación Universitaria, 13(1), 93– 102. https://doi.org/10.4067/S0718-50062020000100093
[14] Kira, K., & Rendell, L. A. (1992). A practical approach to feature selection. In Proceedings of the ninth international workshop on Machine learning (pp. 249-256). Morgan Kaufmann.
[15] Ahmad, A., & Dey, L. (2007). A method to compute distance between two categorical values of same attribute in unsupervised learning for categorical data set. Pattern Recognition Letters, 28(1), 110–118. https://doi.org/10.1016/J.PATREC.2006.06.006
[16] Moya, R. (n.d.). Selección del número óptimo de clusters. Recuperado el 25 de agosto de 2024, de https://jarroba.com/seleccion-del-numero-optimo-clusters/
[17] Fin de Máster, T., Tutor, Q., Pérez, J., Cotutor, Á. A., & Bernabeu, P. (n.d.). Métodos de clasificación con Python: Aplicaciones empresariales. Universitat Politècnica de València, Escuela Politécnica Superior de Alcoy. Recuperado el 25 de agosto de 2024, de https://riunet.upv.es/handle/10251/195236

Diana Carolina Romero Cárdenas: graduated in Systems Engineering in 2006 at the National Open and Distance University. His research interests include Data Science, Data Analysis, Machine Learning, and Big Data.
ORCID: https://orcid.org/0009-0006-0568-874X

Alejandro Ospina Mejía: received his degree in Systems Engineering in 2011. His research interests include Data Science, Data Analysis, Machine Learning, and Big Data. ORCID: https://orcid.org/0009-0009-6902-8910.

Luis Ariosto Serna Cardona received his undergraduate degree in physical engineering (2017) and his M.Sc. degree in engineering (2021) and PsD Student. He is director of research at CIAF educacation superior and researcher in the department of engineering. Research interests: machine learning and deep learning. ORCID: https://orcid.org/0000-0003-3985-4014

Aplicación de herramientas de automatización robótica de procesos (RPA) en procesos de pentesting para MiPyMEs

Application of robotic process automation (RPA) tools in pentesting processes for MSMEs

M. J. Yepes Díaz; G. E. Taborda Blandón

DOI: https://doi.org/10.22517/23447214.25743
Scientific and technological research paper

***Abstract*— Ethical hacking, also known as penetration testing (pentesting), is an essential practice for identifying vulnerabilities in Information Technology (IT) systems through controlled cyberattack simulations, thereby enhancing IT security. However, the traditional manual approach has limitations due to the exponential growth of technological assets and the increasing complexity of infrastructures. These challenges lead to significant time and resource consumption, as well as the need for specialized technical expertise.This paper examines the integration of Robotic Process Automation (RPA) into pentesting as a solution to streamline and optimize these processes. Through a comparative analysis of documented methodologies and existing RPA tools, we propose a specific tool designed to automate penetration testing within a controlled and secure environment. Experimental results demonstrate that the proposed tool is a viable solution for improving the efficiency, accessibility, and scalability of security audits, offering a practical and robust approach to cybersecurity for a broad range of stakeholders, including both organizations and individuals.**

***Index Terms*— Automated pentesting, Cybersecurity, Penetration testing, Pentesting methodologies, Robotic Process Automation (RPA).**

***Resumen*— El hacking ético, también conocido como pentesting, es una práctica clave para identificar vulnerabilidades en sistemas de Tecnologías de la Información (TI) mediante simulaciones controladas de ataques cibernéticos, lo que permite mejorar la seguridad informática. Sin embargo, el enfoque tradicional, que depende de intervenciones manuales, se enfrenta a limitaciones debido al aumento exponencial de activos tecnológicos y la complejidad de las infraestructuras, lo que implica un alto consumo de tiempo, recursos y la necesidad de experiencia técnica especializada. Este artículo explora la integración de la Automatización Robótica de Procesos (RPA) en el pentesting como una solución para optimizar estos procesos. A través de un análisis comparativo de metodologías documentadas y herramientas RPA disponibles, se propone una herramienta específica para automatizar el pentesting en un entorno controlado y seguro. Los resultados experimentales obtenidos indican que esta herramienta es una alternativa viable para mejorar la eficiencia, accesibilidad y escalabilidad de las auditorías de seguridad, lo que la convierte en una solución efectiva en el ámbito de la seguridad informática.**

***Palabras claves*— Automatización Robótica de Procesos (RPA), Ciberseguridad, Metodologías de pentesting, Pentesting automatizado, Pruebas de penetración.**

I. Introduccion

En el ámbito de la evaluación y pruebas de seguridad informática, el hacking ético, también conocido como pentesting, se refiere a la ejecución de pruebas controladas en sistemas de TI para identificar brechas de seguridad y generar informes que orienten a los administradores en la implementación de medidas preventivas. [1] Esta práctica emula ataques cibernéticos reales, permitiendo evaluar la resiliencia de una infraestructura tecnológica frente a potenciales amenazas [2].

Tradicionalmente, el pentesting ha sido un proceso manual que, debido a su complejidad, se limita a un número reducido de activos para garantizar su efectividad. Sin embargo, el crecimiento exponencial de las redes informáticas ha llevado a la necesidad de automatizar herramientas de pentesting para cubrir un mayor espectro de activos en menor tiempo. Hasta hace poco, las pruebas de penetración dependían exclusivamente de especialistas altamente capacitados con años de experiencia, pero esta limitación, combinada con la escasez de expertos y los altos costos de los procesos manuales, ha evidenciado la necesidad de soluciones más eficientes y accesibles [3] [4].

This manuscript was submitted on December 12, 2024. Accepted on March 11, 2025. A1nd published on March 31, 2025.

This article presents progress on the research project "Functional Prototype of a Computer Platform for Information Security Risk Management and Penetration Testing, Using Automation Technologies and Artificial Intelligence Techniques," developed by the Metropolitan Technological Institute (ITM) and the company Grupo Nex, and funded by the Colombian Ministry of Science, Technology and Innovation.

M. J. Yepes Díaz, Software Development Technology student at the Metropolitan Technological Institute. Cybersecurity Research Group (e-mail: mariayepes309613@correo.itm.edu.co).

G.E. Taborda Blandón, PhD in Complex Systems Thinking, MS in Information Security, MS in Computer Science. Researcher in the Automation, Electronics and Computational Sciences group. OTC Teacher at the Metropolitan Technological Institute (e-mail: gabrieltaborda@itm.edu.co).

Este trabajo presenta avances del proyecto de investigación "*Prototipo funcional de una plataforma informática para la gestión del riesgo de seguridad de la información y pentesting, utilizando tecnologías de automatización y técnicas de inteligencia artificial*", desarrollado por el Instituto Tecnológico Metropolitano de Medellín y la empresa Grupo Nex, que es financiado por el Ministerio de Ciencias, Tecnología e Innovación (MinCiencias. La parte que se centra este artículo está relacionada con la Automatización Robótica de Procesos (RPA) en pentesting y fue apoyada por Maria José Yepes como Joven Investigadora vinculada al proyecto.

En este contexto, se propone un estudio que analiza la integración de la automatización en las auditorías de seguridad informática, especialmente en un entorno donde es crucial fortalecer la ciberseguridad frente a amenazas cada vez más sofisticadas. Este enfoque busca emplear herramientas que optimicen procesos manuales, garanticen respuestas oportunas y hagan más accesible este conocimiento a una audiencia más amplia. El objetivo principal del artículo es explorar la incorporación de la Automatización Robótica de Procesos (RPA) en el pentesting, proponiendo una metodología para la ejecución autónoma de pruebas de seguridad. La propuesta incluye una herramienta de RPA capaz de integrar diversas herramientas utilizadas en diferentes fases del pentesting, recolectar resultados de manera eficiente y presentarlos de forma clara, con un enfoque práctico para su aplicación en pequeñas y medianas empresas (MiPyMEs).

El estudio planteó tres objetivos específicos: analizar metodologías de pentesting para identificar las más adecuadas y automatizables, evaluar herramientas de RPA para seleccionar la más idónea, y validar la eficiencia del modelo de automatización mediante pruebas experimentales en un entorno controlado. Este enfoque se encuentra plasmado en el siguiente trabajo.

II. Estado del arte

La metodología aplicada para la construcción del estado del arte se encuentra incluida dentro del anexo 1 "*Metodología aplicada para el estado del arte*".

A. Pentesting

El pentesting, o prueba de penetración, es un proceso mediante el cual se simulan ataques cibernéticos reales sobre una infraestructura tecnológica para evaluar la solidez de su seguridad y obtener información que podría comprometer el funcionamiento correcto del sistema. Este proceso se basa en un conjunto de métodos, técnicas y estrategias diseñadas para evaluar la robustez de un sistema o red, con el fin de identificar y corregir posibles fallos o vulnerabilidades, poniéndose en el lugar de un atacante que podría intervenir en estos sistemas [5].

En la actualidad, el pentesting ha ganado terreno debido a la creciente relevancia de la seguridad de la información, tanto para empresas como para individuos. Como resultado, muchas organizaciones contratan expertos en ciberseguridad con el propósito de realizar pruebas que evalúan la defensa y mecanismos de seguridad de su infraestructura tecnológica y prevengan posibles ataques que son simulados en estas pruebas, mediante la corrección de los errores detectados y el fortalecimiento de sus sistemas [6].

En este sentido, los resultados retornados por las pruebas de pentesting proporcionan, además información relacionada al cumplimiento de las políticas de seguridad y recuperación ante desastres que son definidas por la organización ante estas situaciones, los mecanismos de mitigación de las brechas de seguridad e incluso la conciencia de seguridad de los empleados de esta [5].

Fig. 1. Fases de un ataque informático (Elaboración Propia)

En Fases de un ataque a un Sistema Informático [7] fig. 1 se revela la complejidad de los ataques cibernéticos, que, como señalan [2], buscan aprovechar las vulnerabilidades de los sistemas informáticos para causar un impacto negativo, incluso tomando el control completo de un sistema. Estos ataques representan una seria amenaza para la integridad y seguridad de los sistemas, y se manifiestan a través de distintas fases, desde el reconocimiento inicial hasta la eliminación de huellas para evitar ser detectados como se describe a continuación.

B. Metodologías de Pentesting

De acuerdo con la estructura de un ataque informático descrita anteriormente, se parte de metodologías y frameworks

de aplicación de Pentesting con métodos específicos y enfoques que pueden variar de acuerdo con las necesidades que se presentan para organizaciones y personas. A continuación, en la tabla I, se detallan algunas de las metodologías más conocidas:

TABLA I
METODOLOGÍAS DE PENTESTING

Metodología	Descripción	Fuente
PTES	Proporciona pasos para auditorías: reconocimiento, análisis de vulnerabilidades, explotación, explotación posterior y elaboración de informes detallados.	[1] [6]
OSSTMM	Ofrece una metodología revisada por pares para pruebas de penetración y auditorías. Cubre la seguridad informática, de procesos, tecnologías, comunicaciones e inalámbrica.	[2] [8] [9] [10]
OWASP	Metodología de pruebas de seguridad en aplicaciones web. Define fases para pruebas a lo largo del ciclo de vida del software: diseño, desarrollo, despliegue y producción.	[11]
ISAAF	Framework de pentesting que divide el proceso en planificación, evaluación y reporte. Se enfoca en requisitos de evaluación de seguridad y gestión de riesgos.	[2] [12]
NIST SP 800-115	Guía técnica para pruebas de seguridad, define fases como planificación, descubrimiento, ataque, notificación de resultados y sugerencias de mejora.	[13]

C. Comparación de Metodologías de Pentesting

En el marco de referencia ISO/IEC 250:10:2013, se definen algunos parámetros claves para definir la calidad de un modelo como son la extensibilidad, mantenibilidad, cobertura de dominio, usabilidad, disponibilidad y confiabilidad. Estos indicadores se utilizarán para medir la metodología más sólida en cuanto a la aplicación de automatización de procesos y a partir de la valoración dada, se seleccionará la mejor valorada según la literatura, que tenga una estructura sólida para el proceso de pentesting y existencia de medidas de contención claras y definidas que permitan la automatización del proceso de pentesting, definiendo los criterios a utilizar para ser medidos de la siguiente manera:

1) *Extensibilidad*
Evalúa si la metodología es clara en sus directrices para extender o personalizar los procedimientos de esta en diferentes escenarios y casos de estudio usando herramientas o técnicas variables.

2) *Mantenibilidad*
Verifica si la documentación de la metodología está bien organizada y es fácil de entender, la frecuencia con la que se actualiza la metodología para reflejar nuevas amenazas y tecnologías y si está dividida en módulos o fases que pueden ser modificados de manera independiente.

3) *Cobertura*
La cobertura revisa la amplitud de áreas de seguridad que la metodología cubre y su adaptación a diferentes industrias y entornos.

4) *Usabilidad*
Evalúa si la metodología contiene materiales de formación y soporte adecuados y qué tipo de errores o problemas encuentran los usuarios al aplicar la metodología

5) *Disponibilidad*
Este criterio verifica si la metodología está fácilmente disponible para los profesionales, evalúa la disponibilidad de herramientas, guías y plantillas que la complementen.

En la Tabla II se sintetiza un comparativo de las metodologías de pentesting con basa a diversos autores.

D. Herramientas de Pentesting

De acuerdo con lo anteriormente descrito sobre pentesting, los tipos existentes y las fases que lo componen, es importante destacar algunas herramientas que en cada fase del pentesting, desde el reconocimiento hasta la explotación, ofrecen utilidades específicas que maximizan la efectividad del proceso.

En la tabla III se hará la descripción de algunas de las herramientas más utilizadas para el proceso de pentesting debido a su reconocimiento, soporte de la comunidad, solidez en el mercado y funcionalidades.

E. RPA

La automatización de procesos ha sido definida como la ejecución por parte de una máquina o un agente virtual de una función que anteriormente fue realizada por un humano, de esta manera se permite la gestión automatizada parcial o total de actividades que son manuales y basadas en reglas [14] [15].

RPA permite pensar sobre robots que ejecutan tareas humanas donde estos, en realidad, corresponden a una solución de software que puede significar una extrapolación de las tareas de un humano ejecutando tareas estructuradas y repetitivas [16].

TABLA II
COMPARACIÓN DE METODOLOGÍAS DE PENTESTING

Factor	PTES	OSSTM	OWASP	ISSAF	NIST SP 800-115
(1)	Reconocida y utilizada ampliamente Potencial para desarrollarse en un framework.	Alta, con resultados medibles y verificables en varios ámbitos de seguridad.	Muy extensible con herramientas y recursos integrables en diversos entornos de desarrollo web	Marco detallado y útil para evaluaciones de seguridad en varias áreas tecnológicas.	Flexible y adaptable a diversos entornos y necesidades organizacionales
(2)	Documentación accesible y estructurada. Última actualización en 2016.	Actualizada por una comunidad activa desde 2010, con un enfoque científico y métrico.	Revisiones y actualizaciones frecuentes por una comunidad global. Última versión en 2020.	Sin actualizaciones activas desde 2006.	Publicada en 2008, última revisión en 2021. Mantenida por NIST, asegurando autoridad y reconocimiento
(3)	Define el alcance en la interacción previa con el cliente. No tan amplia como OSSTMM, ISSAF u OWASP.	Amplia, cubriendo seguridad física, telecomunicaciones, redes, interacciones humanas y aplicaciones.	Amplia variedad de pruebas de seguridad en aplicaciones web. Cubre todo el ciclo de desarrollo de software.	Cobertura integral desde seguridad de infraestructura hasta aplicaciones y evaluación de riesgos.	Amplia cobertura, incluyendo pruebas de penetración, revisiones de configuración y controles de seguridad operacionales.
(4)	Estructura clara con guías técnicas detalladas para cada fase.	Recursos detallados y ejemplos prácticos. Complejidad media y fácil implementación en procesos de reestructuración de seguridad informática	Documentación detallada, guías y herramientas accesibles y fáciles de usar, incluso para principiantes.	Directrices claras y detalladas. Ejemplos prácticos y estudios de caso para mejorar la implementación.	Procedimientos claramente definidos, accesibles y fáciles de seguir. Documentación disponible públicamente.
(5)	Recursos en línea variados, pero sin historial de cambios detallado.	Documentación y recursos disponibles a través de ISECOM y canales autorizados	Herramientas, documentos y foros gratuitos y abiertos a cualquier persona interesada en seguridad de aplicaciones.	Documentación detallada públicamente accesible, pero sin un repositorio oficial fácilmente reconocible	Disponible gratuitamente en el sitio web de NIST, junto con documentación complementaria.

TABLA III
HERRAMIENTAS DE PENTESTING

Herramienta	Fase de Pentesting	Descripción	Fuente
The Harvester	Reconocimiento pasivo	Recolecta información de internet sobre el objetivo, como correos, subdominios y metadatos.	[17]
Nmap	Reconocimiento activo	Escanea puertos y revela información de sistemas operativos, servicios y versiones utilizadas.	[18]
Shodan	Reconocimiento activo	Permite identificar configuraciones de red, puertos expuestos y geolocalización de dispositivos.	[11]
OWASP ZAP	Análisis de vulnerabilidades	Detecta brechas de seguridad mediante escaneo pasivo y ataques de fuerza bruta.	[17]
Metasploit	Explotación	Framework avanzado para pruebas de explotación, análisis de código y ataques de fuerza bruta.	[19]

De esta manera RPA define una manera simple para crear, desplegar y manejar sistemas de software robóticos que imitan los movimientos humanos mientras manipulan la información generada por otros sistemas digitales. Estos robots como las personas pueden comprender lo que se muestra en pantalla, que tipo de teclas deben ser presionadas, a que sistema deben moverse y cómo localizar y extraer información para ejecutar una variedad de otras tareas [20].

La Automatización Robótica de Procesos (RPA) ofrece múltiples ventajas para las empresas, destacando su facilidad de configuración, permitiendo su implementación incluso por desarrolladores sin experiencia en programación. Además, es una tecnología no invasiva que se adapta a los sistemas existentes sin necesidad de costosas plataformas nuevas [16].

RPA incrementa la productividad y precisión, reduciendo errores y perfeccionando procedimientos, lo que mejora la calidad y la satisfacción en las industrias que la adoptan. También alivia la carga laboral del personal, mejora la consistencia de los datos y fortalece la competitividad.

Con un costo inicial bajo y un retorno de inversión alto, RPA fomenta la eficiencia y permite generar ahorros significativos, facilitando iniciativas tácticas [21]. Según [20], combinar RPA con la fuerza laboral humana puede reducir los costos operativos entre un 30% y un 50% en actividades transaccionales.

Según [22], los criterios típicos para los procesos adecuados para RPA son:

1) Requisitos cognitivos bajos

Es difícil que los procesos complejos que contienen una alta cantidad de tareas puedan ser manejados por RPA.

2) Poco acceso a múltiples sistemas

RPA se aplica sobre las aplicaciones existentes o cuenta con integración a unas herramientas específicas y puntuales.

3) Procesos realizados con frecuencia

Los procesos que tienen una alta frecuencia de ejecución o repetición son buenos candidatos para la implementación de RPA.

4) Procesos con alta probabilidad de error humano

Los procesos que requieren un alto nivel de detalle y presentan un elevado riesgo de errores debido a su complejidad, amplitud o manejo de grandes volúmenes de información deberían priorizarse para la implementación de RPA.

F. Herramientas de RPA

RPA es considerada como la tecnología que "automatiza la automatización", debido a su enorme potencial para automatizar procesos. Dicha automatización se realiza mediante agentes de software denominados robots, los cuales se encargan de realizar la ejecución de las tareas; por ejemplo, establecer comunicación entre las interfaces gráficas de usuario de dos aplicaciones de manera autónoma sin la intervención humana. Entre estas herramientas podríamos mencionar algunas como lo son: *Uipath, Automation Anywhere, Blue Prism, WorkFusion.* [17].

Las herramientas RPA son básicamente programas de software que operan sobre la interfaz gráfica de otros sistemas informáticos como si fueran humanos. Hay muchas herramientas disponibles en el mercado que se utilizan para desarrollar RPA como UiPath, Automation Anywhere, Blue Prism, etc [20].

La elección de la herramienta de Robotic Process Automation (RPA) adecuada es crucial para el éxito de la automatización de procesos en una organización. Una selección incorrecta puede resultar en un retorno de inversión (ROI) deficiente, ya que la herramienta elegida podría no ser capaz de manejar las especificidades de los procesos que se desean automatizar. [23] Algunas de las herramientas más destacadas se presentan en la tabla IV.

TABLA IV
HERRAMIENTAS DE RPA

Herramienta	Descripción	Fuente
UiPath	Herramienta basada en C# y vb.net para construir y desplegar robots de software. Ofrece funcionalidades de RPA con IA como reconocimiento de imágenes y minería de texto. Interfaz intuitiva y fácil de usar.	[20] [24] [25] [26] [27]
Automation Anywhere	Plataforma enfocada en escalabilidad y seguridad. Proporciona controles centralizados, integración con ERP, capacidades de IA, y opciones para automatización compleja.	[25] [26] [27] [28]
Blue Prism	Basada en .NET, con arquitectura cliente-servidor. Enfocada en gobernanza y seguridad, ideal para empresas con datos sensibles. Permite modelar y diagramar procesos de automatización con énfasis en eficiencia y seguridad.	[23] [26] [27] [29]
Robocorp	Herramienta open-source basada en Python (Robot Framework). Versátil y asequible, permite crear robots personalizados y utilizar editores como Visual Studio. Ideal para organizaciones que buscan adaptabilidad y reducción de costos.	[24] [30] [31]

G. Comparación de Herramientas de RPA

Para la elaboración del análisis comparativo se definen los siguientes criterios de comparación

1) Arquitectura

Este criterio se refiere a la estructura técnica de la herramienta, ya sea basada en cliente-servidor o por medio de un orquestador en la web, esto permite definir de manera clara el proceso de implementación y despliegue de RPA para pentesting.

2) Integración

La integración evalúa la capacidad de la herramienta para interactuar con otras aplicaciones y sistemas existentes. Una buena integración es crucial para la automatización fluida de procesos que involucran múltiples plataformas como es el caso de las pruebas de pentesting.

3) Procesos que pueden ser automatizados

Indica si la herramienta es adecuada para la automatización de procesos de oficina, atención al cliente o procesos más complejos en diferentes sectores.

4) Precio

Se considera el costo de la licencia del software o si es de código abierto, lo que es un factor determinante para la ejecución de pruebas en entornos que tengan herramientas y funcionalidades avanzadas.

5) Soporte y Comunidad

Se evalúa la calidad del soporte que se ofrece con la herramienta y la existencia de una comunidad activa que proporcione documentación y de razón de la mantenibilidad de la herramienta.

6) Popularidad

Este criterio considera la popularidad de la herramienta a través del análisis de búsquedas y de herramientas que proporcionan distintas métricas con el fin de determinar su grado de aceptación y reconocimiento por parte del público objetivo.

A continuación, se presenta una comparación de las herramientas UiPath, Automation Anywhere, Blue Prism y Robocorp, en relación con estos criterios definidos:

UiPath

- *Arquitectura:* Basada en un orquestador web con componentes como UiPath Studio, Orchestrator y Robots, para administrar robots en la nube o localmente [23] [29].
- *Integración:* Compatible con herramientas de ofimática (Word, Excel, correo), plataformas como AWS, Oracle, Microsoft y SAP, facilitando la interacción con objetos en pantalla [26] [32].
- *Procesos que automatiza:* Inicialmente, desarrolló bibliotecas para IBM, Google y Microsoft, enfocándose en sectores como BFSI, salud, telecomunicaciones y retail [23] [29].
- *Precio*: Ofrece una edición gratuita (Community) con limitaciones para distribución de bots y una empresarial con costos iniciales de $3,000–$5,000 USD, con una prueba gratuita de 60 días [26] [28].
- *Soporte y Comunidad:* Más de 1.5 millones de descargas, 750,000 desarrolladores y 250 socios tecnológicos respaldan la herramienta [28].
- *Popularidad*: Tendencia de crecimiento estable, con alta popularidad en Norteamérica, Europa y Asia [34].

Automation Anywhere

- *Arquitectura*: Cliente-servidor con tres componentes principales: Bot Creator, Control Room y Bot Runner, que permiten medir el rendimiento del robot [23] [29].
- *Integración*: Compatible con Google, Salesforce, SAP, Azure y otras APIs. [26]
- *Procesos que automatiza:* Utilizada en procesos de sectores como BFSI, salud, manufactura y telecomunicaciones (p.e.j., General Motors, JP Morgan Chase) [23] [29].
- *Precio*: Cuenta con una versión Community gratuita y opciones empresariales desde $9,000 USD anuales. No incluye pruebas dedicadas [24] [26].
- *Soporte y Comunidad:* Academia en línea, eventos, webinars y soporte especializado [33].
- Popularidad: Posición destacada en Norteamérica, Asia y Latinoamérica [34].

Blue Prism

- *Arquitectura*: Cliente-servidor, con componentes para diagramado, modelado y despliegue en la nube, local o híbrido [23] [29] [33].
- *Integración*: Compatible con Microsoft Power Platform, Azure, Salesforce, SAP y Google, aunque con mayor complejidad en la identificación de objetos [26] [35].
- *Procesos a Automatizar:* Sectores regulados como BFSI, telecomunicaciones y manufactura, con enfoque en gobernanza y cumplimiento normativo [27].
- *Precio:* Learning Edition gratuita por 180 días, luego una prueba de 30 días. Costos anuales de $15,000 USD [26] [35].
- *Soporte y Comunidad:* Foro Digital Exchange (DX) para compartir herramientas y certificaciones en línea [33] [36].
- *Popularidad*: Base de usuarios fieles en Europa y Norteamérica, especialmente en sectores regulados [34].

Robocorp

- *Arquitectura*: Nativa en la nube, soporta despliegue en instalaciones propias, nube o entornos híbridos [25] [36].
- *Integración*: Compatible con SAP, Salesforce, HubSpot, navegadores, APIs, herramientas ofimáticas y funciones de IA [36].
- *Procesos a Automatizar:* Flexible para sectores financieros, consultoría y salud, con casos como Grant Thornton LLP [36].
- *Precio:* Open-source y gratuito para proyectos individuales, con opciones empresariales para orquestación en la nube [24] [36].
- *Soporte y Comunidad:* Comunidad activa, documentación extensa y soporte para pequeñas y medianas empresas [33].
- *Popularidad:* Crecimiento sostenido en Norteamérica y Europa [24] [34].

En función de los criterios anteriormente mencionados y la descripción de las herramientas, se muestra a continuación en la fig. 2 la valoración de las características:

Herramienta	Arquitectura	Integración	Procesos	Precio	Comunidad / Soporte	Popularidad
UiPath	5	5	5	3	5	5
Automation Anywhere	4	4	4	3	4	4
Blue Prism	4	3	4	2	4	3
Robocorp	5	5	5	4	4	3

Fig. 2. Valoración por medio de criterios de Comparación de herramientas de RPA (Elaboración Propia)

En la fig. 3 se presenta la gráfica de búsqueda en Google de la diferente herramienta de RPA, que apoya el criterio de popularidad.

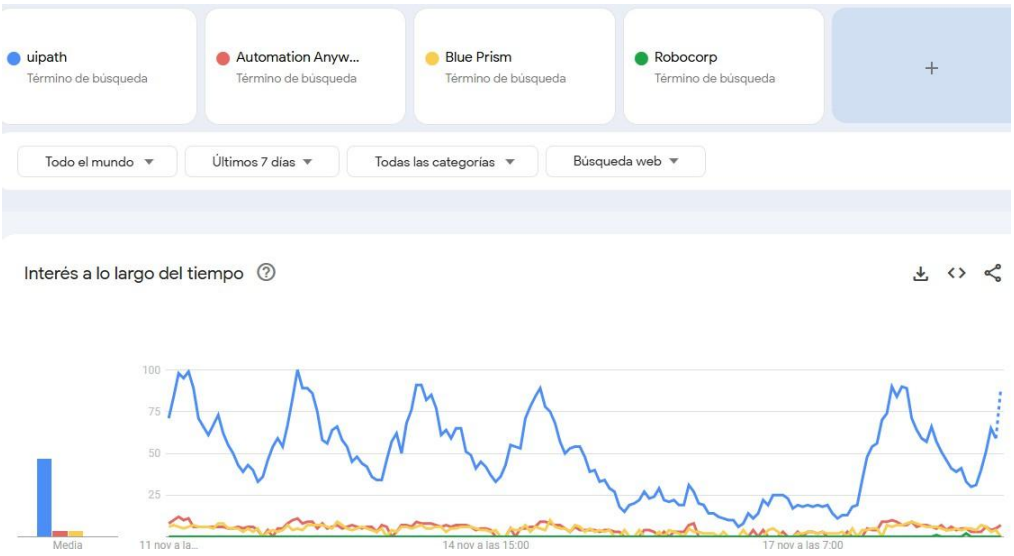


Fig. 3. Tendencias de búsqueda de herramientas de RPA por Google Trends [33]

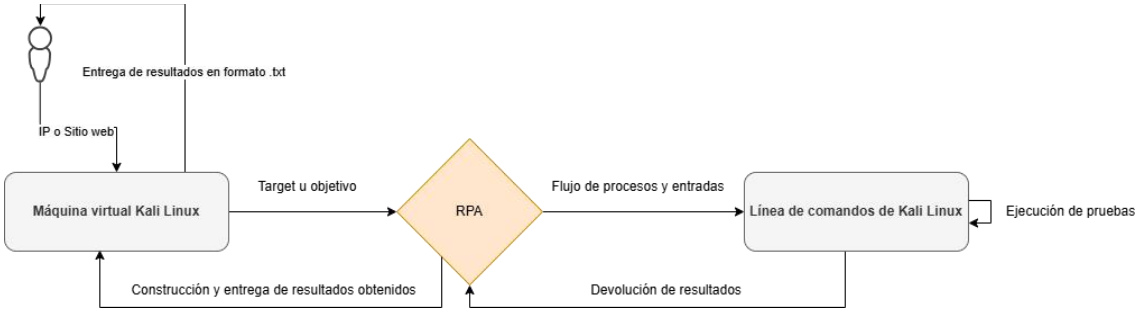


Fig. 4. Modelo conceptual del funcionamiento de RPA en ambiente controlado de pruebas (Elaboración Propia)

La necesidad de construir un prototipo de RPA que permita la automatización de tareas de pentesting de manera eficiente, con capacidad de escalabilidad y de tipo open-source, se decide que la herramienta más adecuada es Robocorp.

Robocorp ofrece una arquitectura nativa de la nube que permite un despliegue versátil y eficiente, ya sea en la nube, en las instalaciones o con configuraciones híbridas; además de que el modelo de código abierto de la herramienta reduce significativamente los costos asociados al desarrollo y despliegue de robots.

El modelo de código abierto de Robocorp permite escalar las operaciones de manera más eficiente, con una implementación personalizada que integra herramientas de pentesting, en contraste con otras ofertas donde las operaciones bajo esquemas de licencia pueden resultar prohibitivas.

H. Pentesting Automatizado con RPA

En el trabajo de Matzenberger, [31] Robocorp se usa junto a Process Mining para identificar procesos repetitivos en organizaciones y desarrollar robots que los ejecuten eficientemente. Este enfoque asegura que los robots optimicen el flujo de trabajo mediante un ciclo de análisis, pruebas y monitoreo continuo, maximizando la eficiencia en la operación de tareas estructuradas.

El artículo de Delilovic [30] expone cómo Robocorp puede integrarse de manera segura con Amazon AWS, utilizando el AWS Security Token Service (STS) para la autenticación con credenciales temporales. Este tipo de integración permite automatizar tareas en la nube sin comprometer la seguridad, al cumplir con requisitos de autenticación multifactor y acceso controlado. Esta implementación ofrece un modelo extensible para cualquier entorno de nube que demande altos estándares de seguridad.

En [37] Robocorp ha demostrado ser una herramienta efectiva para automatizar pruebas de seguridad. Correia presenta una implementación donde Robocorp se integra con el escáner de vulnerabilidades ZAP (Zed Attack Proxy), utilizando el lenguaje ASL4RPA para definir y ejecutar pruebas de seguridad en aplicaciones web. Esta configuración permite simular ataques de inyección SQL y cross-site scripting (XSS), documentando los resultados en tiempo real.

Esta integración optimiza el proceso de auditoría al reducir en un 75-85% procesos manuales, permitiendo que los

especialistas en seguridad se enfoquen en el análisis de vulnerabilidades en lugar de la configuración de herramientas. La combinación de Robocorp con ZAP no solo incrementa la eficiencia en pruebas de seguridad, sino que facilita la adopción de RPA como parte de las estrategias de ciberseguridad, permitiendo auditorías no asistidas o parcialmente asistidas.

III. METODOLOGIA

Robocorp, como herramienta de automatización robótica de procesos (RPA), ha sido ampliamente adoptada en distintas industrias para optimizar tareas repetitivas y complejas. Su flexibilidad le permite adaptarse tanto a la automatización de procesos convencionales como a aplicaciones avanzadas en entornos seguros.

Con el propósito de la construcción de una herramienta de automatización de pentesting por medio de automatización robótica de procesos, se propone el marco de la herramienta Robocorp y los procesos descritos por la metodología PTES dada la secuencialidad y síntesis de esta.

Para respaldar y dar un acercamiento a la herramienta con necesidades reales, se construye una encuesta la cual tiene por objetivo, evaluar dentro de un rango de expertos en ciberseguridad las herramientas con más potencial de automatización en el marco de RPA y que procesos de pentesting serían eficientemente automatizados por herramientas de RPA. Los resultados detallados de la encuesta se encuentran en el anexo 2 "*Tablero de Power BI con resultados de la encuesta*" De acuerdo con esta encuesta, fue posible recolectar la siguiente información:

- El nivel de experiencia de área de los encuestados se sitúa en un nivel Avanzado e Intermedio comprendiendo el uso de técnicas de pentesting o la familiaridad con procesos de seguridad informática en un período desde los 2 años a más de 5 años, esto con un porcentaje del 38% respectivamente para cada uno, y de un 24% de principiantes para conocimiento inferior a 2 años.

- De estos expertos, la segmentación de áreas donde trabajan se ubica dentro del área de la gestión de vulnerabilidades en un 20%, la gestión de riesgos de ciberseguridad, pruebas de pentesting y auditorías de seguridad en una mayor medida con un porcentaje de 17,78% cada una. También se detallan algunas otras áreas de actividad de las personas encuestados a continuación.

- Se evidencia que el 72,73% de los encuestados no ha utilizado herramientas de pentesting automatizado para ejecutar pruebas de seguridad en contraste con un 27,2% que afirma haber utilizado estas herramientas, de esta manera es posible inferir que no es muy conocido el uso de automatización dentro de los procesos de pentesting.
- Para los encuestados que han utilizado estas herramientas, se destaca la funcionalidad de automatización de escaneos, en la que sobresalen Nessus, OpenVAS y Tennable, cada una con un 20% de preferencia. Estas herramientas permiten realizar análisis de vulnerabilidades sobre direcciones IP o dominios, eliminando la necesidad de intervenciones manuales para obtener información de seguridad.

- Otras herramientas mencionadas en las respuestas incluyen: Burp Suite que permite automatizar pruebas de penetración y detectar vulnerabilidades, Checkmarx que realiza análisis estático de código (SAST) de forma automatizada, Immuni Web para facilitar escaneos automáticos en aplicaciones web, móviles y API y Nikto para escanear servidores web en busca de configuraciones inseguras.

- Se logra evidenciar además una tendencia dentro del uso de herramientas de pentesting, destacándose como herramientas principales Nmap (12,57%), Nessus (11,98%); Wireshark (10,78%), OWASP ZAP (9,58%) y Metasploit (8,98%).

- Los encuestados proporcionan su visión de algunos de los procesos con potencial de automatización, donde se destacan principalmente el escaneo de puertos y servicios (12,27%), el escaneo de vulnerabilidades (11,66%), la generación de informes (10,43%) y las pruebas de fuerza bruta (9,82%) pero también se presentan otras opciones con potencial de automatización como las pruebas de inyección de código, inyección SQL y de DoS y DDoS.

- De esta manera y en concordancia con lo anterior, se solicita especificar qué herramientas cuentan con un mayor potencial de automatización donde se destaca Nmap (20,22%) como una de las herramientas más aptas para automatización, seguida de Nessus (16,85%), Metasploit (14,61%), OWASP ZAP (8,99%) y BurpSuite (7,87%).

Gracias a la encuesta realizada fue posible determinar algunas perspectivas claves de cómo algunas herramientas existentes proporcionan una funcionalidad de automatización como es el caso de Nessus, OpenVAS, Burp Suite, donde se es posible por medio de una entrada de información de un target específico ejecutar una serie de pruebas.

También se identifica una tendencia de automatización de procesos de pentesting en torno a procesos que involucran escaneos de puertos, identificación de servicios y configuraciones de red, además de destacarse la realización de pruebas y generación de reportes.

Finalmente, frente a la apreciación descrita para algunas herramientas que tienen potencial de automatización, se logra ver que algunas de las respuestas reflejan utilidades que ya

cuentan con una automatización propia dentro de su estructura como es el caso de Nessus y OWASP ZAP, por lo que da a entender un concepto de automatización por parte de los encuestados que involucre automatización en integración con otras herramientas o que permita aplicar automatización con respecto a otros procesos como los seleccionados por los encuestados. En la Fig. 4 se ilustra el modelo conceptual implementado en esta investigación.

IV. Desarrollo de la solucion

De esta manera, para la ejecución de las pruebas se propone la implementación de un RPA que permita por medio de una entrada de tipo IP o dominio web, realizar diferentes procesos con las herramientas mejor valoradas con potencial de automatización por parte de los encuestados y que logren cubrir en su mayoría los procesos descritos.

Se seleccionan en orden de relevancia, la herramienta Nmap utilizada para el proceso de escaneo de puertos y reconocimiento de servicios, Nessus para el escaneo integral de vulnerabilidades y configuraciones, Metasploit en la realización de pruebas de explotación y SQLMap con participación activa en pruebas de inyección de código SQL. Estas fases están descritas dentro de la metodología PTES como las etapas principales para la elaboración de pentesting, todo esto finalizando con la recolección de los resultados de estas pruebas por parte del RPA en un archivo que pueda sintetizarlos.

El RPA se encontrará dentro de una máquina virtual de Kali Linux, lo cual le proporcionará la capacidad de elaborar la integración con herramientas de pentesting de manera mucho más eficiente y donde recorrerá de manera ordenada cada uno de los procedimientos con las herramientas utilizando la herramienta Robocorp dentro del entorno de desarrollo de Visual Studio Code en el lenguaje de programación de Python. El diagrama de flujo que describe el proceso del RPA se describe a continuación.

Dentro del entorno de desarrollo controlado que se dispuso dentro de la máquina virtual, se implementó en Python un script que realiza mediante diferentes pasos, procedimientos con las diferentes herramientas seleccionadas. En primer lugar, se realiza para la fase de reconocimiento y escaneo de puertos un escaneo con Nmap fig. 5 que incluye los servicios que se están ejecutando sobre los puertos que el mismo software logra encontrar y se almacenan dentro de un archivo con los resultados:

```python
def scan_with_nmap(target):

    print(f"Ejecutando escaneo Nmap en el target: {target}")
    result = subprocess.run(["nmap", "-sV", "-A", target], capture_output=True, text=True)
    save_results("Nmap", result.stdout)
```

Fig. 5. Código Python utilizado en Robocorp para ejecución de escaneo en Nmap.

Seguidamente, se ejecuta el proceso con el escáner de vulnerabilidades de Nessus fig.6, haciendo una petición a la API para poder ejecutar un análisis de red básico, y recolectar también los resultados de la operación

```python
def scan_with_nessus(target):
    """Crear, lanzar y obtener resultados de un escaneo en Nessus"""
    scan_data = {
        "uuid": "731a8e52-3ea6-a291-ec0a-d2ff0619c19d7bd788d6be818b65",
        "settings": {
            "name": "ScanExample",
            "description": "This is a test scan",
            "text_targets": target,
            "enabled": True
        }
    }

    try:

        response = make_request_with_retries(f"{NESSUS_URL}/scans", data=scan_data, method="POST")
        response_data = response.json()

        if "scan" in response_data:
            scan_id = response_data["scan"]["id"]
            logging.info(f"Escaneo creado con ID: {scan_id}")
        else:
            raise ValueError(f"No se pudo crear el escaneo: {response.text}")

        start_scan = make_request_with_retries(
            f"{NESSUS_URL}/scans/{scan_id}/launch",
            method="POST"
        )

        if start_scan.status_code == 200:

            logging.info("Escaneo lanzado correctamente.")

        else:
            error_msg = start_scan.json().get("error", "Error desconocido")
            logging.error(f"No se pudo lanzar el escaneo: {error_msg}")
            return
```

Fig. 6. Fragmento de código en Python utilizado para realizar escaneo de vulnerabilidades con la Herramienta Nessus.

Para finalizar, se realiza una prueba de *directory listing* por medio de Metasploit sobre el target y una prueba de inyección SQL por medio de la herramienta SQLMap utilizando las siguientes instrucciones como se muestra en la fig. 7.

```python
def exploit_with_metasploit(target):

    print(f"Ejecutando Metasploit en el target: {target}")

    with open("metasploit_commands.rc", "w") as f:
        f.write(f"use auxiliary/scanner/http/dir_scanner\n")
        f.write(f"set RHOSTS {target}\n")
        f.write("run\n")

    result = subprocess.run(["msfconsole", "-q", "-r", "metasploit_commands.rc"], capture_output=True,
    save_results("Metasploit", result.stdout)

def scan_with_sqlmap(target):

    print(f"Ejecutando SQLMap en el target: {target}")
    url = f"http://{target}/DVWA/vulnerabilities/sqli/?id=1&Submit=Submit"
    result = subprocess.run(["sqlmap", "-u", url, "--batch"], capture_output=True, text=True)
    save_results("SQLMap", result.stdout)
```

Fig. 7. Fragmento de código en Python para realizar escaneo de vulnerabilidades con la Herramienta SQLMap

V. Resultados

La ejecución del robot se da por medio del comando *ROBOT_TARGET= {target} rcc run* el cual permite recibir como parámetro una dirección IP o dominio para realizar las pruebas. Para la ejecución de pruebas se utiliza la IP local, sobre una aplicación desplegada dentro de la máquina virtual, corriendo dentro del mismo servidor y que es conocida por su amplia gama de vulnerabilidades para ejecutar pruebas de seguridad.

Los resultados devueltos por cada proceso de las herramientas se encuentran en un documento que recoge los hallazgos principales del escaneo realizado a la aplicación desplegada de manera local. Principalmente se encuentran los resultados de la fase de Reconocimiento con la herramienta de Nmap fig. 8.

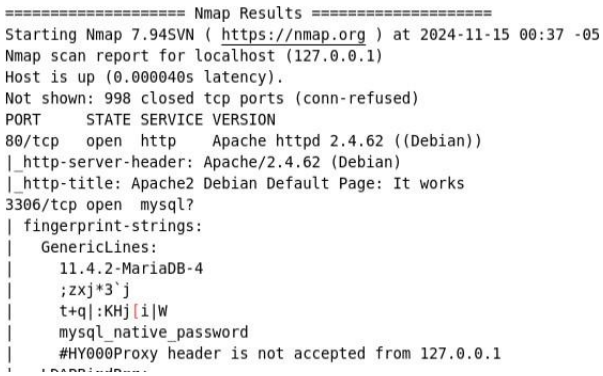


Fig. 8. Resultados del escaneo Nmap por medio del RPA incluidos en un documento de texto.

Los resultados evidenciados anteriormente permiten obtener más información sobre los puertos abiertos dentro del objetivo y algunos de los posibles servicios ejecutándose en los mismos. De manera similar se ejecuta el escaneo de vulnerabilidades utilizando la API de Nessus como insumo principal para dirigir el escaneo.

Se obtiene la lista de vulnerabilidades por parte de Nessus que logra encontrar en el target local, mostrando además dentro del archivo algunos de los controles aplicables como se evidencia en la fig. 9. Comprendiendo la fase de análisis de vulnerabilidades y modelamiento de amenazas propuesta por la metodología PTES.

Fig. 9. Resultados del escaneo Nessus por medio del RPA incluidos en un documento de texto.

De acuerdo con los hallazgos se procede a realizar pruebas de explotación sobre el objetivo utilizando el framework Metasploit y la herramienta SQLMap.

Se completa la prueba retornando algunas de las rutas que fue posible encontrar dentro de la aplicación seguidos de un código que indica la respuesta del target al intentar acceder a estas rutas. fig. 10

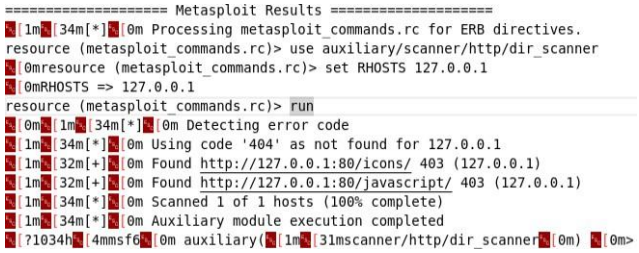


Fig. 10. Resultados de prueba de directory listing utilizando Metasploit Framework por medio del RPA incluidos en un documento de texto.

Finalmente, por medio de la herramienta SQLMap fue posible identificar la tecnología principal que utiliza el target a nivel de base de datos, como se ilustra en la fig. 11 y que es información útil que conlleva a la realización de pruebas combinadas.

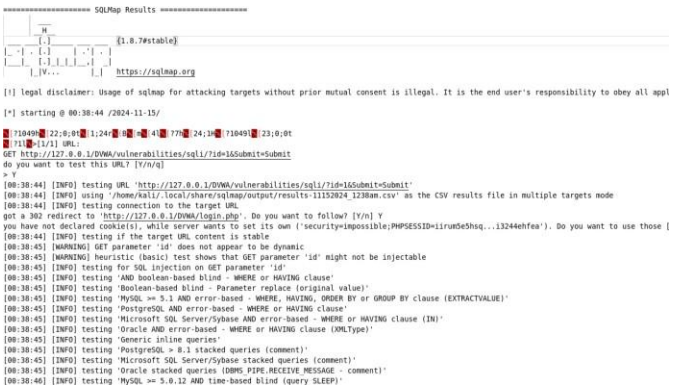


Fig. 11. Resultados de prueba de SQL Injection utilizando SQLMap por medio del RPA incluidos en un documento de texto.

VI. Conclusiones

Dada la creciente importancia de la ciberseguridad en las micro, medianas y pequeñas empresas la implementación de una solución basada en Automatización Robótica de Procesos (RPA) para optimizar tareas de pentesting ha demostrado ser viable y altamente beneficiosa permitiendo que mediante una instrucción puedan ejecutarse pruebas específicas que pueden ser relevantes para una organización y le den resultados que pueda utilizar para comprender qué tipo de información podría tener un agente externo si llegase a darse un ataque a un activo de información.

Mediante el uso del software RPA Robocorp como plataforma base, se logró desarrollar una herramienta de código abierto, escalable y personalizable, adecuada para entornos controlados como una máquina virtual Kali Linux.

Este prototipo ha permitido automatizar de manera eficiente las fases principales de la metodología PTES (desde el reconocimiento hasta la explotación), integrando herramientas clave de ciberseguridad como Nmap, Nessus, Metasploit y SQLMap y algunas pruebas simples como son el escaneo de puertos, escaneo de vulnerabilidades, enumeración de usuarios e inyección de código SQL.

El desarrollo e integración de esta solución destacan por su capacidad para simplificar procesos complejos, reducir significativamente los costos y tiempos de ejecución, además, su flexibilidad permite ampliar el alcance de las auditorías mediante la integración de nuevas herramientas y procesos, ofreciendo un enfoque escalable y adaptable a diferentes escenarios.

Este trabajo no sólo valida el uso de RPA en el contexto del pentesting, sino que también abre nuevas posibilidades para la automatización en ciberseguridad, especialmente en procesos definidos. Sin embargo, queda mucho terreno por explorar en la automatización de tareas más dinámicas y no estructuradas, lo que subraya la necesidad de continuar investigando y desarrollando soluciones innovadoras en esta área que puedan ofrecer además una mejor organización de la información y un catálogo de controles que apoyen a las MiPyMEs.

La herramienta desarrollada representa un avance significativo hacia la optimización del pentesting, demostrando que la automatización no solo es posible, sino también práctica y efectiva en entornos controlados. Este enfoque marca un punto de partida para futuras investigaciones y aplicaciones en la automatización de pruebas de seguridad.

ANEXOS

Anexo 1: Método empleado para la revisión sistemática de la literatura, para no hacer muy extenso este manuscrito es ubico en el siguiente enlace: **Metodología aplicada para el estado del arte**

Anexo 2: La encuesta o consulta a expertos de seguridad informática sobre herramientas de pentesting y los resultados obtenidos se encuentran en el enlace: **Tablero de Power BI con resultados de la encuesta**

Anexo 3: Las preguntas que fueron llevadas a cabo en la encuesta hacia los expertos de seguridad informática es posible encontrarlas en el enlace a continuación: **Encuesta sobre Herramientas de Pentesting y su Automatización con RPA**

REFERENCIAS

[1] I. B. Lahmar, Cybersecurity: Hacking and penetration testing techniques and methodologies, 2021.

[2] I. A. Coronel and D. I. Quirumbay, "Seguridad informática, metodologías, estándares y marco de gestión en un enfoque hacia las aplicaciones web," 2022.

[3] M. C. Ghanem, T. M. Chen, and E. G. Nepomuceno, "Hierarchical reinforcement learning for efficient and effective automated penetration testing of large networks," *Journal of Intelligent Information Systems*, vol. 60, 2022. [Online]. Available: https://doi.org/10.1007/s10844-022-00738-0

[4] E. A. Altulaihan, A. Alismail, and M. Frikha, "A survey on web application penetration testing," *Electronics*, vol. 12, no. 5, 2023. [Online]. Available: https://doi.org/10.3390/electronics12051229

[5] A. M. Ortiz, *Introducción a las pruebas de penetración*, 2020.

[6] A. Arce Rendón, A. Samacá Burbano, and C. Urcuqui López, "Artificial intelligence model for the automation of information collection in the recognition phase of pentesting," 2023.

[7] J. Calle Condori, "Fases de un ataque a un Sistema Informático," *Revista PGI. Investigación, Ciencia y Tecnología en Informática*, no. 7, pp. 52-55, 2020.

[8] J. F. Caranqui Allaica, "Auditoría de la seguridad informática siguiendo la metodología Open Source Security Testing Methodology Manual (OSSTMM) para la empresa MEGAPROFER S.A.," 2020.

[9] ISECOM, *OSSTMM 3*, 2010.

[10] C. Núñez Alcalá, *Penetration testing: Auditoría profesional*, 2021.

[11] OWASP, *Web security testing guide. WSTG - Stable OWASP Foundation*. [Online]. Available: https://owasp.org/www-project-web-security-testing-guide/stable/. [Accessed: 2024].

[12] A. Shanley and M. N. Johnstone, "Selection of penetration testing methodologies: A comparison and evaluation," in *Australian Information Security Management Conference*, 2015.

[13] I. M. Raazia, M. Malahayati, B. Basrulb, R. Maliac, and M. Fadhli, "Analysis server security assessment of staffing management information system using the NIST SP 800-115 method at UIN Ar-Raniry Banda Aceh," *Circuit: Jurnal Ilmiah Pendidikan Teknik Elektro*, vol. 8, 2024. [Online]. Available: https://doi.org/10.22373/crc.v8i1.20808

[14] C. A. Bermúdez Irreño, "RPA - Automatización robótica de procesos: Una revisión de la literatura," *Rev. Ingeniería, Matemáticas y Ciencias de la Información*, vol. 8, 2021. [Online]. Available: https://dx.doi.org/10.21017/rimci.2021.v8.n15.a97

[15] E. K. Chiou and J. D. Lee, "Trusting automation: Designing for responsivity and resilience," *Human Factors*, vol. 65, no. 1, 2023. [Online]. Available: https://doi.org/10.1177/00187208211009995

[16] J. G. Enríquez, A. Jímenez-Ramírez, F. J. Domínguez-Mayo, and J. A. García-García, "Robotic process automation: A scientific and industrial systematic mapping study," *IEEE Access*, vol. 8, 2020. [Online]. Available: https://doi.org/10.1109/ACCESS.2020.2974934

[17] S. Maji, H. Jain, V. Pandey, and V. A. Siddiqui, "White hat security: An overview of penetration testing tools," in *2nd International Conference on Advancement in Electronics & Communication Engineering (AECE 2022)*, 2022.

[18] Z. Asrak, *Penetration testing tools: The use of penetration testing tools in Kali Linux*, 2020.

[19] I. U. Haq and T. A. Khan, "Penetration frameworks and development issues in secure mobile application development," 2021. [Online]. Available: https://doi.org/10.1109/ACCESS.2021.3088229

[20] R. Mehta and R. Chaher, "Implementation of robotic process automation (RPA) in digital marketing," in *3rd International Conference for Emerging Technology (INCET)*, 2022.

[21] J. I. Amador Escalera, "Propuesta metodológica para implementar RPA's," 2020.

[22] J. Siderska, "Robotic process automation — A driver of digital transformation?," *Engineering Management in Production and Services*, vol. 12, no. 2, 2020. [Online]. Available: https://doi.org/10.2478/emj-2020-0009

[23] S. Khan, "Comparative analysis of RPA tools - UiPath, Automation Anywhere and BluePrism," 2020. [Online]. Available: https://doi.org/10.47760/ijcsma.2020.v08i11.001

[24] R. Sindhuja, P. T. Modugu, S. A. Goud, E. R. Kumar, G. S. Babu, and R. Reddy, "A comparative analysis of RPA tools: UiPath, Automation Anywhere and Robocorp," in 2024 OPJU International Technology Conference (OTCON) on Smart Computing for Innovation and Advancement in Industry 4.0, Raigarh, India, 2024, pp. 1-6, doi: 10.1109/OTCON60325.2024.10688237.

[25] J. Ribeiro, R. Lima, T. Eckhardt, and S. Paiva, "Robotic process automation and artificial intelligence in Industry 4.0 – A literature review," in *CENTERIS - International Conference on ENTERprise Information System*, 2021. [Online]. Available: https://doi.org/10.1016/j.procs.2021.01.104

[26] B. Axmann and H. Harmoko, "Process & software selection for robotic process automation (RPA)," 2022. [Online]. Available: https://doi.org/10.31803/tg-20220417182552

[27] P. Desai, S. Joshi, Y. Desai, N. Kothari, and D. Sawant, "Leading platforms in robotic process automation: Review," in *Proceedings of the International Conference on Cognitive and Intelligent Computing*, 2022. [Online]. Available: https://doi.org/10.1007/978-981-19-2350-0_62

[28] D. Andrade, "Challenges of automated software testing with robotic process automation RPA - A comparative analysis of UiPath and Automation Anywhere," International Journal of Intelligent Computing Research, vol. 11, pp. 1066-1072, 2020. doi:10.20533/ijicr.2042.4655.2020.0129.

[29] S. Baweja, "Exploring advanced process automation with Blue Prism," 2023.
[30] N. Delilovic, Implementing Advanced Amazon AWS Authentication Capabilities for the Robot Test-Automation Framework, 2022.
[31] R. Matzenberger, *Exploring open-source robotic process automation: The Robocorp approach*, 2022.
[32] UiPath, "Integrations with enterprise applications - Automation partners." [Online]. Available: https://www.uipath.com/partners/technology-alliances. [Accessed: Nov. 21, 2024].
[33] S. Mandvikar, "Indexing robotic process automation products," International Journal of Computer Trends and Technology, vol. 71, pp. 52-56, 2023. doi: 10.14445/22312803/IJCTT-V71I8P108.
[34] Google Trends, "Explore Google Trends." [Online]. Available: https://trends.google.com. [Accessed: Nov. 5, 2024].
[35] Blue Prism, "Blue Prism RPA software." [Online]. Available: https://www.blueprism.com. [Accessed: Nov. 5, 2024].
[36] Robocorp, "Robocorp - Open source RPA for developers." [Online]. Available: https://robocorp.com. [Accessed: Nov. 21, 2024]
[37] C. Correia, A. Silva, and V. Lobo, "Cybersecurity test automation: Experiences with RPA tools and ZAP technologies using ASL4RPA," in *2024 International Conference on Emerging Computing and Engineering Technologies (ICECET)*, 2024, pp. 1-6, doi: 10.1109/ICECET61485.2024.10698536.

Maria José Yepes Díaz: Egresada del programa Tecnología en Desarrollo de Software del Instituto Tecnológico Metropolitano y estudiante de Ingeniería de Sistemas. Actualmente es miembro activa del semillero de investigación en Ciberseguridad, participando en proyectos enfocados en la automatización de procesos para mejorar la eficiencia en pruebas de penetración (pentesting). https://orcid.org/0009-0000-7369-7687.

Gabriel Enríque Taborda Blandón: PhD en Pensamiento Complejo de Multiversidad Mundo Real Edgar Morin (México), Master en Seguridad Informática por la Universidad Internacional de La Rioja (España) y Master en Computer Science de la Atlantic International University (EEUU). Profesor investigador en el área de Ingeniería de Software y Seguridad Informática. Profesor en el área de Sistemas en educación superior. Actualmente vinculado al Instituto Tecnológico Metropolitano (docente OTC). Liderando línea de investigación en Ciencias Computacionales del Grupo de Investigación en Automatización, Electrónica y Ciencias computacionales del Instituto Tecnológico Metropolitano. https://orcid.org/0000-0002-8067-1490.

Python programming and algebra: Some special elements in Gaussian integers modulo a prime

Programación en Python y álgebra: Algunos elementos especiales en los enteros gaussianos módulo un primo

J. Ávila, J. D. Liévano-González, O. E. Trujillo-Niño

DOI: https://doi.org10.22517/23447214.25748
Scientific and Technological Research Paper

***Abstract* — We consider the complexification of a commutative ring with unity and specialize this construction to $\mathbb{Z}_p[i]$, with p a prime of the form $4k+1$. Since this ring is commutative with unity and is not a field, it is feasible to study various classes of special elements such as invertibles, zero-divisors, idempotents, and nilpotents. The method for this study consists of developing computer programs in Python, through which the lists of special elements in $\mathbb{Z}_p[i]$ are generated for different values of p. The patterns that characterize these lists are sought, in addition to the cardinality of each of these sets. Subsequently, conjectures of mathematical type are stated for each of these classes of elements, which reflect the observed patterns and properties. Finally, formal mathematical proofs of all the conjectures found are made based on various concepts and results of the theory of numbers, groups, and rings. Thus, we show that Python programming, properly used as part of a method, becomes an important tool to identify patterns, properties, and characteristics of several abstract concepts, typical of algebra.**

***Index Terms*— 2-nilpotent; idempotent; invertible; Python; zero-divisor.**

***Resumen* — En este trabajo presentamos la complejificación de un anillo conmutativo con unidad y especializamos esta construcción al anillo de los enteros gaussianos $\mathbb{Z}_p[i]$, con p un primo de la forma $4k+1$. Como este anillo es conmutativo con unidad y no es un cuerpo, resulta viable estudiar diversas clases de elementos especiales como invertibles, divisores de cero, idempotentes y nilpotentes. La metodología seguida para este estudio consiste en desarrollar programas computacionales en Python, mediante los cuales se generan las listas de elementos especiales en $\mathbb{Z}_p[i]$ para distintos valores de p; luego con estas listas se buscan los patrones que caracterizan a los elementos invertibles, idempotentes, divisores de cero y 2-nilpotentes, además del cardinal de cada uno de estos conjuntos. Posteriormente, para cada una de las clases de elementos anteriores se enuncian conjeturas de tipo matemático, las cuales reflejan los patrones y propiedades observadas. Finalmente, apoyados en diversos conceptos y resultados de la teoría de números, grupos y anillos, se hacen las demostraciones matemáticas formales de todas las conjeturas halladas. Mostramos así que la programación en Python, usada adecuadamente en la metodología, se convierte en una herramienta importante para identificar patrones, propiedades y características de diversos conceptos abstractos, propios del álgebra.**

***Palabras clave*— 2-nilpotente; divisor de cero; idempotente; invertible; Python.**

I. Introduction

COMPLEX numbers, since their appearance in the 16th century due to Girolamo Cardano [1], have served as inspiration for many studies in mathematics, engineering, and physics, including countless applications in control theory, electromagnetism, fluid dynamics, signal processing, quantum mechanics, cosmology and cartography, among many others [2,3]. In some cases, it has been necessary and fundamental to consider subsets of the complex numbers, such as the Gaussian integers or Eisenstein integers, since they are the appropriate environment to study various problems in number theory, such as the laws of quadratic and cubic reciprocity [4], or certain subfields of the complex numbers, important for studying the roots of polynomials, extensions of fields and Galois correspondences [5,6]. Currently, the formal construction of the field of complex numbers is based on taking the complete set of ordered pairs with real entries and considering as operations the usual component by component sum and a special product, which is the one that characterizes the complex numbers [7]. The importance of this construction lies in the fact that any ring can be taken as a basis, as shown in [8] and developed in detail in [9,10], which means that virtually any commutative ring with unity can be complexified. This not only allows the construction of new rings, extended via the complex numbers, but also provides another way to introduce some special sets of the complex numbers, such as the ring of Gaussian integers, the field of rational complex numbers, and also finite rings such as the Gaussian integers modulo n, which will be studied in this paper.

This work was submitted on December 20, 2024. Accepted on March 25, 2025. Published on March 31, 2025. This work was partially funded by the University of Tolima, Ibagué, Colombia.

J. Ávila is a professor in the Department of Mathematics and Statistics at the University of Tolima, Ibagué, Colombia (email: javila@ut.edu.co).

J. D. Liévano-González works with RST Asociados, Bogotá D.C., Colombia (email: mineriadedatos@rstasociados.com.co).

O. E. Trujillo-Niño is a Master's student in Mathematical Statistics at the University of Puerto Rico, Mayagüez Campus, Puerto Rico (email: oscar.trujillo1@upr.edu).

On the other hand, the current importance of programming and more broadly of computational thinking lies, among many other things, in the fact that it is a fundamental tool for solving complicated problems, automating tasks, and facilitating multiple aspects of contemporary society, a process that has accelerated due to the emergence and development of artificial intelligence. According to [11],

the concept of computational thinking was implicit in Papert [12], where he spoke of its importance in the framework of his educational proposal known as constructionism, through the work with the robot "turtle" and the programming language LOGO. Papert recognized the importance of an education in which technology should be immersed, so he considered it important to include robotics and programming from an early age, which was not achieved in the following decades. However, according to [11], the work that marks a turning point in the current conception of computational thinking is that developed by Wing [13], where she shows that this concept provides new meanings and dimensions for the human being, with great potential to be developed in educational environments. For [14], computational thinking is a term coined by Wing [13], to describe a set of skills, habits, and comprehensive approaches to solving problems related to programming, which are not only limited to computer use. Thus, in a certain way, we can understand that computational thinking is a set of processes that allow active interaction between a person and a computer, to solve problems of various kinds and make use of patterns, algorithms, models, etc.

Given the above, the Ministries of Education and Ministries or Institutes of Technology in many countries have been creating special initiatives aimed at young people, to foster computational skills, including programming. Even in countries such as England, Spain, the United States, Costa Rica, Ecuador, and Argentina, curricular proposals have been developed to enhance computational and technological skills, including computational thinking and programming [15]. In the Colombian case, the Ministry of Information Technologies and Communications has a Digital Government Policy [16], through which it is developing the Colombia Program and Green Code Strategy initiatives. The purpose of the former is to generate resources and opportunities for teachers' professional development to promote computational thinking in official educational institutions in Colombia, with a focus on gender equity. The second consists of the first learning ecosystem for the development of computational thinking skills for children and young people in public and private schools in Colombia. The ICT Ministry also has the following four proposals for Free Digital Training: Talento Tech, Senatic, Avanza Tech, and Talento GovTech.

As we have seen, programming, and more broadly computational thinking, has become a fundamental tool in practically all areas of knowledge. In mathematics, this tool has been important not only for research but also for education. Currently, several programs for mathematical and statistical use are booming, due to their potential, ease of learning, and because they are open source, allowing any user to use them without any subscription costs for their use. In particular, Python programming has provided countless possibilities in research and teaching, which is why we highlight in this work the use of this language for the study of various algebraic concepts.

Our principal goal in this work is to show the importance of programming to obtain results in algebra. This paper is organized as follows. After the introduction, in Section II, we present the construction of the complexification of any commutative ring with unity. In Section III, we use Python programming to continue the work developed in [9,10]. We study the zero-divisors, idempotents, 2-nilpotents, and invertible elements of the ring $\mathbb{Z}_p[i]$ with p a prime of the form $4k+1$. We want to emphasize the importance of programming to obtain results in algebra. Thus, the implemented methodology consists of developing computer programs in Python language. This leads to obtaining the lists of such elements for different values of the prime p. In the next step, we find the corresponding patterns that characterize these classes of elements, which are presented as conjectures. Finally, we proceed to find the formal proofs of the conjectures obtained, most of which were originally developed by the authors. This same process is done to determine the cardinality of each of these classes.

II. The Complexification Of $\mathbb{Z}_p$, With p Prime

The classical construction of the complex numbers from the real numbers is a topic that is usually addressed in the exercises of an algebra textbook or presented in a course on complex variables. In addition, the three classical representations of the complex numbers, the usual one, the one by matrices, and the one by a quotient, are often overlooked by many students of both engineering and mathematics. The work done in [9,10] recovers these constructions and representations and brings them into more general contexts. Following these two works, we present below the construction of the complexification of any commutative ring with unity.

Let A be a commutative ring with unit element 1. On the set $A \times A$, we define the sum component by component. The product is given by

$$(a,b)(c,d) = (ac - bd, ad + bc),$$

for any $(a,b),(c,d) \in A \times A$.

Affirmation 1. If A is a commutative ring with unit 1, then the set $A \times A$ with the operations given above is a commutative ring with unit.

Proof. Since the sum of pairs is component by component, it is easily observed that $(A \times A, +)$ is an abelian group. Moreover, the product is commutative, since A is commutative. Now for $(a,b),(c,d),(e,f) \in A \times A$ it follows that

$$\begin{aligned}(a,b)\big((c,d)(e,f)\big) &= (a,b)(ce - df, cf + de)\\ &= (ace - adf - bcf - bde,\\ &\quad acf + ade + bce - bdf) \quad (1)\end{aligned}$$

and

$$\begin{aligned}\big((a,b)(c,d)\big)(e,f) &= (ac-bd, ad+bc)(e,f)\\ &= (ace-bde-adf-bcf,\\ &\quad acf-bdf+ade+bce). \quad (2)\end{aligned}$$

Since (1) = (2), then $(a,b)\big((c,d)(e,f)\big) = \big((a,b)(c,d)\big)(e,f)$. That is, the product in $A \times A$ is associative.

Moreover, for $(a,b),(c,d),(e,f) \in A \times A$ it follows that

$$\begin{aligned}(a,b)\big((c,d)+(e,f)\big) &= (a,b)(c+e.d+f)\\ &= (ac+ae-bd-bf,\\ &\quad ad+af+bc+be).\\ &= (a,b)(c,d)+(a,b)(e,f).\end{aligned}$$

That is, the product distributes with respect to the sum on the left. The other distributive property is a consequence of the previous distributivity and commutativity. Finally, it is easy to see that the pair $(1,0)$ is the identity element for the product. Therefore, the set $A \times A$ with the indicated operations is a commutative ring with unity. ■

The ring $(A \times A, +, \cdot)$, described above will be denoted by $A \boxtimes A$. In addition, seeking to make this work complete for the reader, we present below some particularities of this ring.

Affirmation 2. Let A be a commutative ring with unit element 1. Then:

1. The function $\varphi : A \to A \boxtimes A$ defined by $\varphi(a) = (a,0)$ for all $a \in A$ is an injective homomorphism of rings.

2. The element $i = (0,1) \in A \boxtimes A$ commutes with $(m,0)$, for each $m \in A$.

According to Affirmation 2, the First Isomorphism Theorem allows us to conclude that the ring $A \boxtimes A$ contains a subring isomorphic to the ring A or we can also say that any element $a \in A$ is biunivocally identified with the pair $(a,0) \in A \boxtimes A$. Then, for each $(a,b) \in A \boxtimes A$ we have $(a,b) = (a,0) + (b,0)(0,1)$. Moreover, note that the element $i = (0,1)$ satisfies $i^2 = (-1,0)$ and i commutes with each element $(m,0)$. Therefore, using the aforementioned identification we can conclude that $(a,b) = a + bi$ where$a, b \in A$, $i = (0,1)$ is such that $i^2 = -1$ and i commutes with each element$(m,0)$.

Thus, $A \boxtimes A$ can be seen as the "**Complexification of the ring *A***'' [8]. That is, $A \boxtimes A$ coincides with the ring,

$$\mathrm{A}[i] = \{a + bi : a, b \in A, i^2 = -1,\ im = mi, \forall m \in A\}.$$

As particular cases, it is clear that when $A = \mathbb{R}$, the field of complexes $\mathbb{C}$ is obtained; if $A = \mathbb{Z}$, one obtains the ring of Gaussian integers $\mathbb{Z}[i] = \{a + bi : a, b \in \mathbb{Z}\}$ [5,6]; and if $A = \mathbb{Z}_n$, one obtains the Gaussian integers modulo n, $\mathbb{Z}_n[i] = \{a + bi : a, b \in \mathbb{Z}_n\}$, which can also be constructed as the quotient of the ring $\mathbb{Z}[i]$ by the ideal $\langle n \rangle$ in $\mathbb{Z}[i]$ generated by n [17,18].

Since by considering as a base ring the field of reals $\mathbb{R}$, the field of complex numbers $\mathbb{C}$ is obtained, one can generalize this construction by considering finite fields. This is the case in [9,10], where they consider as a base ring the field $\mathbb{Z}_p$, with p a prime, construct the ring $\mathbb{Z}_p[i]$ and show that it is a field if, and only if, p is not a sum of two squares or equivalently p is of the form $4k + 3$. Furthermore, they prove that in this case three isomorphic representations of $\mathbb{Z}_p[i]$, analogous to those obtained from the reals, are obtained. They are obtained using the matrix ring $M(\mathbb{Z}_p) = \left\{\begin{pmatrix} a & b \\ -b & a \end{pmatrix} : a, b \in \mathbb{Z}_p\right\}$, the quotient ring $\mathbb{Z}_p[x]/\langle x^2 + \overline{1}\rangle$ and the one we have developed in this section $\mathbb{Z}_p \boxtimes \mathbb{Z}_p$ (see Figure 1). Thus, we will denote the ring of Gaussian integers over $\mathbb{Z}_p$ as $\mathbb{Z}_p[i]$ or $\mathbb{Z}_p \boxtimes \mathbb{Z}_p$.

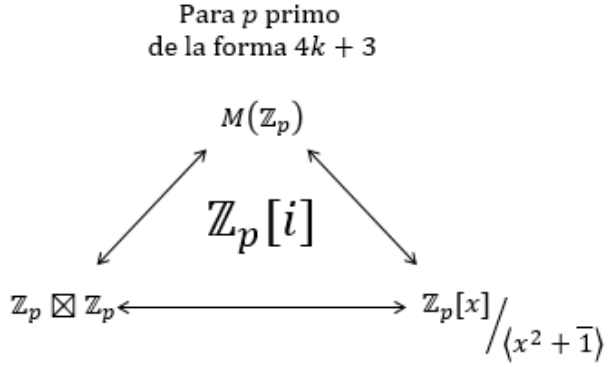


Fig. 1. Isomorphic representations of $\mathbb{Z}_p[i]$ for p prime of the form $4k + 3$.

III. Some Algebraic Concepts In $\mathbb{Z}_p[i]$, With p A Prime Of The Form $4k + 1$

According to the previous section, if p is a prime of the form $4k + 1$, or equivalently p is an odd prime which is the sum of two squares, then $\mathbb{Z}_p[i]$ is a commutative ring with unity, which is not a field. This means that in these rings it is feasible to find various elements important for ring theory such as invertibles, zero-divisors, idempotents, and nilpotents, among many others [5, 8, 19].

Although several of these classes of elements have been characterized in various works [17, 18], the way this has been done is by directly using previous results or theoretical aspects of algebra and number theory to finally, on some occasions, verify these results using programming languages. In this work, we want to emphasize the importance of programming to obtain results in algebra. By continuing the work developed in [9,10], we will use Python to study the zero-divisors, idempotents, 2-nilpotents, and invertible elements of the ring $\mathbb{Z}_p[i]$ with p a prime of the form $4k + 1$. Thus, the process we will follow in this work consists of developing programs in Python and generating the previously described lists of elements of the ring $\mathbb{Z}_p[i]$, for all primes p of the form $4k + 1$ less than 100. Then,

we proceed to identify the patterns or characteristics of each class of these elements, state the conjectures obtained in mathematical terms, and make the corresponding formal proofs. This process is also done to determine the cardinality of each of the sets defined by the different classes of elements. All the results presented here are part of a more general study presented in [20].

Figure 2 shows the program that allows one to perform and visualize the product of elements in $\mathbb{Z}_p[i]$.

```
def simulacionCZp(p):
    lista=list(range(0,p))
    for i in range(0,p):
        for j in range(0,p):
            for k in range(0,p):
                for l in range(0,p):
                    x=[lista[i],lista[j]]
                    y=[lista[k],lista[l]]

                    z=[(x[0]*y[0]-x[1]*y[1])%p,\
                       (x[0]*y[1]+x[1]*y[0])%p]
                    print(x,"*",y,":",z)

    print("multiplicacion de complejos")
```

Fig. 2. Python program for the product in $\mathbb{Z}_p[i]$.

Definition 1. A nonzero element $(\bar{a},\bar{b}) \in \mathbb{Z}_p \boxtimes \mathbb{Z}_p$ is a zero-divisor if there exists a nonzero element $(\bar{c},\bar{d}) \in \mathbb{Z}_p \boxtimes \mathbb{Z}_p$ such that $(\bar{a},\bar{b})(\bar{c},\bar{d}) = (\bar{0},\bar{0})$.

The Python program in Figure 3 was designed to display the zero-divisors for each prime and indicate the number of zero-divisors in each case.

```
def simulacionCZp(p):
    b=[0,0]
    dcero=0
    dcerodist=0
    valores=[]
    lista=list(range(0,p))
    for i in range(0,p):
        for j in range(0,p):
            for k in range(0,p):
                for l in range(0,p):
                    x=[lista[i],lista[j]]
                    y=[lista[k],lista[l]]

                    z=[(x[0]*y[0]-x[1]*y[1])%p,\
                       (x[0]*y[1]+x[1]*y[0])%p]


                    if z==b and x!=b and y!=b:
                        dcero=dcero+1
                        print(x,y)
                        if dcero%(p-1)==1:
                            dcerodist=dcerodist+1


    print("cardinal de la lista:",dcero)
    print("Div. cero distintos:",dcerodist)
```

Fig. 3. Python program for the zero-divisors in $\mathbb{Z}_p[i]$.

The values obtained for $p = 5,\ 13,$ and 17 are shown in Table I. In each pair found by the program, a very interesting particularity is observed. For example, for $p = 5$ if we take the pair of zero-divisors $(\bar{1},\bar{2})$ and $(\bar{1},\bar{3})$, then one has that $\overline{1^2+2^2} = \bar{0}$ and $\overline{1^2+3^2} = \bar{0}$. If for $p = 13$ one takes the pair $(\bar{8},\overline{12})$ and $(\bar{1},\bar{5})$, one has that $\overline{8^2+12^2} = \bar{0}$ and $\overline{1^2+5^2} = \bar{0}$. This pattern is observed for all other pairs of zero-divisors, for all odd primes, sum of two squares less than 100. We can then conjecture that a nonzero element $(\bar{a},\bar{b}) \in \mathbb{Z}_p \boxtimes \mathbb{Z}_p$ is a zero-divisor if, and only if, $\overline{a^2+b^2} = \bar{0}$. Before proving this conjecture, we must prove a preliminary result.

Affirmation 3. Let p be an odd prime which is a sum of two squares. If $(\bar{a},\bar{b})$ is a zero-divisor in $\mathbb{Z}_p \boxtimes \mathbb{Z}_p$, then $\bar{a} \neq \bar{0}$ and $\bar{b} \neq \bar{0}$.

Proof. By assumption, there exists $(\bar{c},\bar{d}) \in \mathbb{Z}_p \boxtimes \mathbb{Z}_p$ nonzero such that $(\bar{a},\bar{b})(\bar{c},\bar{d}) = (\overline{ac-bd},\overline{ad+bc}) = (\bar{0},\bar{0})$. If $\bar{a} = \bar{0}$, then $\overline{-bd} = \bar{0}$ and $\overline{bc} = \bar{0}$. Since $\mathbb{Z}_p$ is an integral domain, one has that $\bar{c} = \bar{0}$ and $\bar{d} = \bar{0}$, which is a contradiction. The same is true if $\bar{b} = \bar{0}$, so we conclude that $\bar{a} \neq \bar{0}$ and $\bar{b} \neq \bar{0}$. ■

TABLE I
ZERO-DIVISORS IN $\mathbb{Z}_5[i]$, $\mathbb{Z}_{13}[i]$, AND $\mathbb{Z}_{17}[i]$

p = 5	p = 13	p = 17
32; distintos 8	288; distintos 24	512; distintos 32
[1, 2] * [1, 3]	[1, 5] * [1, 8]	[1, 4] * [1, 13]
[1, 2] * [2, 1]	[1, 5] * [2, 3]	[1, 4] * [2, 9]
[1, 2] * [3, 4]	[1, 5] * [3, 11]	[1, 4] * [3, 5]
[1, 2] * [4, 2]	[1, 5] * [4, 6]	[1, 4] * [4, 1]
[1, 3] * [1, 2]	[1, 5] * [5, 1]	[1, 4] * [5, 14]
[1, 3] * [2, 4]	[1, 5] * [6, 9]	[1, 4] * [6, 10]
[1, 3] * [3, 1]	[1, 5] * [7, 4]	[1, 4] * [7, 6]
[1, 3] * [4, 3]	[1, 5] * [8, 12]	[1, 4] * [8, 2]
[2, 1] * [1, 2]	[1, 5] * [9, 7]	[1, 4] * [9, 15]
[2, 1] * [2, 4]	[1, 5] * [10, 2]	[1, 4] * [10, 11]
[2, 1] * [3, 1]	[1, 5] * [11, 10]	[1, 4] * [11, 7]
[2, 1] * [4, 3]	[1, 5] * [12, 5]	[1, 4] [12, 3]
[2, 4] * [1, 3]	[1, 8] * [1, 5]	[1, 4] * [13, 16]
[2, 4] * [2, 1]	[1, 8] * [2, 10]	[1, 4] * [14, 12]
[2, 4] * [3, 4]	[1, 8] * [3, 2]	[1, 4] * [15, 8]
[2, 4] * [4, 2]	[1, 8] * [4, 7]	[1, 4] * [16, 4]
[3, 1] * [1, 3]	[1, 8] * [5, 12]	[1, 13] * [1, 4]
[3, 1] * [2, 1]	[1, 8] * [6, 4]	[1, 13] * [2, 8]
[3, 1] * [3, 4]	[1, 8] * [7, 9]	[1, 13] * [3, 12]

Definition 2. Let p be a prime and a an integer relatively prime to p, a relation which is denoted by $(a,p) = 1$. We say that a is a **quadratic residue** or a **square modulo p**, if there exists $b \in \mathbb{Z}$ such that $b^2 \equiv a\ (mod\ p)$.

Note that if a is a quadratic residue modulo p, then there exists $\bar{x} \in \mathbb{Z}_p$ such that $\bar{x}^2 = \bar{a}$. We can then say that $\bar{x}$ is a square root of $\bar{a}$ modulo p and $\overline{-x}$ is also. Then, in this case, we can write $\bar{x} = \pm\sqrt{\bar{a}}$.

For the study of quadratic residues, the Legendre symbol becomes fundamental. It is defined below.

Definition 3. Let p be an odd prime and $a \in \mathbb{Z}$ such that $(a,p) = 1$. Legendre's symbol is defined as:

$$\left(\frac{a}{p}\right) = \begin{cases} 1, & \text{if } a \text{ is a quadratic residue} \\ -1, & \text{if } a \text{ is not a quadratic residue} \end{cases}$$

The following proposition presents some properties of the Legendre symbol. The proofs can be found in many classical texts on number theory [21,22].

Proposition 1. Let p be an odd prime and $a, b \in \mathbb{Z}$ such that $(a,p) = (b,p) = 1$. Then:

1. $\left(\frac{a}{p}\right) \equiv a^{\frac{p-1}{2}} \ (mod\ p)$.
2. $\left(\frac{a^2}{p}\right) = 1$.
3. $\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right)$.
4. If $a \equiv b\ (mod\ p)$, then $\left(\frac{a}{p}\right) = \left(\frac{b}{p}\right)$.
5. $p \equiv 1\ (mod\ 4)$ if, and only if, -1 is a quadratic residue modulo p.
6. $p \equiv 3\ (mod\ 4)$ if, and only if, -1 is not a quadratic residue modulo p.

Affirmation 4. Let p be an odd prime sum of two squares and $(\bar{0},\bar{0}) \neq (\bar{a},\bar{b}) \in \mathbb{Z}_p \boxtimes \mathbb{Z}_p$. Then $(\bar{a},\bar{b})$ is a zero-divisor if, and only if, $\overline{a^2 + b^2} = \bar{0}$.

Proof. If $(\bar{a},\bar{b})$ is a zero-divisor, then there exists $(\bar{c},\bar{d}) \in \mathbb{Z}_p \boxtimes \mathbb{Z}_p$ nonzero such that $(\bar{a},\bar{b})(\bar{c},\bar{d}) = (\bar{0},\bar{0})$. Note then that $(\bar{c},\bar{d})$ is also a zero-divisor and by Affirmation 3, it follows that $\bar{c} \neq \bar{0}$ and $\bar{d} \neq \bar{0}$. From this equality there follows $(\overline{ac - bd}, \overline{ad + bc}) = (\bar{0},\bar{0})$ and thus we obtain

$$\overline{ac - bd} = \bar{0} \qquad (1)$$

$$\overline{ad + bc} = \bar{0} \qquad (2)$$

Multiplying equation (1) by $\bar{a}$, equation (2) by $\bar{b}$ and summing yields $\overline{(a^2 + b^2)c} = \bar{0}$. Since $\bar{c} \neq \bar{0}$, we conclude that $\overline{a^2 + b^2} = \bar{0}$.

Conversely, since $(\bar{0},\bar{0}) \neq (\bar{a},\bar{b}) \in \mathbb{Z}_p \boxtimes \mathbb{Z}_p$ and $\overline{a^2 + b^2} = \bar{0}$, then $(\bar{0},\bar{0}) \neq (\bar{b},\bar{a}) \in \mathbb{Z}_p \boxtimes \mathbb{Z}_p$ and $(\bar{a},\bar{b})(\bar{b},\bar{a}) = (\bar{a}\bar{b} - \bar{b}\bar{a}, \overline{a^2 + b^2}) = (\bar{0},\bar{0})$. That is, $(\bar{a},\bar{b})$ is a zero-divisor. ■

Having characterized the zero-divisors, we also ask how many of them are there, in terms of the prime p. If we denote the cardinality of this set as $|Div(\mathbb{Z}_p[i])|$, then, according to the lists provided by the program (see Table I), it can be seen that for $p = 5$ we have $|Div(\mathbb{Z}_5[i])| = 8 = 2(5) - 2$; for $p = 13$ it follows that $|Div(\mathbb{Z}_{13}[i])| = 24 = 2(13) - 2$; and for $p = 17$, that $|Div(\mathbb{Z}_{17}[i])| = 32 = 2(17) - 2$. This same pattern holds for all odd primes that are a sum of two squares, less than 100. This allows us to conjecture that $|Div(\mathbb{Z}_p[i])| = 2(p-1)$, which is formally proven below.

Affirmation 5. If p is an odd prime which is a sum of two squares, then $|Div(\mathbb{Z}_p[i])| = 2(p-1)$.

Proof. Let $(\bar{a},\bar{b}) \in \mathbb{Z}_p \boxtimes \mathbb{Z}_p$ be a zero-divisor. Then $\bar{a}$ and $\bar{b}$ are nonzero and $\overline{a^2 + b^2} = \bar{0}$. This implies that $\overline{b^2} = \overline{-a^2}$ and by Proposition 1, we have that for all $\bar{a} \in \{\bar{1}, \bar{2}, \ldots, \overline{p-1}\}$:

$$\left(\frac{-a^2}{p}\right) = \left(\frac{-1}{p}\right)\left(\frac{a^2}{p}\right)$$

$$= \left(\frac{-1}{p}\right).1$$

$$= 1.$$

Which signifies that $-a^2$ is a quadratic residue modulo p. Equivalently, $\overline{-a^2}$ has two square roots in $\mathbb{Z}_p$ and they are $\bar{b}$ and $\overline{-b}$. Moreover, $\bar{b}$ and $\overline{-b}$ are different because otherwise you would have $\overline{2b} = \bar{0}$, which leads to $\bar{b} = \bar{0}$, which is a contradiction.

In conclusion, if $(\bar{a},\bar{b}) \in \mathbb{Z}_p \boxtimes \mathbb{Z}_p$ is a zero-divisor, the possibilities for $\bar{a}$ are $p - 1$ and for $\bar{b}$ are 2. That is, there exist $2(p-1)$ zero-divisors. ■

Definition 4. An element $(\bar{a},\bar{b}) \in \mathbb{Z}_p \boxtimes \mathbb{Z}_p$ is called nilpotent if there exists $n \in \mathbb{N}$ such that $(\bar{a},\bar{b})^n = (\bar{0},\bar{0})$. We will say that $(\bar{a},\bar{b})$ is 2 −nilpotent if $(\bar{a},\bar{b})^2 = (\bar{0},\bar{0})$.

The Python program in Figure 4 is designed to find the 2-nilpotent elements for each prime p and at the same time indicate the number of 2-nilpotents in each case.

```
def simulacionCZp(p):
    b=[0,0]
    nilp=0
    lista=list(range(0,p))
    for i in range(0,p):
        for j in range(0,p):
            x=[lista[i],lista[j]]
            y=[lista[i],lista[j]]

            z=[(x[0]*y[0]-x[1]*y[1])%p,\
               (x[0]*y[1]+x[1]*y[0])%p]

            if z==b:
                print(x,"*",y,":",z)
                nilp=nilp+1

    print("lista nilpotentes")
    print("numero de elementos 2-nilpotentes:",nilp)
```

Fig. 4. Python program for the 2-nilpotent elements in $\mathbb{Z}_p[i]$.

The values obtained for $p = 5, 13, 17, 29,$ and 37 are given in Table II. Note that in all cases, there is only one 2-nilpotent

element, the trivial $(\bar{0},\bar{0})$. This is also observed for the other odd primes that are sums of two squares and are less than 100. So, we can conjecture that only the null element of $\mathbb{Z}_p[i]$ is $2-$nilpotent, which is proved below.

TABLE II
2-NILPOTENT ELEMENTS IN $\mathbb{Z}_p[i]$, FOR $p = 5, 13, 17, 29,$ AND 37

Primo	Lista Nilpotentes	Total Lista
p = 5	[0, 0] * [0, 0] = [0, 0]	1
p = 13	[0, 0] * [0, 0] = [0, 0]	1
p = 17	[0, 0] * [0, 0] = [0, 0]	1
p = 29	[0, 0] * [0, 0] = [0, 0]	1
p = 37	[0, 0] * [0, 0] = [0, 0]	1

Affirmation 6. Let p be an odd prime which is a sum of two squares. Then, the only $2-$nilpotent element of $\mathbb{Z}_p \boxtimes \mathbb{Z}_p$ is $(\bar{0},\bar{0})$.

Proof. If $(\bar{a},\bar{b})$ is $2-$nilpotent, then $(\bar{a},\bar{b})(\bar{a},\bar{b}) = (\bar{0},\bar{0})$. That is, $(\overline{a^2-b^2}, \overline{ab+ba}) = (\bar{0},\bar{0})$ and we obtain the system

$$\overline{a^2-b^2} = \bar{0} \quad (3)$$

$$\overline{2ab} = \bar{0} \quad (4)$$

Equation (4) implies that $\bar{a} = \bar{0}$ or $\bar{b} = \bar{0}$. Replacing either of the two options in equation (3) yields $\bar{b} = \bar{0}$ or $\bar{a} = \bar{0}$, respectively. In conclusion, $\bar{a} = \bar{0}$ and $\bar{b} = \bar{0}$. ■

Definition 5. An element $(\bar{a},\bar{b}) \in \mathbb{Z}_p \boxtimes \mathbb{Z}_p$ is called idempotent if $(\bar{a},\bar{b})^2 = (\bar{a},\bar{b})$.

It is clear that $(\bar{0},\bar{0})$ and $(\bar{1},\bar{0})$ are idempotents of $\mathbb{Z}_p \boxtimes \mathbb{Z}_p$, which are called trivial idempotents.

The Python program in Figure 5 was designed to show the idempotent elements for each prime p and at the same time indicate the number of these elements.

```
def simulacionCZp(p):
    idemp=0
    lista=list(range(0,p))
    for i in range(0,p):
        for j in range(0,p):
            x=[lista[i],lista[j]]
            y=[lista[i],lista[j]]

            z=[(x[0]*y[0]-x[1]*y[1])%p,\
               (x[0]*y[1]+x[1]*y[0])%p]


            if x==y and x==z:
                idemp=idemp+1
                print(x,"*",y,":",z)

    print("idempotentes:",idemp)
```

Fig. 5. Python program for the idempotent elements in $\mathbb{Z}_p[i]$.

The results given by the above program for $p = 5, 13, 17, 29, 37,$ and 41 are shown in Table III. It can be observed that the nontrivial idempotents for $p = 5$ are $(\bar{3},\bar{1})$ and $(\bar{3},\bar{4})$. In both cases the first component is $\bar{3}$, $\bar{3} = \overline{\left(\frac{5+1}{2}\right)}$ and furthermore $\overline{3^2+1^2} = \bar{0}$ and $\overline{3^2+4^2} = \bar{0}$. For $p = 13$ they are $(\bar{7},\bar{4})$ and $(\bar{7},\bar{9})$. In both cases the first component is $\bar{7}$, $\bar{7} = \overline{\left(\frac{13+1}{2}\right)}$ and moreover $\overline{7^2+4^2} = \bar{0}$ and $\overline{7^2+9^2} = \bar{0}$. This same pattern is observed in the other primes in the table and in all other odd primes less than 100 that are the sum of two squares. Thus, we conjecture that $(\bar{a},\bar{b})$ is a nontrivial idempotent if, and only if, $\bar{a} = \overline{\left(\frac{p+1}{2}\right)}$ and $\overline{a^2+b^2} = \bar{0}$. This is proved below.

TABLE III
IDEMPOTENT ELEMENTS IN $\mathbb{Z}_p[i]$, FOR $p = 5, 13, 17, 29, 37,$ AND 41

Primo	Lista Idempotentes	Total Lista
p = 5	[0, 0] * [0, 0] = [0, 0] [1, 0] * [1, 0] = [1, 0] [3, 1] * [3, 1] = [3, 1] [3, 4] * [3, 4] = [3, 4]	4
p = 13	[0, 0] * [0, 0] = [0, 0] [1, 0] * [1, 0] = [1, 0] [7, 4] * [7, 4] = [7, 4] [7, 9] * [7, 9] = [7, 9]	4
p = 17	[0, 0] * [0, 0] = [0, 0] [1, 0] * [1, 0] = [1, 0] [9, 2] * [9, 2] = [9, 2] [9, 15] * [9, 15] = [9, 15]	4
p = 29	[0, 0] * [0, 0] = [0, 0] [1, 0] * [1, 0] = [1, 0] [15, 6] * [15, 6] = [15, 6] [15, 23] * [15, 23] = [15, 23]	4
p = 37	[0, 0] * [0, 0] = [0, 0] [1, 0] * [1, 0] = [1, 0] [19, 3] * [19, 3] = [19, 3] [19, 34] * [19, 34] = [19, 34]	4
p = 41	[0, 0] * [0, 0] = [0, 0] [1, 0] * [1, 0] = [1, 0] [21, 16] * [21, 16] = [21, 16] [21, 25] * [21, 25] = [21, 25]	4

Recall that if e is a nontrivial idempotent in a ring with unity 1, then e is a zero-divisor since one has $e(e-1) = 0$.

Affirmation 7. Let p be an odd prime which is the sum of two squares and $(\bar{a},\bar{b}) \in \mathbb{Z}_p \boxtimes \mathbb{Z}_p$. Then $(\bar{a},\bar{b})$ is a nontrivial idempotent if, and only if, $\bar{a} = \overline{\left(\frac{p+1}{2}\right)}$ and $\overline{a^2+b^2} = \bar{0}$.

Proof. If $(\bar{a},\bar{b})$ is a nontrivial idempotent, then $(\bar{a},\bar{b})$ is a zero-divisor and by Affirmation 3, $\bar{a} \neq \bar{0}$ and $\bar{b} \neq \bar{0}$. The following system of equations results:

$$\overline{a^2 - b^2} = \bar{a} \quad (5)$$

$$\overline{2ab} = \bar{b} \quad (6)$$

From Equation (6), one has $\bar{2}\bar{a} = \bar{1} = \overline{p+1} \Longrightarrow \bar{a} = \overline{\left(\frac{p+1}{2}\right)}$. Substituting in Equation (5) we obtain

$$\begin{aligned} \overline{a^2 + b^2} &= \overline{2a^2 - a} \\ &= \overline{2\left(\frac{(p+1)^2}{4}\right) - \left(\frac{p+1}{2}\right)} \\ &= \overline{\left(\frac{p+1}{2}\right) - \left(\frac{p+1}{2}\right)} \\ &= \bar{0}. \end{aligned}$$

On the other hand, let $(\bar{a}, \bar{b}) \in \mathbb{Z}_p \boxtimes \mathbb{Z}_p$ with $\bar{a} = \overline{\left(\frac{p+1}{2}\right)}$ and $\overline{a^2 + b^2} = \bar{0}$. Then, $\overline{b^2} = \overline{-a^2}$ and moreover,

$$\begin{aligned} (\bar{a}, \bar{b})^2 &= \left(\overline{a^2 - b^2}, \overline{2ab}\right) \\ &= \left(\overline{2a^2}, \bar{b}\right) \\ &= \left(\overline{2\left(\frac{(p+1)^2}{4}\right)}, \bar{b}\right) \\ &= \left(\overline{\left(\frac{p+1}{2}\right)}, \bar{b}\right) \\ &= (\bar{a}, \bar{b}). \end{aligned}$$

That is, $(\bar{a}, \bar{b})$ is idempotent. ■

For the number of nontrivial idempotent elements, it suffices to look at Table III. The results indicate that independently of the prime p, exactly two nontrivial idempotents are always found. That is, if we denote the cardinality of this set by $\left|Id(\mathbb{Z}_p[i])\right|$, then $\left|Id(\mathbb{Z}_p[i])\right| = 2$. This will be proved below.

Affirmation 8. If p is an odd prime which is a sum of two squares, then $\left|Id(\mathbb{Z}_p[i])\right| = 2$.

Proof. Let $(\bar{a}, \bar{b}) \in \mathbb{Z}_p \boxtimes \mathbb{Z}_p$ be a nontrivial idempotent. Then $\bar{a} = \overline{\left(\frac{p+1}{2}\right)}$ and $\overline{a^2 + b^2} = \bar{0}$. Thus, $\bar{a}$ has a fixed value that depends on p and $\bar{b}^2 = \overline{-a^2}$. By Proposition 1 and the proof of Affirmation 7, $\overline{-a^2}$ has two different square roots in $\mathbb{Z}_p$. That is, $\bar{b}$ takes two different values, which implies that $\left|Id(\mathbb{Z}_p[i])\right| = 2$. ■

To finish this work, it remains to study the invertible elements of $\mathbb{Z}_p[i]$.

Definition 6. An element $\bar{a} + \bar{b}i \in \mathbb{Z}_p[i]$ is invertible if there is a $\bar{c} + \bar{d}i \in \mathbb{Z}_p[i]$ with $(\bar{a} + \bar{b}i)(\bar{c} + \bar{d}i) = 1$.

The Python program in Figure 6 was designed to display the invertible elements for each prime p and at the same time indicate the number of these elements.

```
def simulacionCZp(p):
    b=[1,0]
    unos=0

    lista=list(range(0,p))
    for i in range(0,p):
        for j in range(0,p):
            for k in range(0,p):
                for l in range(0,p):
                    x=[lista[i],lista[j]]
                    y=[lista[k],lista[l]]

                    z=[(x[0]*y[0]-x[1]*y[1])%p,\
                       (x[0]*y[1]+x[1]*y[0])%p]

                    if z==b:
                        unos=unos+1
                        print(x,"*",y,":",z)

    print("Inversos:",unos)
```

Fig. 6. Python program for the invertible elements in $\mathbb{Z}_p[i]$.

The results obtained by the above program for $p = 5$ and $p = 13$ are observed in Tables IV and V. We can observe that the only invertible pairs $(\bar{a}, \bar{b})$ are those with $\overline{a^2 + b^2} \neq \bar{0}$. The same result is observed for all odd primes less than 100 that are the sum of two squares. The proof of this is analogous to that presented in [9, 10], we include it for completeness of this paper and as a benefit to the reader.

TABLE IV
INVERTIBLE ELEMENTS IN $\mathbb{Z}_5[i]$

p = 5	
Total inversos = 16	
[0, 1] * [0, 4] = [1, 0]	[2, 2] * [4, 1] = [1, 0]
[0, 2] * [0, 2] = [1, 0]	[2, 3] * [4, 4] = [1, 0]
[0, 3] * [0, 3] = [1, 0]	[3, 0] * [2, 0] = [1, 0]
[0, 4] * [0, 1] = [1, 0]	[3, 2] * [1, 1] = [1, 0]
[1, 0] * [1, 0] = [1, 0]	[3, 3] * [1, 4] = [1, 0]
[1, 1] * [3, 2] = [1, 0]	[4, 0] * [4, 0] = [1, 0]
[1, 4] * [3, 3] = [1, 0]	[4, 1] * [2, 2] = [1, 0]
[2, 0] * [3, 0] = [1, 0]	[4, 4] * [2, 3] = [1, 0]

Affirmation 9. Let p be an odd prime which is the sum of two squares and $\bar{a} + \bar{b}i \in \mathbb{Z}_p[i]$. Then, $\bar{a} + \bar{b}i$ is invertible if, and only if, $\overline{a^2 + b^2} \neq \bar{0}$.

Proof. If $\overline{a^2 + b^2} = \bar{0}$, by Affirmation 4 we have that $\bar{a} + \bar{b}i$ is a zero-divisor and thus it is not invertible.

On the other hand, if $\overline{a^2+b^2} \neq \bar{0}$, then it is easy to see that the multiplicative inverse of $\bar{a}+\bar{b}i$ is $(\bar{a}+\bar{b}i)^{-1} = (\overline{a^2+b^2})^{-1}(\bar{a}-\bar{b}i) \in \mathbb{Z}_p[i]$. ■

TABLE V
INVERTIBLE ELEMENTS IN $\mathbb{Z}_{13}[i]$

p = 13	
Total inversos = 144	
[0, 1] * [0, 12] = [1, 0]	[1, 7] * [6, 10] = [1, 0]
[0, 2] * [0, 6] = [1, 0]	[1, 9] * [10, 1] = [1, 0]
[0, 3] * [0, 4] = [1, 0]	[1, 10] * [4, 12] = [1, 0]
[0, 4] * [0, 3] = [1, 0]	[1, 11] * [8, 3] = [1, 0]
[0, 5] * [0, 5] = [1, 0]	[1, 12] * [7, 7] = [1, 0]
[0, 6] * [0, 2] = [1, 0]	[2, 0] * [7, 0] = [1, 0]
[0, 7] * [0, 11] = [1, 0]	[2, 1] * [3, 5] = [1, 0]
[0, 8] * [0, 8] = [1, 0]	[2, 2] * [10, 3] = [1, 0]
[0, 9] * [0, 10] = [1, 0]	[2, 4] * [4, 5] = [1, 0]
[0, 10] * [0, 9] = [1, 0]	[2, 5] * [5, 7] = [1, 0]
[0, 11] * [0, 7] = [1, 0]	….....
[0, 12] * [0, 1] = [1, 0]	[12, 6] * [7, 3] = [1, 0]
[1, 0] * [1, 0] = [1, 0]	[12, 7] * [7, 10] = [1, 0]
[1, 1] * [7, 6] = [1, 0]	[12, 9] * [3, 1] = [1, 0]
[1, 2] * [8, 10] = [1, 0]	[12, 10] * [9, 12] = [1, 0]
[1, 3] * [4, 1] = [1, 0]	[12, 11] * [5, 3] = [1, 0]
[1, 4] * [10, 12] = [1, 0]	[12, 12] * [6, 7] = [1, 0]

As for the number of invertible elements, the program shows in Tables IV and V, that for $p = 5$ there are 16 invertibles and for $p = 13$ there are 144 invertibles. We note then that $16 = (5-1)^2$ and $144 = (13-1)^2$. Moreover, the same can be observed for the other odd primes less than 100 that are the sum of two squares. Thus, if we denote the cardinality of this set by $|Inv(\mathbb{Z}_p[i])|$, then $|Inv(\mathbb{Z}_p[i])| = (p-1)^2$. This is proved below.

Affirmation 10. If p is an odd prime which is the sum of two squares, then $|Inv(\mathbb{Z}_p[i])| = (p-1)^2$.

Proof. According to Affirmations 4 and 9, it can be concluded that every nonzero element $\bar{a}+\bar{b}i \in \mathbb{Z}_p[i]$, is a zero-divisor or invertible: this depends on whether $\overline{a^2+b^2} = \bar{0}$ or $\overline{a^2+b^2} \neq \bar{0}$, respectively. That is, the set $\mathbb{Z}_p[i]$ is partitioned into three classes, the zero-divisors, the invertibles, and the zero element. By Affirmation 5, $p^2 = |Inv(\mathbb{Z}_p[i])| + 2(p-1) + 1$, which implies that $|Inv(\mathbb{Z}_p[i])| = p^2 - 2(p-1) - 1 = (p-1)^2$, which is what we wanted to prove. ■

IV. CONCLUSIONS

We present below the most relevant aspects that emerged during the development of this work and at the same time we would like to make some recommendations to continue with this study, providing new elements for discussion and research.

- The process employed in this work allows using computational programming to conjecture results in algebra. This shows that programming is not only useful in engineering or applied sciences but also allows interesting computational studies in algebra, an abstract area. Consequently, many of the known processes in mathematical problem-solving using programming were evidenced, such as: problem understanding, exploration, case study, program design and implementation, desktop testing, and evaluation.

- We emphasize the pedagogical importance of the process followed in this work to obtain the results. This allows, through computer programming, to obtain results, which in turn lead to a differentiated mental development in terms of the observation of patterns, formulation of hypotheses, and finally the formal proofs of the assertions. In the same way, other mental processes are developed in the student as a consequence of the deep understanding of the set being studied together with its structure, the programming of the different algebraic concepts, analysis of the results, observation of patterns, formulation of hypotheses, and their theoretical proof.

- The computer programs developed in this work can be modified to study other important elements in a ring such as nilpotent in general, regular, associated, and irreducible, among many others [5, 8, 19]. In this case, one could also consider various rings of integers modulo n or some subclasses as $\mathbb{Z}_p, \mathbb{Z}_{p^\alpha}, \mathbb{Z}_{pq}$ with p, q primes and the corresponding complexification of each of them. Even other sets of integers modulo n such as Eisenstein, Hurwitz and Lipschitz integers could be considered [23, 24].

- As a continuation of this work and also relying on computational programming, additional studies on the group of invertible elements, its generators, and the cardinality of this set can be considered. As for the zero-divisors we can observe that the simulations found can be used to determine which and how many are the pairs $(\bar{c}, \bar{d})$ such that $(\bar{a}, \bar{b})(\bar{c}, \bar{d}) = (\bar{0}, \bar{0})$, where $(\bar{a}, \bar{b})$ is a fixed zero-divisor. That is, in the language of graphs we would be thinking about determining which and how many vertices are connected to the given vertex $(\bar{a}, \bar{b})$. This would lead to the study of the zero-divisor graph of the ring $\mathbb{Z}_p[i]$ with p an odd prime of the form $4k+1$ or more generally of the rings $\mathbb{Z}_n[i]$ and $\mathbb{Z}_n$ [8]. Finally, as for idempotents, this paper shows that in the case of the rings $\mathbb{Z}_p[i]$ with p an odd prime of the form $4k+1$, only two nontrivial idempotents result, which does not allow us to go deeper into this ring. However, when considering more general rings such as $\mathbb{Z}_n$, $\mathbb{Z}_n[i]$ and even quaternions modulo n, a nontrivial number of idempotents arise [20]. This makes viable a deeper study of them in terms of characterization, cardinality, classes of these, associated ordered set, and all notions arising from this order [19].

REFERENCES

[1] I. Stewart, *Significant Figures: The Lives and Works of Greats Mathematician*. New York, NY, USA: Basic Books, 2007.

[2] R. Penrose, *The Road to Reality: A Complete Guide to the Laws of the Universe*. New York, NY, USA: Alfred. A. Knopf, 2005.

[3] S. Ponnusamy and H. Silverman, *Complex Variables with Applications*. Boston, MA, USA: Birkhäuser, 2006.

[4] I. Kleiner, "From numbers to rings: The early history of ring theory," *Elem. Math.*, vol 53, pp. 18-35, 1998, DOI 10.1007/s000170050029.

[5] D. S. Dummit and R. M. Foote, *Abstract Algebra*. 3rd ed. New Delhi, IND: Wiley India Pvt. Ltd., 2016.

[6] J. B. Fraleigh, *A First Course in Abstract Algebra*. 7th ed. New York, NY, USA: Pearson, 2002.

[7] J. B. Conway, *Functions of One Complex Variable*. 2nd ed. New York, NY, USA: Springer New York, 1978.

[8] K. H. Spindler, *Abstract Algebra and Applications Vol. II, Rings and Fields*. New York, NY, USA: Marcel Dekker Inc., 1994.

[9] J. Ávila, J. S. Correa-Amaya y E. Cupitra-Vergara, "Números complejos sobre anillos," *Scientia Et Technica*, vol. 23, no. 04, pp. 581-585, 2018, DOI 10.22517/23447214.18171.

[10] J. S. Correa-Amaya y E. Cupitra-Vergara, "De los complejos a $\mathbb{Z}_p \times \mathbb{Z}_p$, p primo," Tesis de Pregrado, Departamento de Matemáticas y Estadística, Universidad del Tolima, Ibagué, Colombia, 2016.

[11] M. D. M. Sánchez-Vera, "La robótica, la programación y el pensamiento computacional en la educación infantil," *Infancia, Educación y Aprendizaje (IEYA)*, vol. 7, no. 1, pp. 209-234, 2021, DOI 10.22370/ieya.2021.7.1.2343.

[12] S. Papert, "Mindstorms: Children, Computers and Powerful ideas," New York, NY, USA: Basic Books Inc. Publishers, 1980. [Online]. Available: https://worrydream.com/refs/Papert_1980_-_Mindstorms,_1st_ed.pdf. Accessed on: Dec 9, 2024.

[13] J. M. Wing, "Computational thinking," *Communications of the ACM*, vol. 49, no. 3, pp. 33-35, 2006, DOI 10.1145/1118178.1118215.

[14] L. A. Sánchez-Ruíz, "Comprendiendo el pensamiento computacional: Experiencias de programación a través de Scratch en colegios públicos de Bogotá," Tesis de Maestría, Departamento de Psicología, Universidad Nacional de Colombia, Bogotá, Colombia, 2016. https://repositorio.unal.edu.co/bitstream/handle/unal/57424/luisalfredosanchezruiz.2016.pdf?sequence=1&isAllowed=y Accessed on: Mar 28, 2025.

[15] E. Coronel-Díaz y G. Lima-Silvain, "El pensamiento computacional. Nuevos retos para la educación del siglo XXI," *Virtualidad, Educación y Ciencia*, año 11, no. 20, pp. 115-137, 2020, DOI 10.60020/1853-6530.v11.n20.27451.

[16] Ministerio de Tecnologías de la Información y las Comunicaciones. *Decreto no. 767 de 16 mayo 2022, Por el cual se establecen los lineamientos generales de la Política de Gobierno Digital y se subroga el Capítulo 1 del Título 9 de la Parte 2 del Libro 2 del Decreto 1078 de 2015, Decreto Único Reglamentario del Sector de Tecnologías de la Información y las Comunicaciones*. [Online]. Available: https://gobiernodigital.mintic.gov.co/692/articles-272977_Decreto_767_2022.pdf. Accessed on: Mar 28, 2025.

[17] A. A. Allam, M. J. Dunne, J. R. Jack, J. C. Lynd, and H. W. Ellingsen Jr., "Classification of the group of units in the Gaussian integers modulo N," *Pi Mu Epsilon Journal*, vol. 12, no. 9, pp. 513-519, 2008. [Online]. Available: http://www.jstor.org/stable/24345265. Accessed 28 Mar. 2025.

[18] O. Emad-Abu, A. A. Salah, and J. Nafiz-Abu, "Zero divisor graph for the ring of Gaussian integers modulo n," *Commun. Algebra*, vol. 36, no. 10, pp. 3865-3877, 2008, DOI 10.1080/00927870802160859.

[19] T. Y. Lam, *A First Course in Noncommutative Rings*. New York, NY, USA: Springer-Verlag, 2001.

[20] J. D. Liévano y O. E. Trujillo, "Algunos conceptos algebraicos en los complejos y cuaterniones sobre $\mathbb{Z}_p$, p primo," Tesis de Pregrado, Departamento de Matemáticas y Estadística, Universidad del Tolima, Ibagué, Colombia, 2021. [Online]. Available: https://repository.ut.edu.co/server/api/core/bitstreams/a48ec74b-3d68-4dc3-b751-791f1da1f930/content. Accessed on: Mar 28, 2025.

[21] R. Jiménez, E. Gordillo y G. Rubiano, *Teoría de Números para Principiantes*. 2a ed. Bogotá, COL: Universidad Nacional de Colombia, 2014.

[22] I. M. Vinogradov, *Fundamentos de la Teoría de los Números*. Moscú, URSS: Mir, 1977.

[23] A. Osama and Emad-Abu, "On Eisenstein integers modulo n," *Int. Math. Forum*, vol. 5, no. 22, pp. 1075-1082, 2010. [Online]. Available: https://www.m-hikari.com/imf-2010/21-24-2010/index.html. Accessed on: Mar 28, 2025.

[24] N. Tsopanidis, "The Hurwitz and Lipschitz integers and some applications," Tese de Doutoramento, Departamento de Matemática, Universidade do Porto, Porto, Portugal, 2020. [Online]. Available: https://repositorio-

aberto.up.pt/bitstream/10216/131608/2/437378.pdf. Accessed on: Mar 28, 2025.

J. Avila. He was born in Líbano, Tolima - Colombia, in 1972. He graduated in Mathematics and Physics from the Universidad del Tolima in 1995. He studied a Master's Degree in Mathematics at the Universidad Nacional de Colombia, Bogotá, and graduated in 2002. PhD. in Mathematics from the Universidade Federal do Rio Grande do Sul, Brazil, and graduated in 2008. He currently works as a professor at the Universidad del Tolima. His areas of interest are algebra and topology.https://orcid.org/0000-0002-8713-2449

J. D. Liévano-González. He was born in Ibagué, Tolima - Colombia, in 1993. He completed undergraduate studies at the Universidad del Tolima, obtaining a degree of Professional in Mathematics with Emphasis in Statistics in 2021. Subsequently, he took certified courses in software development and data science. Currently, he works as a data scientist in the collection management industry.https://orcid.org/0009-0001-1898-7494

O. E. Trujillo-Niño. He was born in Ibagué, Tolima-Colombia, in 1996. He graduated in Mathematics with Emphasis in Statistics from the Universidad del Tolima in 2022. He is currently studying for a master's degree in mathematics-statistics at the University of Puerto Rico, Mayagüez campus. He is also working as a data scientist in the company Lumni Colombia S.A. https://orcid.org/0009-0003-8421-9914

Bibliometric analysis of water resource quality conservation in agroecological structures

Análisis bibliométrico de la conservación de la calidad del recurso hídrico en estructuras agroecológicas

K. J. Meléndez-Segura; H. Reyes-Pineda; J. F. Betancur-Pérez

DOI: https://doi.org/ 10.22517/23447214.25744
Scientific and technological research paper

***Abstract*— Water conservation is essential for agricultural sustainability and adaptation to climate change. Agroforestry systems, which integrate trees and sometimes animals in the same production unit, offer hydrological benefits superior to those of conventional agricultural systems, positioning them as a key strategy for water resource management in agricultural areas. However, in Colombia, publications on the subject are limited. Understanding their dynamics is crucial to preserving water quality, especially in sectors such as fish farming, which is vital for biodiversity and ecosystem services. This study aimed to perform a bibliometric and systematic analysis of the Web of Science (WoS) database; the data were examined using graph theory and specialized tools such as VOSviewer and Tree of Science, investigating in diverse perspectives that explore the association between agroecological structures, water quality and conservation. The analysis was structured in three categories: roots (classic), trunk (structural) and branches-leaves (recent); inclusion and exclusion criteria were also applied, adopting certain guidelines of the PRISMA protocol (Preferred Reporting Items for Systematic Reviews and Meta-Analyses); 61 relevant references were identified, highlighting Chen, Hung-Chih as the most cited author, and the United States and China as the leading countries in research on the subject. It was concluded that agroforestry systems are fundamental to conserve and improve water quality, promoting ecological agriculture and fish farming, welfare and sustainability, contributing to regional economic development; furthermore, there are knowledge gaps and it is necessary to have policies for the construction, research and implementation of these systems.**

***Index Terms*— Agroecological systems; ecosystem services; pollutants; water conservation; water quality.**

***Resumen*— La conservación del recurso hídrico es esencial para la sostenibilidad agrícola y la adaptación al cambio climático. Los sistemas agroforestales, que integran árboles y en ocasiones animales en una misma unidad de producción, ofrecen beneficios hidrológicos superiores a los de los sistemas agrícolas convencionales, posicionándolos como una estrategia clave para la gestión de recursos hídricos en áreas agrícolas. Sin embargo, en Colombia, las publicaciones sobre revisión del tema son limitadas. Comprender su dinámica es crucial para preservar la calidad del agua, especialmente en sectores como el piscícola, vital para la biodiversidad y los servicios ecosistémicos. Este estudio tuvo como objetivo realizar un análisis bibliométrico y sistemático de la base de datos Web of Science (WoS); los datos fueron examinados utilizando la teoría de grafos y herramientas especializadas como VOSviewer y Tree of Science, investigando en diversas perspectivas que exploran la asociación entre estructuras agroecológicas, calidad y conservación del agua. Se estructuró el análisis en tres categorías: raíces (clásicos), tronco (estructurales) y ramas-hojas (recientes) de igual forma se aplicó criterios de inclusión y excusión adoptando ciertas directrices del protocolo PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses); se identificaron 61 referencias relevantes, destacando a Chen, Hung-Chih como el autor más citado, y a Estados Unidos y China como los países líderes en investigaciones sobre el tema. Se concluyó que los sistemas agroforestales son fundamentales para conservar y mejorar la calidad del agua, promoviendo la agricultura y piscicultura ecológica, el bienestar y la sostenibilidad, aportando al desarrollo económico regional; además que existen brechas de conocimiento y es necesario que existan políticas para la construcción, investigación e implementación de estos sistemas.**

***Palabras claves*— Calidad del agua; conservación del agua; contaminantes; servicios ecosistémicos; sistemas agroecológicos.**

I. INTRODUCTION

WATER is an essential resource for life and agricultural production, the quality of which is critical to the health of ecosystems and human communities. Without clean water, fundamental activities such as irrigation, livestock raising, fish farming and the maintenance of ecosystem services are seriously compromised.

However, water quality can deteriorate due to various human activities, especially those related to intensive agriculture. Excessive use of fertilizers, pesticides, and other agrochemicals leads to leaching of nutrients into groundwater, surface runoff that contaminates water bodies with sediments and chemicals, and soil erosion, which increases water turbidity and reduces sunlight penetration. In addition, the use of untreated wastewater for irrigation represents a significant risk to water security and public health.

This manuscript was submitted on December 12, 2024. Accepted on March 06, 2025. And published on March 31, 2025.

This article was developed as part of the research entitled “Integration of knowledge for the conservation of water quality in the fish farming sector” of the PhD in Sustainable Development of the University of Manizales.

Kelly Johana Meléndez Segura, PhD student in Sustainable Development. E-mail: kmelendez105065@umanizales.edu.co

Henry Reyes Pineda, professor, University of Manizales, CIMAD Research Group. E-mail: hreyes@umanizales.edu.co

Jhon Fredy Betancur, professor, Universidad de Manizales, CIMAD Research Group. E-mail: jbetancur@umanizales.edu.co

All of the above leads to the quest for water quality conservation, which is crucial to ensure agricultural

sustainability, protect the health of populations and preserve ecosystems. Aquatic systems require clean water to maintain biodiversity and provide essential services such as climate regulation and natural water filtration. In addition, the availability of quality water plays a key role in climate change adaptation, as healthy ecosystems are more resilient to extreme weather events.

To mitigate the negative impact of agricultural activities on water quality, several sustainable management strategies have been developed. These include agroforestry, integrated pest management, efficient fertilizer use, wastewater recycling, and riparian zone protection. The implementation of these practices is essential to ensure access to clean water in the future and to promote the sustainability of agricultural and natural systems.

In this context, bibliometric analysis is presented as a valuable tool to consolidate and evaluate scientific knowledge in this area. By means of statistical methods, it makes it possible to analyze academic production, identify research trends and evaluate the impact of published studies. According to Gómez, Gutiérrez and Pinzón (2005), cited in Gaitán Sánchez et al. [1], this approach facilitates obtaining relevant data on citations, authors, institutions and countries with the highest scientific production in a specific topic.

Bibliometric indicators fulfill two key functions: descriptive, which characterizes the state of knowledge in each area, and evaluative, which assesses this state from a specific perspective (Gómez, Gutiérrez & Pinzón, 2005, cited in Gaitán Sánchez et al., [1]).

To develop this research, a search was conducted in the Web of Science (WoS) database using the search terms: "Agroecological Structure*" OR "Main Agroecological Structure" OR "Agroecological Planning" OR "Agroecological System*" OR "Agroforestry System*" (Topic) and "water quality" OR "water conservation". From this search, 61 references were obtained and analyzed using the WoS "Results Analyzer" tool, which made it possible to examine frequencies and generate relationship matrices, subsequently exported to Excel for detailed analysis. Additionally, the VOSviewer tool facilitated the graphical visualization of interaction networks between authors, countries and other key descriptors.

The search strategy was precisely designed to gather relevant scientific literature on the relationship between agroecological structures and water conservation and quality, to explore the next two research questions:

What has been the evolution of scientific knowledge on agroforestry systems and contributions to water quality and/or conservation?

What are the lines of research and future perspectives between agroforestry systems and water quality and/or conservation?

The objective of this research was to conduct a bibliometric and systematic analysis on water quality and/or conservation in agroecological structures, using the WoS database as the main source. To organize the documents obtained, the Tree of Science tool [2] was used, whose tree diagram categorizes the publications according to their relevance: the roots represent the classic documents, the trunk groups the structural studies, and the branches and leaves correspond to the most recent and emerging articles on the subject.

This study contributes to the understanding of the current state of research at the intersection between agroecology and water quality, providing a solid foundation for future lines of research and the development of sustainable management strategies in the agricultural sector.

Despite the increasing attention on agroecology and water management, the scientific literature still presents a significant gap in the detailed understanding of how agroecological structures can optimize water conservation at different spatial and temporal scales. In particular, there is a lack of studies that quantitatively and qualitatively integrate the impact of these structures on improving water quality and water efficiency in diverse agricultural systems. In addition, the interaction between agroecological practices and climate change in relation to water conservation is still a developing area of research.

This study is relevant both locally and globally, as challenges related to agricultural water conservation are common in multiple regions of the world. In a local context, the results can guide water management policies and sustainability strategies tailored to specific ecosystems, contributing to improve the resilience of agricultural communities. At the global level, bibliometric analysis has facilitated the exchange of knowledge between different regions and productive contexts.

II. Type of Study and Analysis Material

A. Type of study and analysis material

The bibliometric analysis was performed using data extracted from the Web of Science (WoS) database, chosen for being one of the most recognized sources for the evaluation of scientific production. This platform includes journals of great prestige and high visibility in multiple areas of knowledge, Archambault, (2009), cited in Gaitán Sánchez et al., [1]

The search was carried out using the following equation: "Agroecological Structure*" OR "Main Agroecological Structure" OR "Agroecological Planning" OR "Agroecological System*" OR "Agroforestry System*" (Topic) and "water quality 'OR 'water conservation" (Topic) and Article or Review Article (Document Types), taking into account only articles and review articles, in addition to a window of observation between 2014 and 2023 to broaden the panorama and analyze research and innovations.

B. Bibliometric Variables

In the initial phase, the following bibliometric indicators were defined: number of publications, countries, authors and their connections, academic institutions and total number of citations. Relationship and collaboration indicators were also considered, with the aim of creating thematic maps reflecting the interactions between authors and countries, as well as the co-occurrence of the selected keywords.

C. Bibliometric Data Collection and Analysis

Data collection was carried out using the established search equation, followed by downloading the records in plain text format from the Web of Science (WoS) database. To manage

bibliometric indicators, the WoS "Analyze Results" and VOSviewer tools were used, while frequency calculations and visual representations were performed using tables and graphs in Microsoft Excel.

D. Systematic Analysis of Collected Documents

With the 61 records obtained in WoS, metrics related to the degree of input, degree of output, and intermediation were evaluated, which made it possible to classify the research using the metaphor of trees [3]. From this analogy, three key categories emerge. The first is the "Root" (high centrality), which refers to classic literature or research that possesses dominant theoretical importance within the field of study. These publications are frequently cited, although not necessarily referenced by other authors [4]. The second category corresponds to the "Main Body" (high intermediation), which includes those articles that are not only cited, but also serve as references in works cited by others [5], this section constitutes a structured work that combines fundamental classical theory with current research. Last are the "Branches and Leaves" (high out-degree), which represent recent articles focused on citing other studies and reflecting current trends within the field's research framework. These publications, also called "perspectives," delineate emerging research fronts and articles [4]. This methodology has been previously validated and applied in previous studies [6] [7]. Likewise, the results extracted from the WoS database were systematized in an Excel spreadsheet, analyzing key information such as titles, abstracts, keywords, main findings and number of citations. Finally, with the information collected and its analysis, the main challenges to advance in the application and conservation of water quality in agroecological structures were identified.

E. Inclusion And Exclusion Criteria

In order to ensure accuracy, transparency and reproducibility in the selection of studies, this research adopted certain guidelines of the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) protocol [8]. The inclusion and exclusion criteria applied in the process are detailed below:

1. Inclusion criteria

Type of publication

Only research and review articles published in peer-reviewed scientific journals were selected, thus ensuring the validity and reliability of the data analyzed. Other forms of publication, such as conference proceedings, book chapters and gray literature, were discarded due to the absence of a rigorous peer review process.

Thematic relevance

Only studies that explicitly addressed the relationship between agroecological structures and water quality conservation were included. This selection criterion ensured that the literature analyzed was aligned with the research objectives, allowing for a focused and relevant analysis.

Language

Initially, no language restrictions were established in the search process. However, the 61 studies finally selected were published in English, since this language predominates in the international scientific literature.

Time frame

The review addressed studies published between 2014 and 2023, to include the most recent developments in water quality conservation and the role of major agroecological structures.

2. Exclusion criteria

Gray literature

Gray literature, such as reports from non-governmental organizations and conference proceedings, was discarded in order to prioritize studies published in peer-reviewed journals, thus ensuring scientific rigor.

Selection process

Study selection was carried out following the phases established in the PRISMA protocol, with the complementary incorporation of a bibliometric analysis through the Tree of Science platform. This integration made it possible to structure the selection process in a systematic and chronological manner, optimizing identification.

Identification

A search was carried out in the Web of Science database, applying the search equation specified above. Initially, a set of studies directly linked to synergies between pollinators and floral stripes was identified. Subsequently, the selected articles were subjected to a detailed analysis using the Tree of Science algorithm, which made it possible to evaluate their impact and relevance.

Detection

Tree of Science analyzed the initially identified studies together with their citations, generating a structured selection of key articles. These were classified into three levels: roots (seminal publications), trunks (seminal studies) and leaves (recent research). Subsequently, an independent review of titles and abstracts was conducted to discard those that did not meet the predisposed criteria.

Eligibility

Studies that passed the screening phase were subjected to a rigorous evaluation to verify their relevance and scientific quality. Additional exclusion criteria were applied to ensure that only the most relevant and methodologically sound papers were included in the final analysis. This process involved a detailed review of the objectives, methodological approaches and findings of each study, as well as confirmation of its publication.

Final inclusion

Research that met all the inclusion criteria and satisfactorily passed the previous phases were incorporated into the systematic and bibliometric analysis.

Minimum citation threshold

No minimum citation threshold was established as a requirement for the inclusion of studies in the systematic review. We sought to integrate both seminal research with a high impact on the scientific literature and recent studies that, although they have not yet accumulated a significant number of citations, represent advances and emerging trends in the field. During the bibliometric analysis, citation metrics were used to identify the most influential papers, classifying them within the Tree of Science (ToS) categories: roots (seminal publications), trunks (structural studies) and leaves (recent research).

III. RESULTS

A total of 61 documents were identified in this study, in which research articles predominated with 80.33% (49 articles) followed by subject reviews with 19.67% (12 articles). It is important to mention that the search period was from January 01, 2014, to December 31, 2023.

A. Production indicators.

Fig 1 shows the scientific production in the defined period, where a growing trend is evident in the dissemination of research specifically from the year 2019 to 2020 and 2022 to 2023, but at the same time a decline could be observed in the years 2017 and 2021; the year 2023 illustrates the highest number of publications reaching 11 articles being 18.033% of the reported publications. The scientific community has interest in the field of knowledge with an annual growth rate of 10.65%, from the first year of review in 2014 to 2023. The citations in the year 2014 was 1 being the lowest in the range of time taken, reaching the maximum in 2023 with a total of 364.

When analyzing the scientific production by country, the United States leads the scientific production with 15 publications, representing 24.6%, followed by China with 18%, and finally England, Greece and Spain with 6.6% each, Table 1 shows the global impact of academic production on research associated with water quality and/or conservation in agroecological structures.

In relation to the authors, 10 most representative authors

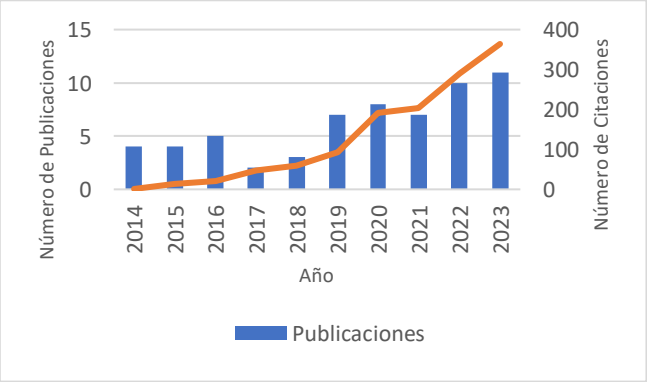


Fig. 1. Scientific production on water quality and/or conservation in agroecological structures and its products per year.

TABLE I
PRODUCTION OF ARTICLES BY COUNTRY

No	*País*	*Publicaciones*	*%*
1	USA	15	24.6%
2	CHINA	11	18.0%
3	FRANCE	8	13.1%
4	INDIA	8	13.1%
5	GERMANY	7	11.5%
6	BRAZIL	6	9.8%
7	COSTA RICA	4	6.6%
8	ENGLAND	4	6.6%
9	GREECE	4	6.6%
10	SPAIN	4	6.6%

TABLE II
MOST RELEVANT AUTHORS

No	*Autor*	*Publicaciones*	*Citaciones*	*Indice H*	*Universidad*
1	Chen, Hung-Chih	4	1810	25	Kunming University
2	Liu, Wenjie	4	2356	30	University of Chinese Academy of Sciences
3	Nettles, Jami	2	285	10	Weyerhaeuser Company
4	Tian, Shiying	2	649	11	North Carolina State University
5	Chescheir, George M.	2	1856	24	North Carolina State University
6	Jiang, Xiao-Jin	2	1277	20	Northeast Forestry University - China
7	Cacho, Julian F.	2	72	4	North Carolina State University
8	Zhu, Xiai	2	664	14	University of Chinese Academy of Sciences
9	Wu, Junen	2	844	15	University of Chinese Academy of Sciences
10	Youssef, M. A. S.	2	1344	19	National Authority for Remote Sensing & Space Sciences (NARSS)

were found Table II, who are categorized by the number of documents published in the database, Chen, Hung-Chih from Kunming University, who has had the most publications, followed by Liu, Wenjie, and Nettles, Jami from China University and Weyerhaeuser Company respectively, in addition, their H-index (H-index), which is used to describe the scientific output of researchers, is correlated [9], where Liu,

Wenjie and Chen, Hung-Chih were found with an index of 30 and 25 respectively.

The journal with the highest impact in the search is SUSTAINABILITY, in second place AGROFORESTRY SYSTEMS and AGRICULTURAL SYSTEMS in third place, the journals in this review are indexed in the databases and all are part of quartile 1. Within the top 10 and with more importance are journals from Switzerland in first place, the Netherlands in second and third place, and the United Kingdom in second place (Table III).

TABLE III
UNITS FOR MAGNETIC PROPERTIES

NO	REVISTAS	PUBLICACIONES	PORCENTAJE	CUARTIL	SJR	H-INDEX	PAÍS
1	SUSTAINABILITY	4	6.6%	Q1	0.67	169	Switzerland
2	AGROFORESTRY SYSTEMS	3	4.9%	Q1	0.51	92	Netherlands
3	AGRICULTURAL SYSTEMS	2	3.3%	Q1	1.59	134	United Kingdom
4	AGRICULTURE BASEL	2	3.3%	Q1	0.61	66	Switzerland
5	AGRICULTURE ECOSYSTEMS ENVIRONMENT	2	3.3%	Q1	1.74	212	Netherlands
6	ANIMALS	2	3.3%	Q1	0.7	75	Switzerland
7	CATENA	2	3.3%	Q1	1.5	164	Netherlands
8	GEODERMA	2	3.3%	Q1	1.76	203	Netherlands
9	LAND	2	3.3%	Q1	0.73	54	Switzerland
10	LAND DEGRADATION DEVELOPMENT	2	3.3%	Q1	1.16	105	United Kingdom

Figure 2 illustrates the four elements of relevance that are part of the bibliographic analysis, in the first box Fig. 2A is the network of collaboration between authors, each node representing an author. In this case, the authors included in the network are Cacho, Julian F, Chescheir, George M., Tian, Shiying, Nettles, Jami E., and Youssef, Mohamed A. The size of the nodes reflects the number of publications or collaborations that each author has in this specific network. All authors appear to have similar node size, indicating an equal contribution in terms of collaborations.

Lines connect the nodes, representing collaborations between authors, the thickness of the lines could indicate the frequency or strength of collaborations between authors. In this graph, all authors are connected to each other, suggesting a very cohesive collaborative team.

There are no clear subgroups in this network, indicating that all authors collaborate closely with each other. The distribution of connections is even, which may indicate that there is no "lead author" in the network, but a more horizontal collaborative approach.

In this way, it can be indicated that this collaborative network is typical of a well-integrated research team, where all members collaborate directly with each other. The lack of subgroups or clusters indicates that this team probably works on highly interconnected projects or on a single common project.

In the second (Fig. 2B) is the co-citation network between authors; the lines connecting the nodes represent co-citations, i.e., how many times have two authors been cited together in other papers. The thickness and color of the lines can indicate the strength of the co-citation relationship. Thicker or more intensely colored lines usually represent more frequent co-citation. Authors Jose, S, Nair, Pkr and Lal, r appear to be related through co-citations, although not directly between Jose, S and Lal, r. This could indicate that Nair, pkr acts as a bridge between Jose, S and Lal, r. The intensity of the connections suggests that these authors have been cited together on several occasions, which could indicate a scholarly collaboration or that

They work on similar topics that are co-cited together.

The country collaboration network Fig. 2C highlights the United States, Brazil, India, China, Germany and France, the size of the nodes could indicate the number of publications or

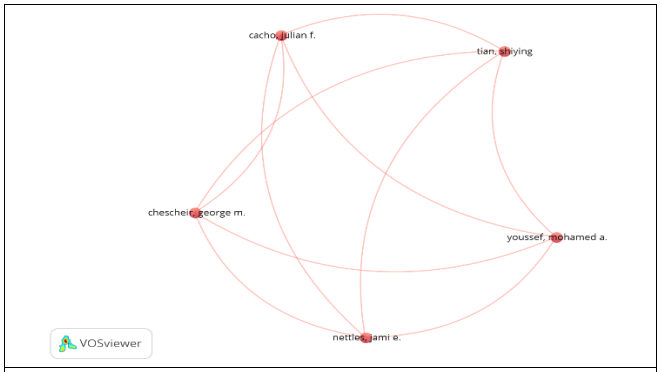


Fig. 2A Network of collaboration between authors

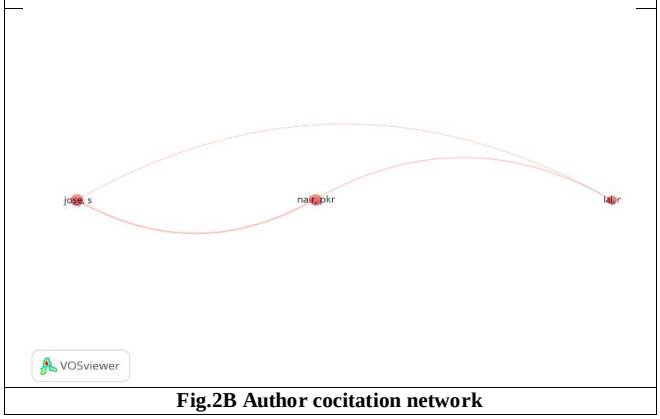


Fig.2B Author cocitation network

international collaborations in which each country is involved. In this case, the United States appears to be a central and largest node, suggesting a dominant role in collaborations.

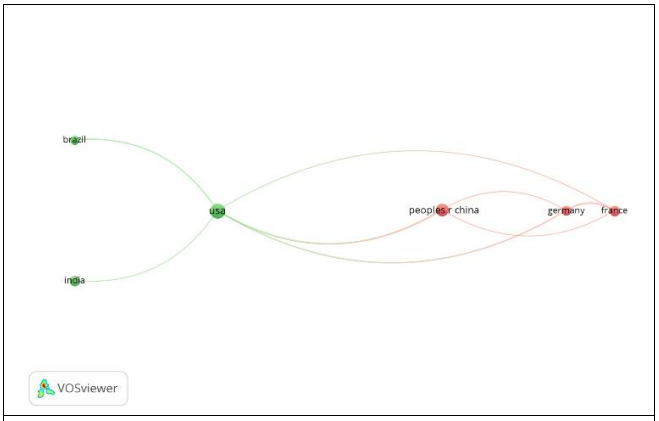


Fig 2C. Country collaboration network

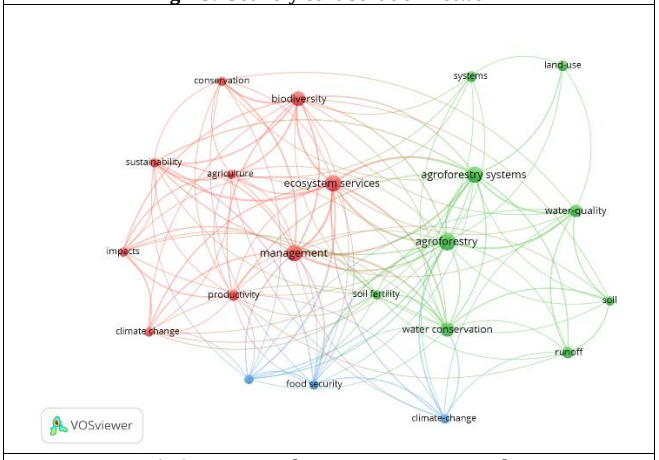


Fig 2D. Keyword co-occurrence network

Fig. 2: networks.

The lines connecting the nodes represent collaborations between countries, i.e., joint publications or projects. The United States is a central node collaborating with both countries such as Brazil and India on one side of the network, and China, Germany, and France on the other. This reflects the central role of the US in global scientific and academic collaboration. Brazil and India are more strongly connected to the U.S. but do not appear to have direct relationships with the other countries in the network. This could indicate that their international collaboration is mainly concentrated in the United States. China, Germany, and France are interconnected with each other, in addition to their connection to the U.S. This suggests an axis of collaboration between these countries, with China playing an important role within this sub-network. The absence of direct links between Brazil, India, and European countries or China could reflect that these countries' collaborations are mediated primarily through the United States. Finally, in the keyword co-occurrence network fig. 2D, the nodes and lines are colored to reflect different clusters or groups of keywords that are most closely related to each other, the red cluster focuses on themes of "biodiversity", "ecosystem services", "conservation", and "management". This suggests a group of keywords related to the management and conservation of biodiversity and ecosystem services. The green cluster groups words such as "agroforestry", "land-use", "soil fertility", and "water conservation". This cluster seems to be related to sustainable agriculture and land management, especially in the context of agroforestry. The blue cluster contains words such as "food security" and "climate change". This cluster seems to focus on food security and how it is affected by climate change.

The network shows a strong interaction between concepts related to ecosystem management, sustainable agricultural practices, and climate change impacts.

The "management" node acts as a central bridge, connecting several important themes. This suggests that management is a key concept linking diverse areas such as biodiversity conservation and land use sustainability. The presence of well-defined clusters indicates that although there is overlap between topics (e.g., the relationship between agroforestry and ecosystem services), there are also more specialized areas of research within the general field of environmental sustainability and management.

B. Network analysis

Through this analysis, the most relevant documents on the topic can be identified, and the metaphor of a scientific tree was used to select documents with the highest metrics for review and organization [2]. Five classic (root), five structural (trunk) eleven (branches), and eight recent (leaves) [8].

C. Classic documents (Root)

The research articles presented below, related to water quality and/or conservation and agroecological structures, are rooted in this literature review, standing out for their relevance in the field. In this sense, five (5) fundamental registers are analyzed which, as described above, present classic dispositions and dominances in the literature.

In this context, agroforestry systems constitute an innovative approach to achieve sustainable agriculture, allowing high crop yields and the protection of soil and water resources. In the present study, the efficiency of these systems in reducing contaminants in groundwater and surface water was determined. The results indicated an attenuation of nutrient leaching to groundwater of up to 97.7% and 90% for nitrogen and phosphorus, respectively, and up to 100% attenuation for both pollutants in surface runoff. In addition, several studies evidenced the capacity of agroforestry systems to reduce the presence of pesticides, with pollutant retention of up to 100% for various types of herbicides and fungicides, although only in runoff mitigation. However, insufficient research has yet been conducted to evaluate soil and groundwater protection against leaching of agrochemicals, especially pesticides. Therefore, further studies and policy implementation are required to maximize the practical benefits of these systems for agriculture, the environment, and ultimately human health and well-being [10].

According to Jose [11], in conventional farming systems, crops absorb less than half of the applied nitrogen and phosphorus. Therefore, excess fertilizer is removed through surface runoff or leached into the subsoil, thus contaminating water sources and decreasing their quality. In this context, agricultural surface runoff can lead to excess sediment, nutrients and pesticides in receiving water bodies, contributing to eutrophication in the Gulf of Mexico.

Based on this, agroforestry practices have proven to be an effective strategy for providing potable water. Among these, agroforestry systems include riparian buffers that contribute to the cleansing of runoff water by slowing its velocity, promoting infiltration, sediment deposition, and nutrient retention. A buffer of switchgrass (Panicum virgatum) and woody stem removed 20% more nutrients. In addition, trees with deep root systems can improve groundwater quality by acting as a "safety net", recycling nutrients through root turnover and litterfall, which improves the efficiency of nutrient use in the system. Studies have reported this mechanism in both tropical and temperate regions, suggesting that agroforestry systems could play a substantial role in mitigating water quality problems generated by intensive agricultural practices [11].

On the other hand, in the Brazilian semi-arid region, inadequate soil management practices have exacerbated erosive processes. In this context, agroforestry systems have been identified as a viable alternative to reduce water erosion. The evaluation of the impact of two agroforestry systems (one traditional and one intensive) in comparison with natural vegetation and a conventional agricultural system, revealed that agroforestry systems were more efficient in reducing water erosion, reducing contamination and loss of water quality. Therefore, their adoption is recommended as a sustainable technical alternative for food production in the region [12].

Likewise, the integration of trees into pastures has proven to be an effective strategy to mitigate water pollution. Studies comparing three types of pastures - one without trees (Paspalum notatum), a pasture under 20-year-old pines (Pinus elliotti) and a pasture of native vegetation under pines - concluded that silvopastoral systems allow a more efficient uptake of nutrients, especially phosphorus, compared to pastures without trees. In addition, the capacity of soils under these systems to receive additional phosphorus is greater, thus reducing nutrient leaching to surface water and mitigating water pollution [13].

Finally, the adaptation of agricultural systems to climate change is crucial, given that this phenomenon can generate negative impacts on agricultural production. According to Lin [14], the resilience of agricultural systems can be improved through greater crop diversification. However, there are barriers such as economic incentives to produce specific crops, the focus on biotechnological strategies and the perception that monocultures are more productive. In this regard, crop and landscape simulation models can help farmers find optimal strategies to maintain production and profitability. Understanding the potential for greater diversity within agricultural systems is essential for coping with climate variability. By adopting practices that foster ecosystem services for pest control, disease and climate resilience, farmers can reduce the risk of production losses and strengthen their capacity to adapt to environmental changes.

In summary, this review has addressed, in the first instance, the definition of agroforestry systems, followed by the problems derived from conventional agriculture and, finally, the specific benefits of agroforestry practices, highlighting their relevance in climate change adaptation and water conservation.

D. Structural documents (Trunk)

Within the structural documents of the knowledge tree, key trends in research development are identified, particularly in the following areas:

Agroforestry systems, which combine trees with crops or pastures, have been widely implemented in temperate and tropical regions due to their effectiveness in reducing water, soil and nutrient loss, as well as mitigating water pollution generated by agricultural activities. However, despite their widespread use, there are still few scientific reviews that comprehensively evaluate their efficiency and scope, considering factors such as soil type, management practices, climatic conditions and the hydrological processes involved. Therefore, it is essential to develop systematic studies that allow the generalization of agroforestry design and its adaptability in regions with similar climatic, geographic, ecological and socioeconomic characteristics worldwide [15].

The progressive deterioration of surface and groundwater quality in recent decades has increased interest in identifying sources of contamination. Agricultural intensification, driven by the need for high quality crops and high yields, has led to excessive use of fertilizers and pesticides, resulting in negative impacts on the environment, especially on soil and water bodies. A study conducted in experimental agricultural fields in the Mediterranean, in which N, P and K ions, as well as the herbicides pendimethalin, its metabolite M455H001 and s-metolachlor, together with the insecticide chlorpyrifos, were analyzed, showed that agroforestry systems, such as corn-poplar and potato-poplar associations, can significantly decrease water pollution. In particular, tree roots have the capacity to absorb excess agrochemicals, preventing them from leaching into groundwater by leaching or being transported to surface water by runoff [16].

On the other hand, multiple studies have shown that land use patterns significantly influence water infiltration capacity. Increasing infiltration and reducing runoff are fundamental aspects for soil and water resource conservation, especially in semi-arid environments. In this regard, research conducted on the Loess Plateau in China compared three planting systems over 11 years and concluded that agroforestry systems significantly improve soil infiltration and soil sustainability, particularly in semi-arid areas. These findings offer new insights into the applicability of agroforestry in regions with similar conditions around the world [17].

In recent decades, it has become evident that agroforestry not only contributes to the protection of natural resources, but also allows maintaining or increasing agricultural productivity. In a study developed in Xishuangbanna, southwest China, Wu JN [18], evaluated a rubber-based agroforestry system, finding that intercropping with legumes significantly improved water use efficiency and tree tolerance to adverse conditions. The rubber trees showed more stable physiological indices and higher water efficiency, suggesting that this strategy is highly beneficial for water conservation.

Finally, Panwar's [19], study examined the effectiveness of agroforestry systems for soil and water conservation on sloping land in the Shivalik region of India. By combining silvihortipastoral practices and the implementation of water harvesting structures, a significant reduction in soil loss was

achieved, as well as an increase in runoff retention. These results underscore the potential of agroforestry as a viable strategy for resource conservation in sloping areas, and its integration into land-use planning is recommended as an effective alternative for the development of sustainable agriculture in challenging environments.

E. Recent Perspectives (Branches)

In the review conducted, three branches were identified that encompass specific sub-areas within the knowledge domain analyzed. These branches encapsulate articles focused on diverse topics derived from cluster analysis and allow the identification of relevant trends within the field of study [2]. One of these fundamental perspectives is the interconnection between water resources and energy and biomass production, a topic of growing relevance in the context of environmental sustainability and conservation.

Perspective 1. Water and energy production.

The relationship between water resources and energy generation is crucial for sustainable development. In this context, property rights over natural resources have been a key legislative tool to promote their responsible use and conservation globally. However, the incorporation of ecological property rights could significantly modify farmers' investment behavior in forests and water resources. This approach could strengthen forest protection, optimize water conservation and, consequently, improve water security in urban areas [20].

The progressive depletion of global land and groundwater reserves, resulting from prolonged overexploitation, underscores the need for effective management of these vital resources. The growing demand for water and soil due to accelerated population growth emphasizes the urgency of maintaining their integrity without compromising productivity. In this sense, agroforestry emerges as a promising strategy, since the integration of trees and shrubs into agricultural practices not only improves soil fertility and reduces erosion, but also optimizes water retention and conservation, favoring the soil's absorption capacity and hydraulic properties.

An emblematic example of the interdependence between water and energy is the Three Gorges Dam (TGD) in China, one of the largest hydropower infrastructures in the world. This dam has generated important benefits, such as drinking water supply, irrigation, power generation and flood control. However, it has also caused adverse environmental impacts, such as eutrophication in secondary rivers due to the accumulation of nutrients in the impounded water. This highlights the need for accurate and controlled water management to mitigate negative effects, such as algal blooms, which requires a detailed understanding of the interactions between main streams and their tributaries [21].

Also, large-scale bioenergy production significantly affects the hydrological cycle. According to Watkins et al. [22], it influences processes such as canopy interception, evapotranspiration, infiltration, runoff and aquifer recharge. These impacts vary according to the type of biomass, soil characteristics, agricultural practices and hydroclimatic conditions. In addition, the interaction between bioenergy and water management is intrinsically linked to land use, water availability and competing demands for this resource in watersheds. Therefore, policies related to water and bioenergy should be evaluated not only in terms of efficiency and effectiveness, but also considering their socioeconomic impacts and their effect on vulnerable communities.

Thus, an integrated and multidisciplinary approach is essential to ensure equitable and sustainable management of water resources in the context of energy and biomass production. Only through coordinated strategies will it be possible to avoid exacerbating water conflicts and ensure the long-term viability of these production systems.

Perspective 2. Integration of ecosystem services and ecological modernization in agroforestry systems.

Contemporary agriculture is facing increasing criticism due to its predominantly productivist approach, which often neglects the supporting and regulating services provided by ecosystems. In this context, agroforestry and ecological modernization strategies emerge as key alternatives to promote sustainability and improve human quality of life through the provision of multiple ecosystem services (ES) [23] [24].

Agroforestry systems can provide ecosystem benefits that contribute to both farm sustainability and human well-being. Notaro et al. [23], studied four ecosystem services in coffee agroforestry systems in Nicaragua: coffee production, water quality preservation, carbon sequestration, and biodiversity conservation. Their findings revealed that carbon sequestration depended more on the presence of large trees than on coffee yield, while tree biodiversity favored productivity up to a certain threshold, after which yield decreased. This underscores the importance of a moderate density of shade trees to optimize both production and SE provision.

Water quality and conservation are fundamental elements in agroforestry systems. The preservation of water sources and the optimization of water use guarantee not only the sustainability of the crop, but also the resilience of the ecosystem in the face of climatic changes. Agroforestry, by promoting vegetation cover and water infiltration into the soil, helps to regulate the hydrological cycle and reduce erosion, ensuring long-term water supply.

Additionally, Padovan et al. [25] addressed the impact of land pressure and the need to maximize income, which forces smallholders to cultivate in suboptimal areas. In a study in Nicaragua, they analyzed water use efficiency in agroforestry systems versus full-sun systems, demonstrating that agroforestry allows more efficient water use under adverse conditions. Most of the soil water was used for coffee transpiration rather than being lost to evaporation or consumed by shade trees. Two shade tree species, Tabebuia rosea and Simarouba glauca, were compared, providing valuable information for the selection of species that optimize water use and improve the resilience of the agroforestry system to water

variability.

Therefore from an ecological modernization perspective, Duru [24] identifies two key approaches: (1) efficiency substitution agriculture, which seeks to optimize input use and minimize environmental impacts, and (2) biodiversity-based agriculture, which develops SE through biological diversification. To facilitate this transition, Duru proposes a transdisciplinary conceptual and methodological framework that involves agronomic innovations and coordination among actors in the supply chain and natural resource management. This approach requires technological, social, economic and institutional changes, enabling local actors to design adaptive action plans that foster the diversification and sustainability of agricultural systems.

In this way it can be emphasized that the integration of ecosystem services in coffee agroforestry systems not only improves agricultural productivity and sustainability but also contributes significantly to environmental conservation. Water quality and conservation play a crucial role in the stability of these systems, ensuring both productivity and ecosystem resilience. To maximize these benefits, it is essential to maintain an adequate density of shade trees and select species that optimize water use and biodiversity.

In addition, the transition to biodiversity-based agriculture requires a holistic approach involving innovations at multiple levels and close collaboration between the different stakeholders involved. Implementing these strategies can lead to agroecological intensification that balances agricultural production with the conservation of natural resources, thus ensuring the long-term sustainability of agroforestry systems.

Perspective 3: Heavy metals and emerging contaminants in agroforestry systems.

Mercury (Hg) is a highly toxic global pollutant that persists in aquatic ecosystems Li et al., [26] highlights how mercury, used in large quantities during the Manhattan project in Oak Ridge, Tennessee, still contaminates the surrounding watershed. Soil erosion and rainfall-runoff events contribute to mercury transport from nonpoint sources into aquatic ecosystems. Proper site management, such as improving vegetative cover and reducing slopes, is critical to reducing this mercury transport, where low plants play a crucial role in phytostabilizing the pollutant.

Agroforestry has been shown to be an effective strategy to control soil erosion and improve agricultural sustainability in semi-arid areas. Huang et al., [27] discusses how agroforestry systems influence soil water storage (SWS) and how the proximity of forest plantations can affect this storage, especially at the afforestation-cropland interface (ACI). In particular, species such as S. Japonica are recommended for their lower impact on soil water availability, making them a valuable option for improving the ecological environment and long-term sustainability in these regions.

Zhang et al., [28] stress the importance of “source-sink” landscape pattern analysis for nonpoint source pollution management. Remote sensing is presented as an effective technique to study these patterns and their relationship with water quality, despite technical challenges. Advances in this area have made it possible to optimize landscape management to reduce pollution, which is crucial for the protection of water resources and the construction of ecological security patterns.

Béliveau et al., [29] highlights the negative effects of soil erosion in the Amazon, where traditional agriculture has contributed significantly to soil degradation and the release of natural mercury into water bodies. Agroforestry practices in the Brazilian Amazon have proven to be effective in reducing both soil erosion and mercury mobility, making them a sustainable solution for agricultural management and environmental conservation. These practices not only conserve soil, but also reduce pollution, making agroforestry an essential tool for the protection of Amazonian ecosystems.

Finally, Pascual Aguilar et al., [30] addresses the emerging pollution problem in Mediterranean wetlands, such as L'Albufera de Valencia, where human impact and socioeconomic development have generated a high concentration of pollutants, including pharmaceuticals and pesticides. The research highlights the urgent need to implement measures to mitigate this pollution and protect these ecosystems of great ecological value, thus ensuring water sustainability and ecosystem health for future generations.

F. Leaves

Agroforestry systems (AFS) are consolidated as a key strategy for agricultural sustainability and conservation of natural resources, especially in terms of water and soil quality. According to the review conducted by François et al. [31], the implementation of AFS contributes significantly to the reduction of nutrient losses, such as nitrogen (N) and phosphorus (P), whose accumulation derived from the intensive use of chemical fertilizers has generated serious water pollution problems. In addition, these systems favor the improvement of the physical, chemical and biological properties of the soil, promoting its water retention capacity and reducing erosion. Likewise, PBS have demonstrated a remarkable potential in the elimination of trace metals such as cadmium, aluminum and mercury in contaminated soils, strengthening their role as an integral agroecological tool. However, factors such as surface geology, slope gradient and topographic conditions can negatively influence water quality in watersheds, underscoring the need for strategic planning in their implementation.

In this context, Ntawuruhunga et al. [32] emphasize the importance of climate-smart agroforestry (CSAF) as a comprehensive solution for climate change mitigation and adaptation, particularly in rural areas. This approach combines trees, crops and livestock in sustainable production systems, optimizing water use and contributing to water security. However, the adoption of CSAF still faces significant challenges, especially among smallholder farmers, due to lack of knowledge and technical support. Therefore, the study highlights the need to strengthen evidence-based public policies, foster public-private partnerships and promote multidimensional initiatives that facilitate their effective implementation, thus ensuring their positive impact on food security, poverty reduction and the resilience of water

ecosystems.

In Indonesia, Sudomo et al. [33] highlight agroforestry as a crucial mechanism for improving food security and access to water, especially among smallholder farmers. By adapting agroforestry practices to local conditions and market needs, sustainable incomes are generated that enable rural communities to strengthen their resilience to climate change. In addition, crop diversification in these systems improves water infiltration into the soil and reduces surface runoff, reducing pollution of water bodies and promoting the resilience of water ecosystems.

Srinivasarao et al. [34] address soil degradation in India as a critical problem that threatens water security and agricultural sustainability. The study highlights the need to adopt integrated soil and water conservation technologies aimed at minimizing erosion and fostering the development of resilient climate change communities. To this end, it proposes the strengthening of community capacities and the creation of local institutions to manage and maintain these conservation structures in the long term, thus ensuring the sustainability of water resources in vulnerable rural environments.

From an environmental governance perspective, Li [20] analyzes the impact of ecological property rights on the promotion of sustainable agroforestry practices in the Heihe Reservoir region, Shaanxi, China. The implementation of these rights has proven to be effective in reducing soil erosion and improving water management, in turn facilitating food security and increased income for local communities. However, the study points out that to maximize its impact, it is necessary to clarify the allocation of rights and strengthen government support through financial incentives and technical assistance.

In the livestock area, Pinheiro Machado Filho et al. [35] present the Voisin rational grazing system (VRG) as a sustainable agroecological alternative that optimizes animal productivity while improving soil and water quality. This model integrates multi-species grasslands with SAF, promoting ecosystem regeneration, carbon absorption and water retention in the soil, which contributes to the protection of water sources and biodiversity. Despite its high potential, the implementation of GBV requires a long-term vision and a comprehensive approach to overcome the associated technical and socioeconomic barriers.

On the other hand, Platis et al. [36] emphasize the need to reduce greenhouse gas emissions in agriculture, aligning agroforestry systems with the objectives of the Paris Agreement on climate change. These systems, in addition to minimizing the use of non-renewable energy, improve water use efficiency and significantly reduce the water and carbon footprint of agricultural production. The adoption of methodologies such as life cycle assessment is crucial to measure and mitigate the environmental impact of these systems, strengthening the long-term resilience of agroecosystems.

Finally, Bardule et al. [37] analyze the effectiveness of juvenile hybrid poplar plantations in agroforestry systems in reducing leaching of nutrients such as nitrogen, phosphorus, and potassium in the Baltic Sea region. Despite the use of fertilizers, a significant decrease in pollution of water bodies was evidenced, reaffirming the potential of PBS to improve water quality and promote sustainable agriculture.

Taken together, this research demonstrates that agroforestry systems represent a viable and effective alternative for water quality conservation in agroecological environments. However, their success depends on a strategic implementation that considers ecological, social and economic factors, as well as the strengthening of public policies and governance mechanisms that facilitate their adoption and long-term sustainability.

IV. Discussion

A review of 61 records obtained in WoS was carried out using the metaphor of trees [3]. In agroecological systems, water quality and conservation are fundamental, given that these systems seek agricultural sustainability through the integration of practices that respect and preserve natural resources. Agroforestry systems, which combine trees with crops or pastures, stand out as a key strategy to mitigate the negative impacts of conventional agriculture, such as groundwater and surface water contamination due to the excessive use of fertilizers and pesticides.

A literature review shows that agroforestry systems can reduce nutrient leaching to groundwater by up to 97.7% for nitrogen and 90% for phosphorus and can remove up to 100% of these pollutants in surface runoff [10]. In addition, these systems offer a solution for pesticide depletion, protecting vulnerable water bodies, although more research is needed to fully address agrochemical leaching into the soil.

The adoption of agroforestry practices has also proven to be effective in semi-arid regions, as in the case of Brazil, where they have significantly reduced water erosion, a critical factor in water pollution and loss of water quality [12]. On the other hand, the integration of trees in pastures can prevent the loss of nutrients to water bodies, thus improving surface and groundwater quality [13].

It is also crucial to consider the adaptation of these systems to climate change. Crop diversification, a practice promoted within agroforestry, can increase the resilience of agricultural systems to climate variability, reducing the risk of contamination and deterioration of water resources [14].

In systems where water is a limited resource, fish farming can be integrated to make the most of available water, using the same resource for multiple purposes: fish farming, crop irrigation, and wetland maintenance. By integrating fish farming with other agricultural activities, environmental impacts, such as eutrophication of water bodies due to nutrient runoff, can be reduced, as well-designed systems can recycle these nutrients rather than allowing them to pollute rivers and lakes.

Thus, agroforestry systems have multiple benefits for water quality and conservation in agroecological systems, acting as a natural barrier that reduces pollution, improves water infiltration and reduces soil erosion.

While interest in the relationship between agroforestry systems and water quality has increased, knowledge gaps persist that require additional research. In particular, limitations

have been identified in quantifying the long-term impacts of these systems on soil hydrodynamics and the composition of microbial communities involved in nutrient recycling [17]. Recent studies have pointed out the need to evaluate the capacity of agroforestry systems for biofiltration of emerging pollutants such as pesticides and heavy metals, as well as their impact on improving groundwater quality [27].

On the other hand, variability in the design and management of agroforestry systems poses a challenge for the generalization of their hydrological effects. There is a need to develop comprehensive models to more accurately predict the water efficiency of different agroforestry designs under climate change scenarios, especially in regions vulnerable to aridity [18].

The results of this study present significant implications for the design of public policies and sustainable water management strategies in agricultural contexts. The incorporation of agroforestry systems in territorial planning could improve water security and climate resilience of rural communities, reducing dependence on conventional water sources and mitigating the effects of climate variability [15]. It has been shown that these systems can reduce irrigation water demand by up to 30% by optimizing infiltration and soil moisture retention [13].

However, the transition to these systems requires the establishment of economic incentives, the strengthening of technical training and the development of regulatory frameworks that promote their large-scale implementation [19]. In this sense, payment for environmental services (PES) programs have shown to be an effective tool to incentivize the adoption of sustainable agroforestry practices in various regions of the world.

To strengthen the knowledge base on the relationship between agroforestry systems and water quality, the development of interdisciplinary studies that integrate ecohydrology, biogeochemistry and hydrological process modeling approaches is recommended [23]. In addition, the application of emerging technologies, such as satellite remote sensing and remote sensing water quality monitoring, would allow a more accurate assessment of the dynamics of these systems at different spatial and temporal scales [21]. Long-term studies that analyze the impact of ecological succession on the ecosystem services of agroforestry systems and their contribution to water resilience are required.

On the other hand, it is essential to foster international collaborations that allow the comparison of different geographical and socioeconomic contexts, facilitating the design of strategies adapted to local needs [24]. The integration of agroforestry systems into local and global water governance frameworks should be a priority approach to ensure their effective and sustainable implementation.

V. Conclusions

The bibliometric and systematic study has identified key trends in scientific production on water conservation and quality in agroecological structures. A sustained growth in the amount of research published in the last decade was evidenced, with a predominance of studies coming from the United States and China, suggesting a strong interest in the application of agroforestry systems as a mitigation and adaptation strategy in the face of global environmental challenges.

Results confirm that agroforestry systems play a determining role in improving water quality, reducing pollution by nutrient and agrochemical leaching by up to 97.7% and favoring water retention in soils [10]. In addition, these systems contribute to the regulation of the hydrological cycle, decreasing erosion and improving water infiltration by up to 60% compared to conventional systems. These advantages are crucial in the context of climate change, where water conservation and ecosystem resilience are critical for agricultural sustainability.

Despite these findings, knowledge gaps persist in quantifying the long-term impacts of these systems in different soil types and climates, as well as their effectiveness in biofiltration of emerging pollutants [27]. The development of more accurate hydrological models and the strengthening of interdisciplinary research integrating ecohydrology, biogeochemistry, and remote sensing approaches are recommended to assess the water efficiency of these systems under different environmental scenarios [21].

Thus, it is crucial that policy makers recognize the potential of agroforestry systems in water conservation and encourage their adoption through economic incentives and payment for environmental services programs. The integration of these systems into water governance frameworks will contribute to long-term water security and sustainability of agricultural production in regions vulnerable to water stress.

This study reinforces the importance of agroforestry systems in sustainable water management, highlighting both their potential and the challenges that remain in their implementation. It calls on the scientific community and policy makers to promote applied research and the development of innovative strategies to maximize the benefits of these systems for water conservation. It is recommended to strengthen the generation of scientific evidence on the long-term effects of agroforestry systems on water quality, mitigation of diffuse pollution and water security in contexts of high climate vulnerability.

Finally, given that climate change will continue to exert increasing pressure on global water resources, agroforestry systems represent a viable alternative to improve ecosystem resilience and ensure the sustainability of agricultural production in the future.

References

[1] J. Gaitan Sanchez, E. J. Corpas Iguarán and G. Restrepo Franco, "Tecnología Biofloc: Perspectivas de Aplicación en Procesos de Producción Piscícola Sostenible," *Revista lasallista de investigacion*, vol. 19(1) https://doi.org/10.22507/rli.v19n1a10, p. 165–181, 2022.

[2] M. Zuluaga, S. Robledo, , O. Arbelaez-Echeverri, G. A. Osorio-Zuluaga and N. Duque-Méndez, "Tree of Science - ToS: A web-based tool for scientific literature recommendation.," *Search less, research more! Issues in science and technology librarianship, 100.* , vol. https://doi.org/10.29173/istl2696, 2022.

[3] D. S. Valencia-Hernandez, S. Robledo, R. Pinilla, N. D. Duque-Méndez and G. Olivar-Tost , "SAP algorithm for citation analysis: An improvement to tree of Science," *Ingeniería e Investigación,* no. 40(1) https://doi.org/10.15446/ing.investig.v40n1.77718, p. 45–49. , 2020.

[4] W. D. Wallis, " A beginner's guide to graph theory," *Birkhäuser Boston* , Vols. https://doi.org/10.1007/978-0-8176-4580-9, 2007.

[5] J. Zhang and Y. Luo, "Degree centrality, betweenness centrality, and closeness centrality in social network.," *Proceedings of the 2017 2nd International Conference on Modelling, Simulation and Applied Mathematics (MSAM2017),* pp. https://doi.org/10.2991/msam-17.2017.68, 2017.

[6] S. Buitrago, P. L. Duque and S. Robledo, "Branding Corporativo: una revisión bibliográfica," *Económicas CUC, 41(1),* vol. https://doi.org/10.17981/econcuc.41.1.2020.org.1, p. 143–162, 2019.

[7] G. Torres, S. Robledo and S. Rojas Berrío, "Market orientation: importance, evolution, and emerging approaches using scientometric analysis. .," *Criterio Libre, 19(35)* , Vols. https://doi.org/10.18041/1900-0642/criteriolibre.2021v19n35.8371, p. 326–340, 2022.

[8] M. Page, J. McKenzie, P. Bossuyt and et al, "La declaración PRISMA 2020: una guía actualizada para la presentación de informes sobre revisiones sistemáticas ". BMJ 372 : n71. https://doi.org/10.1136/bmj.n71 .," 2021.

[9] . J. E. Hirsch, "An index to quantify an individual's scientific research output," *Proceedings of the National Academy of Sciences of the United States of America 102(46),* vol. https://doi.org/10.1073/pnas.0507655102, p. 16569–16572. , 2005.

[10] G. Pavlidis and V. A. Tsihrintzis, "Environmental benefits and control of pollution to surface water and groundwater by agroforestry systems: A review. ," *Water Resources Management, 32(1), https://doi.org/10.1007/s11269-017-1805-4,* p. 1–29, 2018.

[11] S. Jose, "Agroforestry for ecosystem services and environmental benefits: an overview.," *Agroforestry Systems, 76(1), https://doi.org/10.1007/s10457-009-9229-7,* p. 1–10. , 2009.

[12] M. I. De Aguiar, S. M. F. Maia, F. A. d. S. Xavier, F. A. , E. De Sá Mendonça, J. A. A. Filho and T. S. De Oliveira, " Sediment, nutrient and water losses by water erosion under agroforestry systems in the semi-arid region in northeastern Brazil.," *Agroforestry Systems, 79(3), https://doi.org/10.1007/s10457-010-9310-2,* p. 277–289, 2010.

[13] V. D. Nair, P. K. R. Nair, R. S. Kalmbacher and I. V. Ezenwa, "Reducing nutrient loss from farms through silvopastoral practices in coarse-textured soils of Florida, USA.," *Ecological Engineering, 29(2), https://doi.org/10.1016/j.ecoleng.2006.07.003,* p. 192–199, 2007.

[14] B. B. Lin, " Resilience in agriculture through crop diversification: Adaptive management for environmental change.," *Bioscience, 61(3), https://doi.org/10.1525/bio.2011.61.3.4* , p. 183–193, 2011.

[15] X. Zhu, W. Liu, J. Chen, L. A. Bruijnzeel, Z. Mao, X. Yang, R. Cardinael, F. R. Meng, R. C. Sidle, S. Seitz, V. D. Nair, K. Nanko, X. Zou, C. Chen and X. J. Jiang, "Reductions in water, soil and nutrient losses and pesticide pollution in agroforestry practices: a review of evidence and processes.," *Plant and Soil, 453 (1–2), https://doi.org/10.1007/s11104-019-04377-3,* p. 45–86. , 2020.

[16] G. Pavlidis, H. Karasali and V. A. Tsihrintzis, "Pesticide and fertilizer pollution reduction in two alley cropping agroforestry cultivating systems.," *Water, Air, and Soil Pollution, 231(5). https://doi.org/10.1007/s11270-020-04590-2,* 2020.

[17] L. Wang, C. Zhong, P. Gao, W. Xi and S. Zhang, "Soil infiltration characteristics in agroforestry systems and their relationships with the temporal distribution of rainfall on the loess plateau in China.," *PloS One, 10(4), e0124767. https://doi.org/10.1371/journal.pone.0124767,* 2015.

[18] J. Wu, W. Liu and C. Chen, "Below-ground interspecific competition for water in a rubber agroforestry system may enhance water utilization in plants.," *Scientific Reports 6(1). https://doi.org/10.1038/srep19502,* 2016.

[19] P. Panwar, "Toposequential agroforestry based land use system for soil and water conservation in sloping lands Range Manag Agrofor.," 2018.

[20] M. Li, A. Sarkar, Y. Wang, A. Khairul Hasan and Q. Meng, "Evaluating the impact of ecological property rights to trigger farmers' investment behavior—an example of confluence area of Heihe Reservoir, Shaanxi, China.," *Land, 11(3), 320. https://doi.org/10.3390/land11030320.,* 2022.

[21] Y. Zhao, B. Zheng, L. Wang, Y. Qin, H. Li and W. Cao , "Characterization of mixing processes in the confluence zone between the three gorges reservoir mainstream and the daning river using stable isotope analysis.," *Environmental Science & Technology, 50(18), https://doi.org/10.1021/acs.est.5b01132,* p. 9907–9914., 2016.

[22] D. W. Watkins, M. M. G. A. Jr de Moraes, H. Asbjornsen, A. S. Mayer, J. Licata, J. G. Lopez, T. G. Pypker, V. G. Molina, G. F. Marques, A. C. G. Carneiro, H. M. Nuñez, H. Önal and B. Da Nobrega Germano, "Bioenergy development policy and practice must recognize potential hydrologic impacts: Lessons from the Americas.," *Environmental Management, 56(6),*

https://doi.org/10.1007/s00267-015-0460-x, p. 1295–1314, 2015.

[23] M. Notaro, C. Gary, J. F. Le Coq, A. Metay and B. Rapidel, "How to increase the joint provision of ecosystem services by agricultural systems. Evidence from coffee-based agroforestry systems.," *Agricultural Systems, 196(103332), 103332. https://doi.org/10.1016/j.agsy.2021.103332*, 2022.

[24] M. Duru, O. Therond and M. Fares, "Designing agroecological transitions; A review.," *Agronomy for Sustainable Development, 35(4), https://doi.org/10.1007/s13593-015-0318-x*, p. 1237–1257, 2015.

[25] M. P. Padovan, R. M. Brook, M. Barrios, J. B. Cruz-Castillo , S. J. Vilchez-Mendoza, A. N. Costa and B. Rapidel, "Water loss by transpiration and soil evaporation in coffee shaded by Tabebuia rosea Bertol . and Simarouba glauca dc. compared to unshaded coffee in sub-optimal environmental conditions.," *Agricultural and Forest Meteorology, 248, https://doi.org/10.1016/j.agrformet.2017.08.036.*, p. 1–14, 2018.

[26] J. Li, L. Wu, L. Chen, J. Zhang, Z. Shi, H. Ling, C. Cheng, H. Wu, A. D. Butler, Q. Zhang, Z. Arslan, E. M. Pierce, Y. Su and F. X. Han, "Effects of slopes, rainfall intensity and grass cover on runoff loss of mercury from floodplain soil in Oak Ridge TN: A laboratory pilot study.," *Geoderma, 441(116750), 116750. https://doi.org/10.1016/j.geoderma.2023.116750*, 2024.

[27] Z. Y. W.-J. L. Y. S. W. L.-V. M. &. W. G.-L. Huang, Z. Huang, W. J. Yang, Y. Liu, W. Shen, M. López-Vicente and G. L. Wu, "Belowground soil water response in the afforestation-cropland interface under semi-arid conditions," *Catena, 193(104660), 104660. https://doi.org/10.1016/j.catena.2020.104660*, 2020.

[28] X. Zhang, Q. Y. Wu, J. T. Cui, Y. Q. Liu and W. S. Wang, ""“Source–sink” landscape pattern analysis of nonpoint source pollution using remote sensing techniques.," *International Journal of Environmental Science and Technology: IJEST, 15(10), 2253–2268. https://doi.org/10.1007/s13762-018-1683-1*, 2018.

[29] A. Béliveau, M. Lucotte, R. Davidson, S. Paquet, F. Mertens, C. J. Passos and C. A. Romana, "Reduction of soil erosion and mercury losses in agroforestry systems compared to forests and cultivated fields in the Brazilian Amazon.," *Journal of Environmental Management, 203, 522–532. https://doi.org/10.1016/j.jenvman.2017.07.037*, 2017.

[30] J. A. Pascual Aguilar, V. Andreu, P. Vázquez and Y. Picó, "Presence and spatial distribution of emerging contaminants (drugs of abuse) in protected agroecological systems (L’Albufera de Valencia Coastal Wetland, Spain).," *Environmental Earth Sciences, 71(1), 31–37. https://doi.org/10.1007/s12665-013-2688-1*, 2014.

[31] M. François, M. C. G. Pontes, A. Lima da Silva and E. Mariano-Neto, "Impacts of cacao agroforestry systems on climate change, soil conservation, and water resources: a review.," *Water Policy, 25(6), 564–581. https://doi.org/10.2166/wp.2023.164*, 2023.

[32] D. Ntawuruhunga, E. E. Ngowi, H. O. Mangi, R. J. Salanga and K. M. Shikuku, "Climate-smart agroforestry systems and practices: A systematic review of what works, what doesn’t work, and why.," *Forest Policy and Economics, 150(102937), 102937. https://doi.org/10.1016/j.forpol.2023.102937*, 2023.

[33] A. Sudomo, B. Leksono, H. L. Tata, A. A. Rahayu, A. Umroni, H. Rianawati, Asmaliyah, Krisnawati, A. Setyayudi, M. M. B. Utomo, L. A. G. Pieter, A. Wresta, Y. Indrajaya, S. A. Rahman and H. Baral, "Can agroforestry contribute to food and livelihood security for Indonesia’s smallholders in the climate change era?," *Agriculture, 13(10), 1896. https://doi.org/10.3390/agriculture13101896*, 2023.

[34] C. Srinivasarao, V. N. S. P. Jasti, V. R. Kondru, V. S. K. Bathineni, R. Mudigiri, G. V. Venati, P. Priyadarshini, P. C. Abhilash and S. Chaudhari, "Land and water conservation technologies for building carbon positive villages in India.," *Land Degradation and Development, 33(3), 395–412. https://doi.org/10.1002/ldr.4160*, 2022.

[35] L. C. Pinheiro Machado Filho, H. L. S. Seó, R. R. Daros, D. Enriquez-Hidalgo, A. V. Wendling and L. C. Pinheiro Machado, "Voisin Rational Grazing as a sustainable alternative for livestock production.," *Animals: An Open Access Journal from MDPI, 11(12), 3494. https://doi.org/10.3390/ani11123494*, 2021.

[36] D. P. Platis, C. D. Anagnostopoulos, A. D. Tsaboula, G. C. Menexes, K. L. Kalburtji and A. P. Mamolos, "Energy analysis, and carbon and water footprint for environmentally friendly farming practices in agroecosystems and agroforestry.," *Sustainability, 11(6), 1664. https://doi.org/10.3390/su11061664*, 2019.

[37] A. Bardule, I. Grinfelde, D. Lazdina, A. Bardulis and T. Sarkanabols, "Macronutrient leaching in a fertilized juvenile hybrid aspen (Populus tremula L. × P. tremuloides Michx.) plantation cultivated in an agroforestry system in Latvia.," *Hydrology Research, 49(2), 407–420. https://doi.org/10.2166/nh.2017.054*, 2018.

[38] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, R. Chou, J. Glanville, J. M. Grimshaw, A. Hróbjartsson, M. M. Lalu, T. Li, E. W. Loder, E. Mayo-Wilson, S. McDonald and D. Moher, "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews.," *BMJ (Clinical Research Ed.), n71. https://doi.org/10.1136/bmj.n71*, 2021.

[39] M. Duru, O. Therond and M. Fares, "Designing agroecological transitions; A review.," *Agronomy for Sustainable Development, 35(4), https://doi.org/10.1007/s13593-015-0318-x*, p. 1237–1257., 2015.

Meléndez-Segura, K. J., received the Bs. Eng in environment Engineering, specialist in environmental pedagogy and master's degree in teaching exact and natural sciences from the National University of Colombia. Currently, a PhD student in sustainable development. His research interests include: water quality, water resource conservation. ORCID: https://orcid.org/0000-0003-1311-7956

Reyes-Pineda, H, PhD Researcher in Chemical and Nuclear Engineering, Master in Membrane Technologies, Electrochemistry, Environment and Nuclear Safety, Specialist in Electrochemical Engineering and Corrosion, Specialist in Environmental Education, Chemical Engineer. Professor of the Master's Degree in Environment and Sustainable Development at the University of Manizales, Manizales, CIMAD Research Group. Caldas (Colombia). ORCID: https://orcid.org/0000-0001-9475-1910

Betancur-Pérez, J. F., Ph.D in Agricultural Sciences. Specialist in Molecular Biology and Biotechnology Researcher Teacher – School of Medicine. Director of the BSI Integrated Biosystems research line. CIMAD Research Group, University of Manizales. ORCID: https://orcid.org/0000-0002-5979-1498