International Journal of Applied Mathematics
Volume 39 No. 1s, 2026
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

ENHANCED INTERPRETABILITY USING UMAP-BASED LOCAL BIPLOTS FOR THE ANALYSIS OF COFFEE PRODUCTION IN HUILA

Anderson Cardozo-Fajardo[1] , Diego Alejandro Manrique-Cabezas[2],

Alejandro Pérez y Soto Dominguez PhD.[3]

[1] Magíster en Economía. Universidad de Manizales, Colombia. acardozo105036@umanizales.edu.co https://orcid.org/0009-0000-4037-1699

[2]Docente Tiempo completo. Universidad Autónoma de Manizales, Manizales, Colombia.

Estudiante Doctorado en Ingeniería Automática. Universidad Nacional de Colombia, sede Manizales, Manizales, Colombia.

diegoa.manriquec@autonoma.edu.co dimanriquec@unal.edu.co https://orcid.org/0000-0002-6493-3657

[3] Docente Asociado. Universidad Nacional de Colombia, Sede Medellín

aperezy@unal.edu.co https://orcid.org/0000-0001-5190-5297

Abstract

This study applies an interpretable machine learning framework based on UMAP dimensionality reduction and visualization through Local Biplots to analyze the structural complexity of specialty coffee production in the department of Huila, Colombia. Using a multivariate database covering the period 2001–2025 and integrating productive, climatic, agronomic, and economic variables, the methodology identifies latent patterns, heterogeneous production regimes, and nonlinear relationships that remain hidden under conventional linear approaches. The UMAP-based Local Biplot reveals compact and coherent clusters that capture local structures in the reduced space, outperforming the classical SVD-Biplot, which exhibits overlap and limited separation among observations. The results show that the relevance of key variables varies substantially across clusters, highlighting the influence of seasonality, solar radiation, production costs, and labor intensity in shaping local production dynamics. Likewise, the Kernel Density Estimation (KDE) applied to temporal segments evidence transitions from concentrated structures in the early periods to more dispersed and diverse configurations in recent years, likely associated with climatic variability, technological adoption, and crop renovation cycles. Overall, the findings demonstrate that nonlinear and interpretable methods provide a deeper understanding of the multidimensional factors that determine coffee productivity, surpassing the limitations of aggregated indicators and linear techniques.

Keywords: UMAP; Local Biplots; interpretability; machine learning; precision agriculture; dimensionality reduction.

1. Introduction

Colombia is one of the world's leading coffee producers, with estimated revenues exceeding 14 trillion Colombian pesos (approximately 3.14 billion dollars) for 2024 [1]. However, the country faces important challenges regarding production dynamics and the long-term sustainability of the sector [2]. Although the quality of Colombian coffee is internationally

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
recognized and constitutes a distinctive feature of national agriculture, production levels remain well below those of the main global competitors—Brazil and Vietnam—whose output surpasses Colombia's by a factor of four to five, with Colombia producing approximately 13.2 million bags [3]. In Colombia, small-scale farming predominates, with coffee farms typically not exceeding five hectares; in fact, around 95% of the 550,000 producers are smallholders. By contrast, Brazil is characterized by large-scale plantations [4]. This distinction highlights the need to strengthen productive efficiency without compromising the quality, flavor and aroma that characterize Colombian beans—attributes that support the added value of specialty coffee in regions such as Huila [5].

Several indicators show that 26% of national rural employment and 7% of agricultural GDP depend directly on coffee production [6]. This economic relevance is further reflected in the territorial magnitude of the crop, underscoring its structural role in land use and agricultural planning [7]. According to the Unidad de Planificación Rural Agropecuaria (2025), Colombia has approximately 26 million hectares of land with agricultural vocation, equivalent to 23% of the national territory. However, only about 5 million hectares are actually used for agricultural purposes, and of these, merely 845,000 hectares correspond to coffee production [8]. This distribution highlights the strategic importance of coffee in the sustainable use of agricultural land and reinforces Huila's position as a central hub for specialty coffee in Colombia, where optimal agroecological conditions, productive innovation and quality differentiation converge within small production units of up to five hectares—forming the backbone of the national coffee model [9].

In this context, Colombia's agricultural sector exhibits high productive potential and substantial crop diversity, with coffee maintaining a leading role as an export commodity and as a driver of rural development [10]. Nevertheless, the development of coffee production—and agricultural production more broadly—has been constrained by structural factors such as limited infrastructure, weak connectivity, and insufficient access to technical services, all of which restrict competitiveness relative to other producing countries [11]. Furthermore, many production processes remain predominantly manual and empirical, characterized by limited adoption of technology, automation and agro-industrial digitalization. This situation is compounded by poor infrastructure, unequal access to land and supply chain inefficiencies, which represent obstacles for both coffee growers and the agricultural sector as a whole [12].

Huila, for its part, has consolidated itself over the last decade as Colombia's top coffee-producing department, contributing 35% of the country's total output [13]. Nationally, an estimated 603 municipalities across 22 departments participate in some stage of coffee activity, demonstrating that a substantial portion of the country maintains direct ties with this value chain [14]. This wide territorial presence is explained by agroecological factors—such as suitable altitude ranges, temperature patterns, precipitation regimes and soil characteristics—as well as historical and socioeconomic factors associated with agrarian colonization, the predominance of small and medium producers, and sectoral policy orientation [15].

As a result of this territorial breadth and the strategic relevance of coffee to rural development, modernization of coffee cultivation and the incorporation of data-driven approaches present opportunities to strengthen productivity and sustainability, especially in high-potential regions

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
such as Huila [12]. Modernization requires coordination between public and private actors given the importance of coffee as Colombia's main agricultural export and its strong influence on the rural economy [16]. Small and medium coffee growers face the challenge of maintaining quality amid climate variability and agroecological heterogeneity, making it necessary to better understand productive patterns and their determinants [17].

Traditionally, analyses of coffee production have focused on aggregate indicators such as yield per hectare, annual productivity or average income per producer, generally interpreted through linear or descriptive models ([18]. Although these approaches have been useful for generating general diagnoses, they tend to oversimplify the structural complexity of coffee production by failing to consider the interactions between agro-climatic, socioeconomic, technological and quality-related factors [17]. Variables such as altitude, slope, soil fertility and post-harvest practices influence productivity and quality in non-linear ways [19].

Over the past decade, the integration of artificial intelligence and machine learning techniques has significantly transformed global agricultural production [20]. These technologies have improved monitoring and decision-making processes through yield prediction, early detection of pests and diseases, modeling of spatial-temporal variability, and automation of management practices using remote sensing, satellite imagery and geographic information systems [21]. Their implementation has enhanced the efficiency of water and fertilizer use, strengthened supply-chain traceability and enabled the identification of productive patterns that were previously difficult to capture with traditional methods [22]. In small-scale farming contexts, these tools have helped reduce uncertainty associated with climate and supported production planning, highlighting the potential of algorithmic analysis to address challenges related to sustainability, climate variability and competitiveness in contemporary agriculture [23]; [24].

In this sense, non-linear dimensionality reduction techniques such as UMAP (Uniform Manifold Approximation and Projection) offer an analytical alternative for exploring complex relationships among productive, environmental and socioeconomic variables. These methods enable the construction of interpretable local biplots that facilitate the identification of productive typologies and improvement strategies adapted to the specific characteristics of each territory [25]. This approach is supported by recent empirical evidence in the field of precision agriculture in Colombia, where the use of machine learning models combined with multivariate data analysis techniques has proven useful for developing practical recommendations aimed at improving crop establishment based on local agroecological and socioeconomic conditions. This confirms the relevance of non-linear data analysis for capturing the complexity of tropical agricultural systems and reinforces the importance of applying UMAP and local biplots to the study of coffee production in Huila [25]; [26].

The main objective of this research is to evaluate the potential of UMAP analysis to identify local productive typologies and determinant variables associated with coffee production in the department of Huila, thereby contributing to evidence-based rural development strategies. Complementarily, the study seeks to apply dimensionality-reduction methods to visualize the internal structure of productive data; generate interpretable biplots integrating economic, agro-climatic and quality-related dimensions; and propose a methodological approach that supports technical decision-making and public policies associated with coffee sustainability in Huila.

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
However, the adoption of agriculture 4.0 tools in Colombia remains at an early stage, largely limited to pilot projects promoted by the Ministry of ICT in different cropping systems, including coffee [27]. This reveals a wide field for research from economic, technological and productive perspectives, where it is essential to analyze the potential impact of emerging technologies on efficiency, sustainability and competitiveness among the country's agricultural producers [28].

2. Related Work

In recent years, precision agriculture (PA) has experienced significant advancements, becoming a consolidated approach that integrates information and communication technologies to optimize agricultural decision-making based on the spatial and temporal variability of crops [29]. This approach combines tools such as field sensors, remote sensing, geographic information systems, Big Data, cloud computing and artificial intelligence to generate precise and timely information that strengthens real-time productive management [30]; [31].

Currently, remote sensing, automation and machine learning are significantly driving the development of precision agriculture by enabling more efficient, site-specific and sustainability-oriented agricultural practices [32]. These technologies require the processing of heterogeneous, high-dimensional datasets to support timely and robust decision-making across diverse agroecological systems [33]. However, to fully harness their potential, two central challenges persist: generating agricultural predictions that are truly interpretable and enabling transparent analyses capable of uncovering complex structures in the data [34]. If these limitations are not properly addressed, the integration of remote sensing and machine learning into agricultural workflows will remain partial and constrained, significantly reducing the practical value of insights derived from data analysis [35].

In this context, demand for transparent and context-sensitive artificial intelligence systems continues to grow across disciplines. In medical imaging, for example, attention-based models facilitate interpretation through visual saliency maps, while graph-based explainable frameworks strengthen traceability in diagnosing chronic diseases [36]. Similar progress is observed in environmental sciences and remote sensing, where explainable models are used to assess air quality, estimate climate risks and analyze socioeconomic impacts associated with pollution [37]. Likewise, studies using hyperspectral imagery show that it is possible to improve model performance without sacrificing interpretability [38]. Consequently, explainability methods have emerged as essential tools in fields requiring high precision and the management of large volumes of data [39]. Despite similar pressures in agriculture, adoption of these methods remains limited, especially under real-world production conditions where datasets are heterogeneous, labels exhibit uncertainty and trust is critical in decision-support systems [40].

In agriculture, the interpretability of machine learning models is fundamental to supporting decisions in real-world environments, where transparency of results is just as important as accuracy [41]. Visualization techniques based on dimensionality reduction allow agronomists, breeders and sector stakeholders to explore patterns, evaluate predictions and extract knowledge from large datasets by translating high-dimensional representations into accessible

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
formats without losing their essential structure [42]. However, challenges such as irrelevant variables, information redundancy and the curse of dimensionality hinder intuitive interpretation and can erode model trust [43]. Additionally, agricultural datasets incorporate variables that interact nonlinearly and exhibit significant temporal variability due to climate fluctuations, management practices and crop developmental cycles [44]. As a result, basic statistical properties such as mean and variance change over time, introducing non-stationarity that complicates the construction of stable and generalizable models. These challenges highlight the need for adaptive analytical frameworks capable of capturing dynamic relationships across time and space, particularly in agricultural systems that are complex and highly sensitive to environmental conditions [45].

To address these complexities, various dimensionality reduction techniques have been applied [46]. Classical methods such as Principal Component Analysis, Linear Discriminant Analysis and Factor Analysis offer interpretable solutions based on linear projections; while effective for simplifying data and preserving a substantial proportion of variance, their linearity assumptions limit their ability to capture the nonlinear relationships that characterize contemporary agricultural systems [47]. In response, nonlinear manifold learning techniques—such as Locally Linear Embedding, Isomap, t-SNE and UMAP—have gained prominence by more faithfully representing local and global structures, facilitating visual exploration and cluster identification in high-dimensional spaces. However, these techniques face persistent challenges, including high sensitivity to noise, scalability issues and limited interpretability at the variable level—factors that have constrained their adoption in real agricultural applications [48].

Other nonlinear approaches such as Self-Organizing Maps and autoencoders have also been used for visualization and interpretation of agricultural information. While these techniques capture complex patterns and produce useful representations, they require detailed parameter tuning and their latent spaces are often opaque without auxiliary tools, limiting transparency and reducing utility in contexts where technical decision support is required [49]. In contrast, biplot visualizations have reemerged as valuable tools because they simultaneously represent observations and variable loadings in reduced space, providing interpretable summaries of multivariate datasets and serving as a bridge between statistical modeling and visual analysis [50]). However, traditional biplots remain largely descriptive and are seldom integrated into predictive workflows, limiting their usefulness within contemporary machine learning pipelines [9].

Recent studies have expanded the scope of biplots toward predictive tasks. Some works have combined UMAP projections with local biplot construction to interpret water stress in crops, while others have demonstrated their usefulness in soil analysis and microbiological studies [25]. These advances illustrate the potential of such techniques to increase model transparency and traceability; nevertheless, coherently integrating visual exploration with predictive performance continues to represent a significant methodological challenge.

Among modern tools, machine learning has become one of the most relevant approaches for processing large volumes of agricultural information and detecting hidden patterns that improve decision-making [51]). Unsupervised methods—such as cluster analysis—support the

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
identification of productive groups or typologies based on agroecological, technological or socioeconomic variables without requiring predefined labels [35]. These techniques help understand the internal diversity of production systems and enable differentiated strategies based on the characteristics of each area or production unit—an especially useful capability in heterogeneous crops such as coffee.

The relevance of such approaches has been demonstrated in various agricultural contexts. For example, Nikhil et al. (2024) [52] developed a study in southern India aimed at predicting crop yields using machine learning models such as Gradient Boosting Regression, KNN and Linear Regression. The analysis integrated meteorological data, agricultural production information and soil characteristics from official and open-access sources. Results showed significant improvements in prediction accuracy and demonstrated the decisive influence of climatic and edaphic factors on agricultural yield, highlighting the importance of data-driven methods for addressing sustainability and productivity challenges.

Similarly, Patil and Mane (2024) [53] proposed a crop recommendation system using algorithms such as Random Forest and Support Vector Machines (SVM), achieving 93.7% accuracy when recommending optimal crops based on local soil and climatic conditions. These studies demonstrate the potential of machine learning for improving production efficiency, optimizing resource use and enhancing the resilience of agricultural systems to climate change. They also underscore the need for high-quality datasets and models adapted to local realities to maximize their utility.

In Latin America, the work by Silva-Peñafiel et al. (2022) [54] in Ecuador stands out for developing an automated diagnostic system for coffee diseases using artificial intelligence and image analysis, enabling early and accurate detection of pests and pathogens. In a later study, the same authors evaluated user perception and adoption of mobile applications with machine learning capabilities among coffee producers, finding limited knowledge of these technologies and highlighting the need to strengthen technological adoption to improve productivity and phytosanitary management [54]

In Colombia, various studies have explored the application of machine learning in agro-productive contexts. Ramírez (2020) [55] used ML models to predict crop conditions based on agronomic and management variables, identifying decision trees as the most efficient approach. Similarly, Corrales et al. (2014) [56] used Support Vector Regression (SVR) and regression trees to model the incidence of coffee rust in Cauca, obtaining high accuracy after dataset optimization. More recently, Tovar et al. (2022) [57] developed the mobile application HumiCafé, based on Convolutional Neural Networks (CNN), to measure grain moisture content with 99% accuracy. Complementarily, Churta (2020) designed a real-time monitoring system for coffee leaf conditions using CNNs, achieving 98.7% accuracy and demonstrating the potential of deep learning for production control.

In this line, the study by Triana-Martínez et al. (2024) [25] is particularly notable for developing the UMAP-based Local Biplot model to identify complex patterns associated with crop water status using remote sensors and field data in Tolima (Colombia) and Belgium. Their results showed that UMAP clearly outperforms traditional linear methods by capturing nonlinear relationships and local structures in agricultural data, improving interpretability and

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
predictive capacity in precision agriculture contexts. This methodological advancement reinforces the relevance of applying UMAP and clustering techniques to the analysis of productive systems such as coffee in Huila.

From this perspective, the present research adopts a UMAP-based approach combined with Local Biplot representations to analyze coffee production in the department of Huila. The objective is to identify productive typologies and structural relationships among agroclimatic, economic, technological and socio-productive variables, contributing to a comprehensive understanding of the regional coffee system. This approach provides empirical tools to strengthen sustainable rural development strategies and support evidence-based public policies in a context characterized by climate variability, territorial heterogeneity and the need to enhance competitiveness in Colombian coffee production.

Consequently, this study adopts a UMAP and Local Biplot framework to analyze the productive structure of coffee in Huila, aiming to identify typologies and key relationships among agroclimatic, technological and socioeconomic variables. Beyond characterizing system heterogeneity, this research opens a methodological pathway for future studies focused on predictive models, decision-support systems and precision agriculture strategies. The study thus represents a relevant contribution to the modernization of the coffee sector and to the consolidation of evidence-based public policies that strengthen the sustainability and competitiveness of Colombian coffee amidst current and emerging challenges.

3. Methodology

3.1 Variables and Data

The database is composed of monthly information for the department of Huila, with a total of 294 observations and 12 variables that integrate productive, climatic, economic and agronomic dimensions. Its structure combines data from institutional records from different entities in order to reconstruct a coherent temporal line describing the behavior of regional coffee production. This type of longitudinal database is essential for analyzing agro-productive dynamics and transitions in permanent crops such as coffee, particularly when exploring long-term relationships between climate, costs, agronomic management and productivity [13]; [58].

The 13 included variables allow for an integral characterization of the coffee system. These include climatic indicators such as precipitation, temperature, relative humidity and solar radiation, generally reported by IDEAM meteorological stations and widely used for modeling productive responses in Andean agricultural systems. The database also incorporates economic variables such as input prices, production costs and labor costs, aligned with DANE statistics and sectoral studies by the National Federation of Coffee Growers. The inclusion of planted area and yield measured in kilograms per hectare complements the analysis by reflecting changes associated with technical decisions, agroecological conditions and interannual variation of the crop [59]; [7].

The diversity and temporal continuity of the variables make this dataset suitable for machine learning models, multivariate analysis and studies of productive typologies. Its composition facilitates exploration of non-linear relationships among climate, costs, management and productivity, which is particularly relevant in Huila's coffee sector, where small-scale

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
production units and high agroecological variability predominate. Moreover, the monthly behavior of the data captures seasonal fluctuations and provides more accurate representations of productive dynamics, supporting research aimed at sustainability and data-driven decision-making [13].

3.2 Model

Unsupervised Nonlinear Relevance Analysis using UMAP-based Local Biplots

Given an input feature matrix $\mathbf{X} = \{\mathbf{x}_n \in \mathbb{R}^P : n \in N\}$, with $\mathbf{X} \in \mathbb{R}^{N\times P}$, Pdimensions and Nsamples, a lower-dimensional UMAP-based embedding $\mathbf{Z} \in \mathbb{R}^{N\times P'}$, with $P' < P$, is obtained through two steps:

i) Construction of a k–nearest neighbors (k-NN) graph using pairwise distance matrix $d(\mathbf{x}_n, \mathbf{x}_{n'}) \in \mathbb{R}$, to measure similarity between instances $\mathbf{x}_n, \mathbf{x}_{n'} \in \mathbf{X}$.

ii) Estimation of a low-dimensional graphical representation based on the graph to preserve the data structure. Specifically, a directed k-NN graph is first constructed from $\mathbf{X}$, with edges defined by local connectivity, through the following set of weights $\mathbf{W} \in [0,1]^{N\times N}$, with elements:

$$w_{nn'} = \exp\left(-\max\left(0, \frac{d(\mathbf{x}_n, \mathbf{x}_{n'}) - \tilde{\rho}_n}{\sigma_n}\right)\right) \qquad (1)$$

where $\sigma_n \in \mathbb{R}^+$serves as a normalization factor defining a local metric around $\mathbf{x}_n$, and $\tilde{\rho}_n \in \mathbb{R}^+$is the distance to the nearest neighbors of $\mathbf{x}_n$. To ensure a symmetric diffuse measure, the weights are normalized as:

$$\breve{w}_{nn'} = w_{nn'} - w_{n'n} - w_{nn'}w_{n'n}. \qquad (2)$$

In turn, a low-dimensional weight matrix $\widetilde{\mathbf{W}}(\mathbf{Z}) \in [0,1]^{N\times N}$ is computed as:

$$\widetilde{w}_{nn'}(\mathbf{Z}) = (1 + \tilde{\alpha}\, d(\mathbf{z}_n, \mathbf{z}_{n'})^2)^{-\iota}, \qquad (5)$$

where $\mathbf{z}_n, \mathbf{z}_{n'} \in \mathbf{Z}$denote points in the embedding space, and $\tilde{\alpha}, \iota \in \mathbb{R}^+$are hyperparameters—typically set to 1—that control the balance between local and global structure preservation.

The UMAP optimization objective is formulated as a cross-entropy loss function:

$$\mathbf{Z}^* = \arg\min_{\mathbf{z}_n \in \mathbf{Z}} \sum_{\substack{n\in N \\ n\neq n'}} \left[\breve{w}_{nn'} \log\left(\frac{v_{nn'}}{\widetilde{w}_{nn'}^{-Z}}\right) + (1 - v_{nn'}) \log\left(\frac{1 - v_{nn'}}{1 - \widetilde{w}_{nn'}^{-Z}}\right)\right], \qquad (3)$$

International Journal of Applied Mathematics
Volume 39 No. 1s, 2026
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
where $v_{nn'}$represents similarities in the observation space expressed through soft distances to nearest neighbors:

$$v_{nn'} = (v_{n|n'} + v_{n'|n}) - v_{n|n'}v_{n'|n}, v_{k|k'} = \exp\left(\frac{d(\mathbf{x}_k, \mathbf{x}_{k'}) - \rho_k}{\sigma_k}\right). \quad (4)$$

The low-dimensional weight $\widetilde{w}_{nn'}^{Z}$ is defined as:

$$\widetilde{w}_{nn'}^{Z} = (1 + \alpha \parallel \mathbf{Z}_{n'} \parallel_2^{2\beta})^{-1},$$

with $\widetilde{w}_{nn'}^{Z} \in \mathbf{W}^{Z} \subset [0,1]^{N\times N}$, and $\alpha, \beta \in \mathbb{R}$are tuning parameters.

UMAP enables clustering in the latent space, where groups become more evident—particularly when using the k-means algorithm:

$$\widetilde{\mathbf{Z}}_d^{*} = \arg\min_{\mu d, Zd} \sum_{n=1}^{N}\sum_{d=1}^{\widetilde{D}} \parallel Z_n - \mu_d \parallel_2^2; \text{s.t.} \widetilde{\mathbf{Z}}_d \cap \widetilde{\mathbf{Z}}_{d'} = 0 \ \forall\, d, d' \in \widetilde{D}, d \neq d' \quad (5)$$

where $\widetilde{D}$is the number of distinct clusters $\{\tilde{\mathbf{Z}}_d \in \mathbb{R}^{N_d\times 2}\}_{d=1}^{\widetilde{D}}$, with $\sum_{d=1}^{\widetilde{D}} N_d = N$, and $\mu_d \in \mathbb{R}^2$.

To improve interpretation of UMAP clusters, this study employs SVD-biplots by combining the cluster partitions $\{\tilde{\mathbf{Z}}_d\}$ with a reduced representation of the original high-dimensional feature matrix. The linear projection for cluster $d$is:

$$\tilde{\mathbf{Z}}_d = \widetilde{\mathbf{X}}_d \mathbf{B}_d, \quad (6)$$

where $\widetilde{\mathbf{X}}_d$is the low-dimensional representation obtained by 2D singular value decomposition:

$$\widetilde{\mathbf{X}}_d^{*} = \arg\min \parallel \mathbf{X}_d - \widetilde{\mathbf{X}}_d \parallel_F^2, \quad (7)$$

$$\widetilde{\mathbf{X}}_d = \widetilde{\mathbf{U}}_d \tilde{\mathbf{S}}_d \widetilde{\mathbf{V}}_d^{\top}, \widetilde{\mathbf{U}}_d \in \mathbb{R}^{N_d\times 2}, \widetilde{\mathbf{V}}_d \in \mathbb{R}^{P\times 2}, \quad (8)$$

where $\|\cdot\|_F$denotes the Frobenius norm.

Thus, the projection term is computed as:

$$\mathbf{B}_d = \widetilde{\mathbf{V}}_d \tilde{\mathbf{S}}_d^{1/2} \in \mathbb{R}^{P\times 2},$$

where $\tilde{\mathbf{S}}_d$is the diagonal matrix containing the singular values.

International Journal of Applied Mathematics
Volume 39 No. 1s, 2026
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
To ensure that nonlinear UMAP reduction preserves feature structure, an affine-like transformation is applied (Pérez, Manrique, Triana, Álvarez & Castellanos, 2025):

$$\gamma_d^*, v_d^* = \arg\min \| \widetilde{\mathbf{B}}_d - \mathbf{B}_d \|_F, \widetilde{\mathbf{B}}_d = \gamma_d \mathbf{B}_d + v_d. \qquad (9)$$

Finally, a local feature classification vector is computed:

$$\lambda_d = \mathbf{B}_d \mathbf{1}. \qquad (10)$$

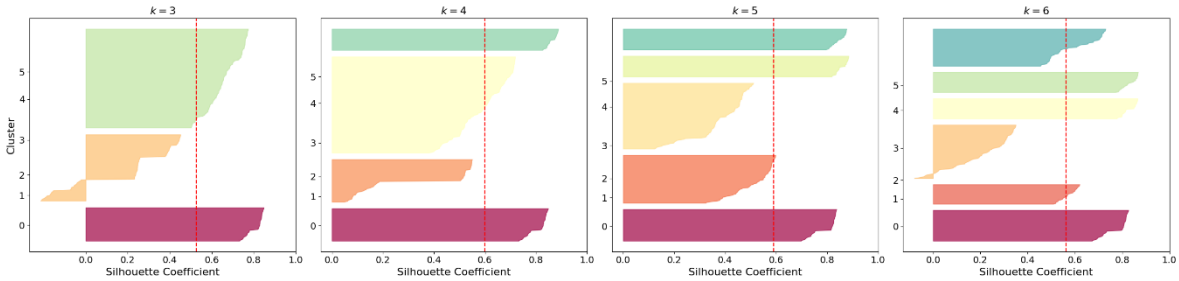
4. Results and discution

Relevance Analysis of Indicators Based on Local Biplot

Figure 1 shows the silhouette coefficients (marked by the vertical red line) obtained for different clustering configurations derived from the UL approach, considering between three and six clusters. These plots serve as a diagnostic mechanism that allows assessing cluster quality and supporting the selection of an appropriate number of partitions. Although the silhouette profiles are generally similar across the evaluated configurations—none falling below the average threshold—the four- and five-cluster solutions exhibit a more balanced distribution of observations per cluster. In contrast, the three- and six-cluster configurations show greater disparity in shape and size.

Although the silhouette coefficient reaches its highest value with five clusters, this improvement does not remain consistent as the number of clusters increases. Therefore, a five-cluster partition is recommended, as it achieves an adequate balance between compactness, separation and internal cohesion, supporting a more robust interpretation of the underlying structure.

Figure 1. Silhouette coefficient based on UL obtained from clustering results.



Source: Authors' elaboration based on results (2025).

The results presented in the figure reveal differentiated patterns of variable relevance among the five clusters identified through the UL-Biplot approach shown in Figure 2. In the input data, a marked heterogeneity in the relative contribution of variables is observed, suggesting that the groups differ substantially in both their productive structure and their climatic and economic conditions. This initial variability becomes more evident in the local biplots, where the representation of each cluster makes it possible to isolate and visualize the specific relationships among variables within each group.

Cluster 1 exhibits generally low relevance levels across almost all variables, reflecting a homogeneous and weakly structured behavior among the observations it contains. Cluster 2, in

International Journal of Applied Mathematics
Volume 39 No. 1s, 2026
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
contrast, shows a more defined pattern, with precipitation, temperature, the input price index, and production cost per hectare standing out as key variables—suggesting environmental and



economic conditions strongly associated with local productivity. Cluster 3 presents the most clearly defined structure, with high relevance in both productive and climatic variables, indicating a more complex behavior likely linked to areas with higher productive dynamism. In Cluster 4, the matrix reveals isolated and sporadic contributions, pointing to a weak internal structure or reduced variability. Finally, Cluster 5 is characterized by an even more concentrated pattern, where only a few variables dominate the local explanation, suggesting highly specific productive processes or unique conditions in this small subset of observations.

Taken together, the results confirm that local representation through UL-Biplot facilitates the identification of distinctive patterns within each cluster, highlighting key variables and enabling a more precise understanding of the internal heterogeneity of the coffee production system. This approach complements the findings of the global analysis and offers a robust tool for interpreting structural relationships within each group.

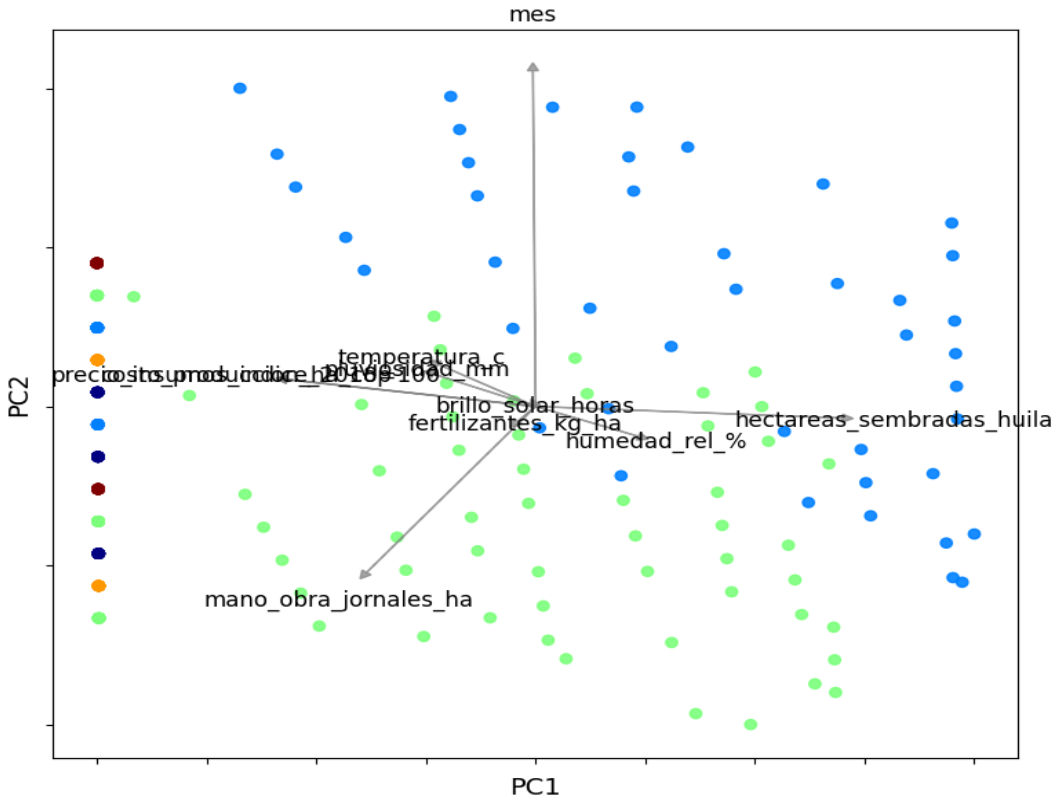
The comparison between the input data and the patterns derived from the UL-Biplot demonstrates the ability of the local approach to filter noise and highlight only the most relevant structural relationships within each cluster. While the matrix corresponding to the original data shows a more dispersed and even redundant distribution of relevances—typical of a system with high multicollinearity and significant climatic and productive variability—the local biplots focus attention on a reduced subset of key variables that explain the behavior of each group. This contrast shows that the local representation not only improves interpretability but also captures nonlinear interactions that remain hidden in the original space. Consequently, the UL-Biplot acts as a structural refinement mechanism capable of revealing the internal productive logic of each cluster, facilitating the identification of critical factors for decision-making, the design of differential strategies and the precise characterization of the territorial heterogeneity of coffee cultivation in Huila.

Illustration 2. Visual inspection results based on the SVD biplot and the Local Cluster Biplot (Clusters 1–5).

International Journal of Applied Mathematics
Volume 39 No. 1s, 2026
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Source: Authors' elaboration based on results (2025).

Illustration 3. Visual inspection results based on the SVD biplot and the PC1–PC2 Biplot.



Source: Authors' elaboration based on results (2025).

Illustration 3 shows the SVD-Biplot projection of the coffee dataset for Huila, where broad dispersion patterns and a complex internal structure are observed. The observations are heterogeneously distributed across the plane defined by the first two principal components, suggesting substantial variability across productive cycles, agroclimatic conditions, and production costs. The arrangement of points reveals considerable overlap, which limits the natural separation of groups and reflects the difficulty of representing nonlinear relationships through a strictly linear model such as the SVD-Biplot.

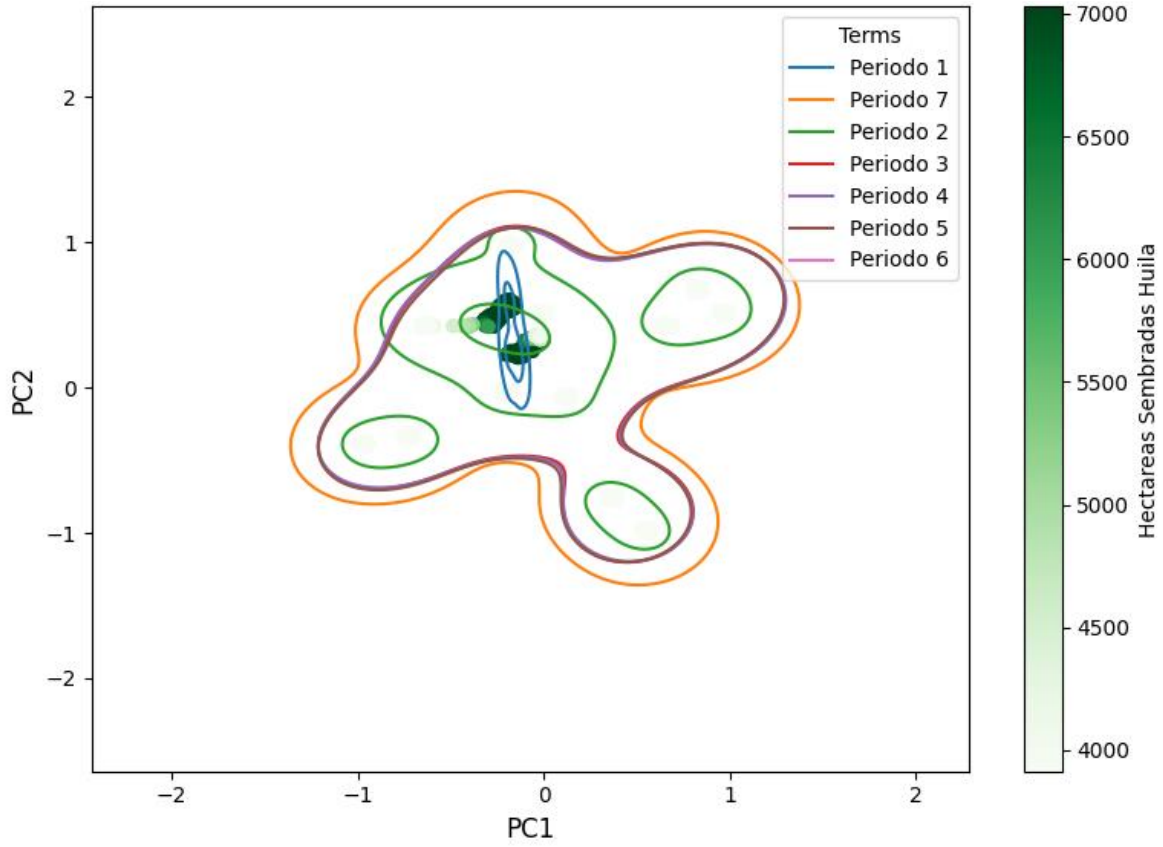
The base vectors provide information on the relative contribution of each variable. In this case, anio and costo_produccion_ha_cop exhibit the largest vector magnitudes and align with the first principal component (PC1), indicating that both factors account for a significant portion of the structural variability associated with temporal changes and production costs in the analyzed period. In the second principal component (PC2), Mes, mano_obra_jornales_ha, and rendimiento_cafe_kg_ha stand out, highlighting the influence of seasonality, labor demand, and productivity on the vertical fluctuations in the map. Likewise, the proximity between anio and fertilizantes_kg_ha suggests a positive association between the two variables, possibly linked to the gradual increase in input use throughout the time series. In contrast, variables such as hectareas_sembradas_huila or brillo_solar_horas display shorter vectors, indicating a reduced contribution to the structure explained by the first components.

From a broader perspective, the structure revealed by these projections confirms that Huila's coffee system exhibits a non-trivial organization, with internal dynamics that likely correspond to productive subgroups, temporal transitions, or non-stationary agroclimatic patterns. These

International Journal of Applied Mathematics
Volume 39 No. 1s, 2026
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
results contrast with those obtained through the UL-Biplot, where the local structure preserved by UMAP produces more compact and coherent groupings, capturing nonlinear relationships that the SVD-Biplot fails to represent adequately. The two-dimensional nonlinear embedding—normalized in the range of −1 to 1—successfully reflects both the global geometry of the dataset and the essential local relationships, delineating subgroups that are more interpretable and consistent with the productive realities of the territory.

The color coding of the vectors provides an additional layer of interpretation: productive and agronomic variables tend to exert a stronger influence on group separation, whereas economic variables do so in a more limited and localized manner. This suggests that variations in coffee yields and crop structure are more closely associated with agroclimatic and management conditions than with the monetary factors included in the dataset. Altogether, the evidence indicates that nonlinear methods such as UMAP offer a more faithful and differentiated representation of the heterogeneity of the coffee system, establishing themselves as key tools for identifying productive typologies and latent patterns.

Illustration 4. KDE analysis by periods.



Source: Authors' elaboration based on results (2025).

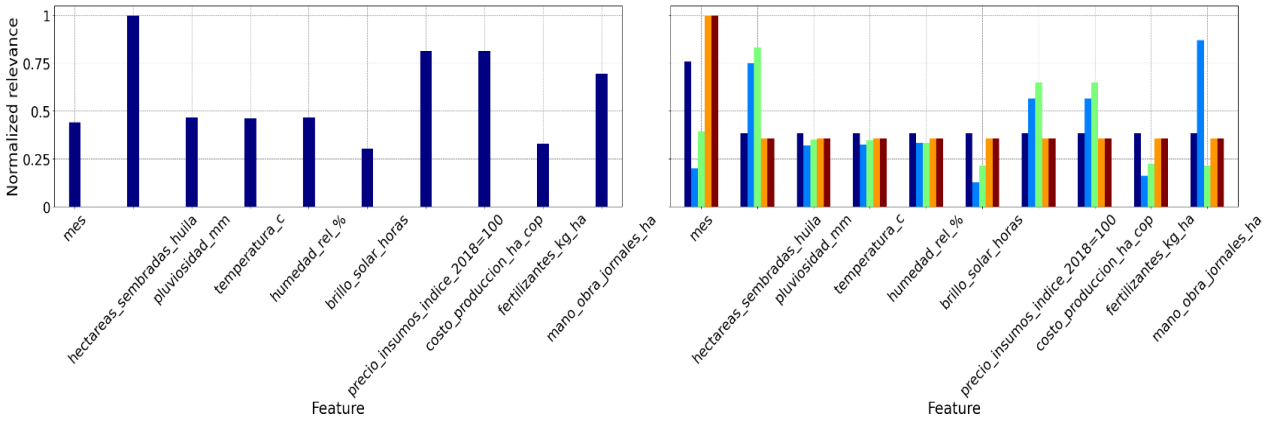
On the other hand, Illustration 4 presents a Kernel Density Estimation (KDE) analysis applied to the two-dimensional projection obtained from the first two principal components (PC1 and PC2), using a four-year segmentation of the database. Each contour represents the distribution of observations belonging to one of the seven defined periods, allowing the visualization of how the productive characteristics of the department of Huila evolve over time. The contours reveal the shape, concentration, and dispersion of each temporal group within the reduced

International Journal of Applied Mathematics
Volume 39 No. 1s, 2026
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
space, where shades of green indicate the magnitude of the hectares planted in each observation.

Overall, the more compact contours—such as those observed in the initial periods—suggest a relatively homogeneous and concentrated productive structure, typical of phases in which variability among productive units was limited. In contrast, periods with broader contours—such as Period 7—exhibit an expansion in productive variability, possibly associated with coffee plantation renewal processes, climatic fluctuations, or increasing differences in the adoption of technologies and agricultural practices. Likewise, the partial overlap between contours indicates that, although differences exist across periods, common elements persist in the productive and agroclimatic dynamics of the department.

The color gradient associated with the planted hectares complements the interpretation of the KDE contours by highlighting the areas of the map where the highest levels of cultivation are concentrated. The darker shades suggest periods characterized by greater expansion of coffee crops or by the presence of larger productive units. Taken together, this visualization facilitates the identification of temporal density patterns, the comparison between productive cycles, and the detection of possible structural transitions in Huila's coffee sector over more than two decades.

Illustration 5. Unsupervised feature relevance analysis.



Source: Authors' elaboration based on results (2025).

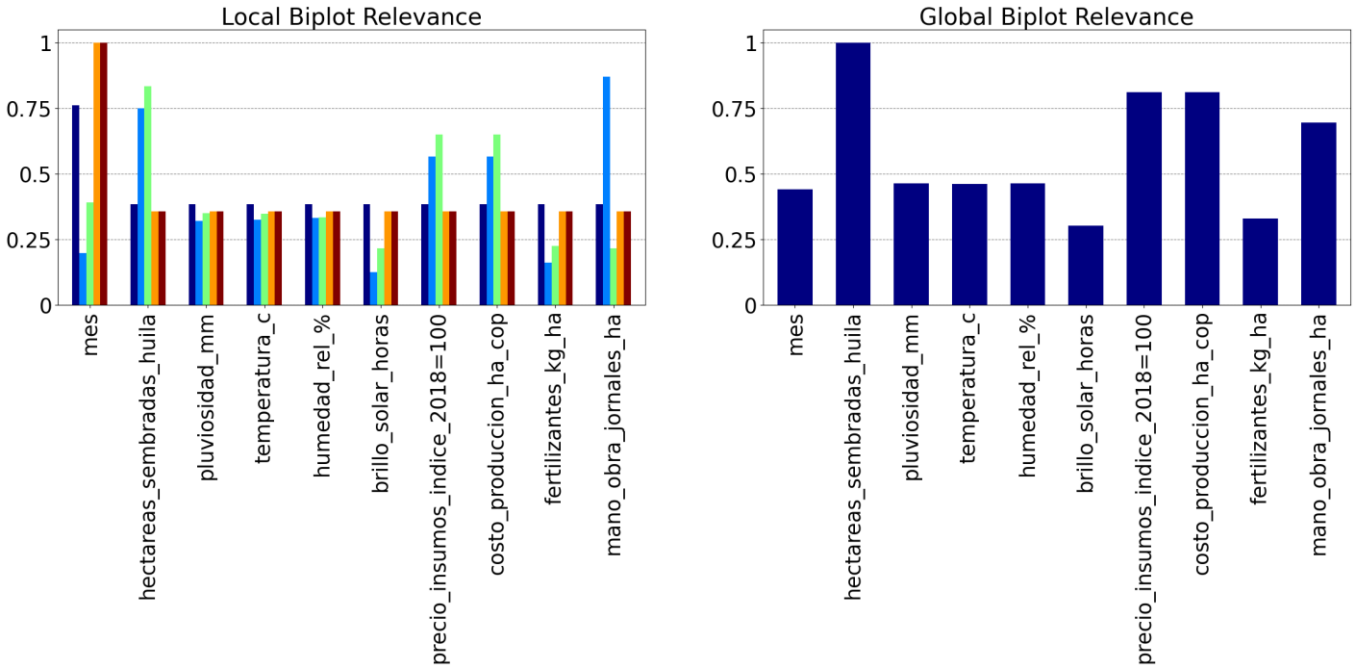
Illustration 5 presents the results of the feature relevance analysis obtained through a fully unsupervised approach. The plot on the left displays the normalized global relevance for each variable in the dataset, revealing which factors exert the greatest influence on the internal structure of the productive system. In this representation, the variables mes, brillo_solar_horas, costo_produccion_ha_cop, and rendimiento_cafe_kg_ha stand out due to their high relative relevance, indicating that they play a central role in the multidimensional configuration of coffee cultivation in Huila. In contrast, variables such as temperatura_c and hectareas_sembradas_huila show lower contributions in this global analysis.

The plot on the right deepens this perspective by disaggregating the relevance for each of the five clusters identified through the UL model. This local perspective allows for the observation of significant variations in variable importance according to the internal dynamics of each group. For example, the variable mes maintains a dominant contribution across all clusters, confirming the strong influence of seasonality on productive processes. Similarly, precipitation and solar radiation gain high relevance in specific clusters, reflecting agroclimatic differences that affect the groups in distinct ways. Meanwhile, economic factors such as precio_insumos_indice_2018=100 and costo_produccion_ha_cop exhibit heterogeneous levels

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
of relevance, indicating that their impact depends on the particular characteristics of each productive segment.

Taken together, both panels demonstrate that the relevance of variables in an agricultural context cannot be assumed to be homogeneous: while some characteristics dominate the global structure of the dataset, others become important only in local contexts. This global–local duality highlights the usefulness of the unsupervised approach for identifying underlying patterns and understanding the heterogeneity of the Huila coffee system, providing a solid foundation for subsequent clustering and predictive modeling analyses.

Illustration 6. Global and Local Variable Relevance: A Multiscale Interpretability Mechanism.



Source: Authors' elaboration based on results (2025).

On the other hand, the relevance analysis shown in Illustration 6 reveals differentiated patterns in the behavior of the variables. In the right panel, the Local Biplot—derived from the embeddings generated by UMAP—shows that the relative importance of each feature varies substantially across clusters, confirming the internal heterogeneity of the productive system. This variability reflects the presence of substructures and nonlinear relationships that only emerge when the data are analyzed from a local perspective. In contrast, the left panel displays the global relevance obtained through the SVD-based Biplot, where certain variables dominate the overall explanation of the dataset but fail to adequately capture the distinctions within groups.

The disparity between global and local relevance underscores the need for adaptive analytical approaches that can account for both the general structure and the internal particularities of each subgroup. In this regard, the UL-Biplot framework enhances interpretability by isolating subsets of observations that share similar behaviors, revealing patterns that remain hidden in traditional linear methods. In this way, the multiscale analysis not only improves the

Volume 39 No. 1s, 2026
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
understanding of the key determinants of the coffee system but also provides a solid foundation for designing differentiated and context-specific strategies.

Conclusions

Over the last decade, the advancement of machine learning models with interpretability capabilities has begun to profoundly transform agricultural production systems worldwide. Unlike purely predictive approaches, explainable methodologies allow for a transparent understanding of the factors that drive crop yield, quality, and sustainability, facilitating informed decision-making in contexts characterized by high climatic variability, territorial heterogeneity, and structural constraints. Contemporary agriculture faces unprecedented challenges—from variability associated with climate change to the need to optimize inputs and improve productive efficiency—which demand tools capable of capturing nonlinear relationships while providing useful interpretations for producers, technicians, and policymakers. In this regard, interpretable machine learning emerges as a strategic alternative that combines analytical precision with transparency, two necessary elements for accelerating technological adoption in the sector.

Moreover, the integration of explainable frameworks such as UMAP, local biplots, sequential attention, or hybrid models strengthens the capacity to detect latent patterns that once remained hidden within large volumes of agroclimatic, economic, and productive data. These approaches enable the decomposition of agricultural system complexity into interpretable components that reveal productive typologies, temporal transitions, and critical factors associated with crop management. By offering a clear reading of multivariable interactions, these models not only improve prediction accuracy but also facilitate the design of personalized recommendations, risk mitigation strategies, and differentiated policies tailored to the specific characteristics of each territory. In this way, agriculture driven by explainable artificial intelligence is emerging as a fundamental axis for advancing toward more efficient, resilient, and sustainable production systems.

Additionally, the results derived from the cluster analysis using the UL technique show that the coffee production structure in Huila exhibits substantial heterogeneity. The joint evaluation of silhouette coefficients, local biplots, and nonlinear projections confirms that the coffee system is composed of differentiated subgroups whose internal dynamics respond to temporal, agroclimatic, economic, and crop management patterns. This structural diversity is not adequately captured by traditional linear techniques, highlighting the relevance of nonlinear methodologies based on embeddings and local analysis.

Second, the comparison between the SVD-Biplot and the UL-Biplot demonstrates that relationships among coffee variables are predominantly nonlinear. While the SVD-Biplot provides a general view constrained by observation overlap and the inability to capture complex curvature in the data space, the UL-Biplot preserves local structures and reveals coherent clusters, more faithfully reflecting the productive reality of the territory. This evidence confirms that the non-stationary and multivariable nature of the coffee sector requires approaches capable of modeling highly irregular geometries.

Third, the feature relevance analysis reveals marked differences between global and local importance. Although factors such as month, solar radiation, production costs, and yield appear as globally dominant influences, their role changes substantially across clusters. This finding indicates that coffee production is not conditioned by a uniform set of variables but rather by specific combinations that emerge depending on the local context, reaffirming the need for differentiated strategies in agricultural analysis and technical management.

Volume 39 No. 1s, 2026
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Fourth, the KDE analysis using four-year periods confirms that the temporal evolution of coffee production in Huila is neither linear nor uniform. More compact contours in early periods suggest more homogeneous production, whereas recent periods display greater dispersion, possibly associated with uneven technological adoption, increasing climatic variability, and structural transformations in the crop. This shows that the territory has undergone processes of diversification and productive transition that must be considered in agricultural planning.

Indeed, the integrated results of this study confirm that the combination of nonlinear dimensionality reduction techniques, local biplot analysis, and unsupervised clustering constitutes a solid framework for understanding the complexity of the coffee system. This approach enables the identification of productive typologies, visualization of latent patterns, and recognition of the characteristics that structure each segment of the territory. Altogether, it offers a more accurate, explanatory, and useful representation for evidence-based rural development strategies.

It is recommended that future research incorporate additional variables related to bean quality, detailed phenological information, and records of differentiated agronomic practices by farm or zone. The inclusion of these dimensions would enhance analytical precision and deepen the identification of determinants of productivity and specialty coffee quality, particularly in a highly heterogeneous territory like Huila. Likewise, integrating high-resolution satellite information and IoT sensor data would increase the spatial and temporal granularity of the analysis.

Furthermore, advancing in the design of predictive models that integrate nonlinear embeddings and local variable relevance with supervised machine learning approaches is suggested. This would allow the construction of hybrid models capable of generating predictions that are both accurate and interpretable, aligned with the explainability principles required in agricultural contexts. Validating these models across different coffee-producing regions of the country would contribute to consolidating a replicable and robust methodological framework for analyzing complex agricultural systems.

Bibliography

[1] F. A. Meza, Tecnologías 4.0 En El Desarrollo Y Crecimiento De Economía Colombiana: Revisión De Literatura Desde El Sector Cafetero, Medellin, 2022.

[2] H. M. D. O. H. &. G.-T. J. C. Salazar Echeverry, The Economic Efficiency of Coffee Growers in the Department of Caldas, Colombia., vol. 11, Economies, 2023, p. 255.

[3] P.-S. H. L. P. García L JC, Recommendations for the regionalizing of coffee cultivation in Colombia: a methodological proposal based on agro-climatic indices., PLoS One, 2014.

[4] A. &. F. A. Samoggia, «Revealing the Governance Dynamics of the Coffee Chain in Colombia: A State-of-the-Art Review.,» *Sustainability,* vol. 15, nº 18, 2023.

[5] E. Teran, «Enhancement of Coffee Quality Attributes by Combining Processing Methods and Varieties. Beverages,» vol. 10, nº 1, 2024.

[6] R. L. M.-R. G. A. &. G. A. Ferrucho, «Integrated Disease Management for the Sustainable Production of Colombian Coffee,» *Agronomy,* vol. 14, nº 6, 2024.

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
[7] R. A. H. Sánchez, «The Crucial Role of Agriculture in the Colombian Economy: Challenges and Opportunities,» *Revista Facultad Nacional de Agronomía Medellín,* vol. 77, nº 3, 2024.

[8] C. Hinojosa, K. Sanchez, A. Camacho y H. Arguello, «AgroTIC: Bridging the gap between farmers, agronomists, and merchants,» *Computers and Society,* 2023.

[9] L. M. Guevara, «Rural Land Formalization's Institutional Dynamics and Contribution to Future Rural Reform in Colombia,» 2023.

[10] P. Mooney, La insostenible agricultura 4.0 Digitalización y poder corporativo en la cadena alimentaria, Ciudad de Mexico, 2020.

[11] A. Mamun, «IoT-Based Agriculture and Smart Farming: Machine Learning Applications: A Commentary,» *ORCID,* vol. 2, nº 1, 2024.

[12] T. Arevalo-Ramirez y F. A. Cheein, Cluster Analysis for Agriculture, Encyclopedia of Smart Agriculture Technologies, 2023.

[13] Y. Hozumi y G.-W. Wei, «Analyzing scRNA-seq data by CCP-assisted UMAP and t-SNE,» *ARXIV,* pp. 1-18, 2023.

[14] R. Taghizadeh-Mehrjardi, K. Nabiollahi, N. M. Kebonye, N. Kakhani, M. Ghebleh-Goydaragh, B. Heung, A. Amirian-Chakan, S. M. T. Hossaini y T. Scholten, «High-performance soil class delineation via UMAP coupled with machine learning in Kurdistan Province, Iran,» *Geoderma Regional journal,* 2024.

[15] U. V. Nikhil, A. M. Pandiyan, S. P. Raja y Z. Stamenkovic, «Machine Learning-Based Crop Yield Prediction in South India:,» *Computers,* vol. 13, nº 137, 2024.

[16] N. A. Patil y S. Mane, «Crop Recommendation in Precision Agriculture Using Machine Learning Techniques,» *Research Square,* 2024.

[17] G. E. Silva-Peñafiel, B. M. Llanos-Bonilla, A. M. Cordova-Vaca y W. C. Cunuhay-Cuchipe, «mplementación de modelo machine learning aplicado al estudio de enfermedades de café en el centro de investigación sacha wiwa,» *Polo del conocimiento,* vol. 8, nº 7, pp. 625-640, 2022.

[18] G. E. Silva-Peñafiel, M. V. Estrada-Velasco, Á. A. Flores-Lescano y D. K. Chicaiza-Angamarca, «Diagnóstico del uso de aplicaciones móviles para el estudio de enfermedades del café,» *Polo del conocimiento,* pp. 1883-1896, 2023.

[19] C. A. Ramírez Gómez, «Aplicación del Machine Learning en agricultura de precisión,» *Revista CINTEX,* vol. 25, nº 2, pp. 14-27, 2020.

[20] D. C. Corrales, A. Ledezma, A. J. P. Q, J. Hoyos, A. Figueroa y J. C. Corrales, «A new dataset for coffee rust detection in Colombian crops base on classifiers,» *Sistemas y Telemática,* vol. 12, nº 29, p. 9, 2014.

[21] Y. T. Tovar, A. F. Calvo y A. Bejarano, «Desarrollo de un sistema de clasificación de imágenes digitales para medir la humedad en granos de café,» *Información tecnológica,* vol. 33, nº 3, pp. 117-128, 2022.

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
[22] A. U. Churta, «Desarrollo De Un Sistema De Reconocimiento Del Estado De Los Cultivos De Café En Tiempo Real Usando Redes Neuronales Artificiales,» *Revista CINTEX,* vol. 25, nº 2, pp. 37-44, 2020.

[23] U.S. Department of Agriculture, «Foreing agriculture service,» 25 octubre 2025. [En línea]. Available: https://www.fas.usda.gov/data/production/commodity/0711100.

[24] Federación Nacional de Cafeteros de Colombia, «Informe de gestión,» FNC , Bogotá D.C, 2024.

[25] Federación Nacional de Cafeteros de Colombia, «Comite departamental de cafeteros del Huila,» 2024. [En línea]. Available: https://huila.federaciondecafeteros.org/cafe-de-huila/.

[26] C. Mora-Villalobos y R. Jara-Rojas, «Una nueva distribución geográfica deproducción de café - Análisis de la concentración y especialización del café producción (Coffea arábica) en municipios de Colombia entre 2007 y 2021,» *Revista de Geografia de Norte Grande,* vol. 92, pp. 1-36, 2025.

[27] C. Ferrucho, «Caficultura, empleo rural y sostenibilidad productiva en Colombia.,» *Revista Colombiana de Economía Agraria,* pp. 45-63, 2024.

[28] J. Meza, «Distribución y uso del suelo agrícola en Colombia: desafíos y oportunidades para el sector cafetero.,» Universidad Nacional de Colombia., Bogotá D.C, 2022.

[29] L. A. Terán, «Café especial y sostenibilidad territorial: dinámicas productivas en el departamento del Huila.,» *Revista de Estudios Agrarios y Rurales,* pp. 77-95, 2024.

[30] L. F. Gómez-Ramírez y M. A. Ríos-Botero, «Distribución territorial de la caficultura y su incidencia en el desarrollo rural colombiano.,» *evista Colombiana de Estudios Agrarios,,* pp. 55-78, 2023.

[31] J. Ovalle-Másmela, F. Romero-Perdomo y C. Uribe-Galvis, «Tecnologías emergentes para el agro y su aplicación en Colombia,» *Corporación Colombiana de Investigación,* 2023.

[32] C. C. Muñoz, C.-A. Cobos y J.-F. Muñoz, Predicción del rendimiento de cultivos de café: un mapeo sistemático, vol. 25, Ingeniería y competitividad, 2023.

[33] F. Ceballos-Sierra y S. Dall'Erba, The effect of climate variability on Colombian coffee productivity: A dynamic panel model approach., vol. 190, Agricultural Systems, 2021.

[34] J. Ríos y D. Saavedra, «Determinantes agroclimáticos y tecnológicos de la productividad del café en sistemas de pequeña escala en Colombia.,» *Revista de Economía y Desarrollo Rural,* pp. 33-52, 2022.

[35] C. L. Hunt y J. R. Ortmann, «Artificial intelligence and machine learning in global agriculture: Emerging trends, applications, and challenges.,» *Journal of Agricultural Systems Innovation,* pp. 155-178, 2023.

[36] A. R. Khamis, J. F. López-Morales y R. K. Gupta, «Data-driven agriculture: Remote sensing, machine learning and decision support systems for crop monitoring and management.,» *Precision Agriculture Advances,* pp. 45-72, 2024.

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
[37] L. Yu, D. M. Hernández-Soto y J.-H. Park, «AI-enabled resource efficiency and supply chain traceability in modern agriculture: Impacts on water use, fertilization and production analytics.,» *nternational Journal of Agricultural Data Science,* pp. 112-134, 2023.

[38] C. Zamudio, «Aplicaciones de inteligencia artificial en la agricultura: avances, desafíos y oportunidades para sistemas agrícolas emergentes.,» *Journal of Global Agricultural Technology,* p. 245–268, 2023.

[39] W. Liu y E.-R. Hartman, «Machine learning for sustainable crop production: Integrating environmental variability, predictive analytics and adaptive management.,» *International Review of Agricultural Informatics,* pp. 51-79, 2024.

[40] J. Triana, A. Álvarez y G. Castellanos, «Enhancing agricultural data interpretability and visualization with TabNet-driven feature extraction and Local Biplots,» *Results in Engineering,* vol. 16, 2025.

[41] L. Talero-Sarmiento, S. Roa-Prada, L. Caicedo-Chacon y O. Gavanzo-Cardenas, «A Data-Driven Approach to Improve Cocoa Crop Establishment in Colombia: Insights and Agricultural Practice Recommendations from an Ensemble Machine Learning Model.,» *AgriEngineering,* vol. 7, nº 1, 2025.

[42] D. Presta-Novello, N. Salazar-Camacho, L. Delgadillo-Mirquez, H. Hernández-Sarabia y M. d. P. Álvarez-Bustos, «Sustainable Development in the Colombian Post-Conflict—The Impact of Renewable Energies in Coffee-Growing Women,» *Sustainability,* vol. 15, nº 2, 2023.

[43] C4IR Colombia – Centro para la Cuarta Revolución Industrial, «Fomento a la adopción de tecnologías de la cuarta revolución industrial en emprendimientos agro 4.0.,» C4IR Colombia – Centro para la Cuarta Revolución Industria, Bogota, 2021.

[44] F. Costa, S. Frecassetti, M. Rossini y A. Portioli, «Industry 4.0 digital technologies enhancing sustainability: Applications and barriers from the agricultural industry in an emerging economy.,» *Journal of Cleaner Production,* vol. 408, 2023.

[45] IDEAM, «Boletín climatológico nacional 2023.,» IDEAM, Bogotá D.C, 2023.

[46] Unidad de Planificación Rural Agropecuaria (UPRA), «Informe técnico sobre vocación y uso del suelo en Colombia,» Ministerio de Agricultura y desarrollo rural, 2025, 2025.

[47] Departamento Administrativo Nacional de Estadística, «Estadísticas del sector agropecuario y precios de insumos agrícolas 2024,» DANE, Bogotá D,C, 2024.

[48] D. Pérez, D. Manrique, J. Triana, A. Álvarez y G. Castellanos, «An Explainable Framework Integrating Local Biplots and Gaussian Processes for Unemployment Rate Prediction in Colombia,» *Computation ,* 2025.

[49] G. Rambauth, «Agricultura de Precisión: La integración de las TIC en la producción Agrícola Precision,» *Theory Applied.,* vol. 3, nº 1, pp. 34-38, 2022.

[50] A. Giraldo, Tan cerca y tan lejos de la agricultura 4.0 en Colombia, vol. 55, Medellín: Universidad EAFIT, 2020, pp. 75-85.

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
[51] J. Botero, V. García, A. Valencia, J. Valencia, E. Reyes, M. Mejía y R. Hernández, «Machine Learning in Sustainable Agriculture: Systematic Review and Research Perspectives,» *Agriculture,* 2025.

[52] Leonardo Talero, «An optimization model for Theobroma cacao yield maximization based on smart farming technologies,» UNAB, Bucaramanga, 2024.

[53] A. Kumar, A. Singh-Gill, A. Sharma y V. Kumar, «Enhancing the Explainability and Interpretability of Crop Yield Prediction Models Through Precision Agriculture,» *Agricultural-Centric Computation,* 2025.

[54] J. Arévalo, F. Flor, J. Latorre, R. Tino, E. Martínez y J. Blanco, «AI Algorithms in the Agrifood Industry: Application Potential in the Spanish Agrifood Context,» *Applied sciences,* 2025.

[55] M. Huang, X.-S. Zhang, A. Bhatti, Y. Wua, Y. Zhang y Y. Ghadi, «An interpretable approach using hybrid graph networks and explainable AI for intelligent diagnosis recommendations in chronic disease care,» *Biomedical Signal Processing and Control,* 2024.

[56] T. Tasioulis y K. Karatzas, «Reviewing Explainable Artificial Intelligence Towards Better Air Quality Modelling,» *Advances and New Trends in Environmental Informatics 2023,* 2023.

[57] D. Hong, W. He, N. Yokoya, j. yao, L. Gao y L.-P. Zhang, «Interpretable Hyperspectral Artificial Intelligence: When nonconvex modeling meets hyperspectral remote sensing,» *IEEE Geoscience and Remote Sensing Magazine,* 2021.

[58] V. Belle y I. Papantonis, «Principles and Practice of Explainable Machine Learning,» *Frontiers,* 2021.

[59] D. Iakovidis, Y. Gadanakis, J. Campos y J. Park, «Optimising decision support tools for the agricultural sector,» *Environment, Development and Sustainability,* 2025.

[60] D. Paudel, A. d. Wit, H. Boogaard, D. Marcos, S. Osinga y I. Athanasiadis, «Interpretability of deep learning models for crop yield forecasting,» *Computers and Electronics in Agriculture,* 2023.

[61] A.-A. Wani, «Comprehensive review of dimensionality reduction algorithms: challenges, limitations, and innovative solutions,» *Computer Science,* 2025.

[62] C. Worel, «The Curse of Dimensionality: When Too Many Features Break Your Machine Learning Model,» *SSRN,* 2023.

[63] M. Semenov y J. Porter, «Climatic variability and the modelling of crop yields,» *Agricultural and Forest Meteorology,* 1995.

[64] A. Dixit y S. Jain, «Contemporary Approaches to Analyze Non-stationary Time-series: Some Solutions and Challenges,» *Recent Advances in Computer Science and Communications,* 2023.

[65] D. A. Pérez-Rosero, D. A. Manrique-Cabezas, J. C. Triana-Martinez, A. M. Álvarez-Meza y G. Castellanos-Dominguez, «An Explainable Framework Integrating Local

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
Biplots and Gaussian Processes for Unemployment Rate Prediction in Colombia.,» *Computation,* vol. 13, nº 5, 2025.

[66] J. Wang, Y. Wang y Z. Qi, «Remote Sensing Data Assimilation in Crop Growth Modeling from an Agricultural Perspective: New Insights on Challenges and Prospects,» *Agronomy,* 2024.

[67] I. Zarza, J. Curto y C. Calafate, «UMAP for Geospatial Data Visualization,» *Procedia Computer Science,* 2023.

[68] B. Kim, J. Park y J. Suh, «Transparency and accountability in AI decision support: Explaining and visualizing convolutional neural networks for text information,» *Decision Support Systems,* 2020.

[69] A. Rowan, «Unravelling black box machine learning methods using biplots,» OpenUCT, París , 2019.