Environmental Science

https://www.aimspress.com/journal/environmental

AIMS Environmental Science, 11(5): 759–775.
DOI: 10.3934/environsci.2024038
Received: 24 June 2024
Revised: 08 August 2024
Accepted: 28 August 2024
Published: 10 September 2024

Research article

Design of a neuro-fuzzy model for agricultural employment in Colombia using fuzzy clustering

Juan Sánchez[1,*], Juan Rodríguez[2] and Helbert Espitia[1]

[1] Facultad de Ingeniería, Universidad Distrital Francisco José de Caldas, Bogotá, Colombia
[2] Facultad del Medio Ambiente y Recursos Naturales, Universidad Distrital Francisco José de Caldas, Bogotá, Colombia

* **Correspondence:** Email: jmsanchezc@udistrital.edu.co; Tel: +573112172889.

Abstract: High levels of poverty in rural areas constitute one of the main challenges for developing countries. Since agricultural employment is the main source of income in these areas, the design of tools that simulate and help public policymakers will be remarkably useful. This work proposes the development of a model for agricultural employment in Colombia, considering input variables such as education, contract, and income, and the output is the amount of agricultural employment. Real data measured in Colombia are used for the design and adjustment of the model. To design the fuzzy system for an agricultural employment model, the methods employed are fuzzy C-means clustering and neuro-fuzzy systems. The systems were tested with different cluster configurations, and a fuzzy system was obtained with an adequate distribution of the fuzzy sets and the respective rules that relate the sets. It was observed that as the clusters increase, the adjustment function decreases. The implementation of neuro-fuzzy systems to model agricultural employment will allow public policymakers to generate guidelines that adjust to their political agendas with a lower degree of uncertainty.

Keywords: agricultural employment; fuzzy clustering; model; neuro-fuzzy; public policymakers

1. Introduction

The Sustainable Development Goals promulgated by the United Nations on September 25, 2015, set forth 17 goals to protect the planet, mitigate poverty, and ensure prosperity for all [1]. Poverty is one of the biggest problems in developing countries such as Colombia. Based on the data reported [2], Figure 1 shows the historical behavior of multidimensional poverty from 2010 to 2022. There is no data on multidimensional poverty in rural areas for 2017.

One of the main metrics for assessing poverty in Colombia is the Multidimensional Poverty Index (MPI) created by the National Administrative Department of Statistics (Departamento

Administativo Nacional de Estadística - DANE) of Colombia, which measures the access of individuals to certain characteristics considered vital. The index includes five dimensions: Household educational conditions, childhood and youth conditions, work, health, and access to domiciliary public services and housing conditions. The index comprises 15 indicators, and households deprived of 33% of these indicators are considered multidimensionally poor [3].

Figure 1 illustrates a decrease in multidimensional poverty, but a notable disparity persists between rural and urban areas. In 2022, multidimensional poverty in urban areas was 8.7% while in rural areas it was 27.3%. This represents a substantial percentage difference of 18.6%. Therefore, it is extremely significant for governments to understand the behavior of poverty in rural areas, with the examination of employment trends in these areas being a key factor.

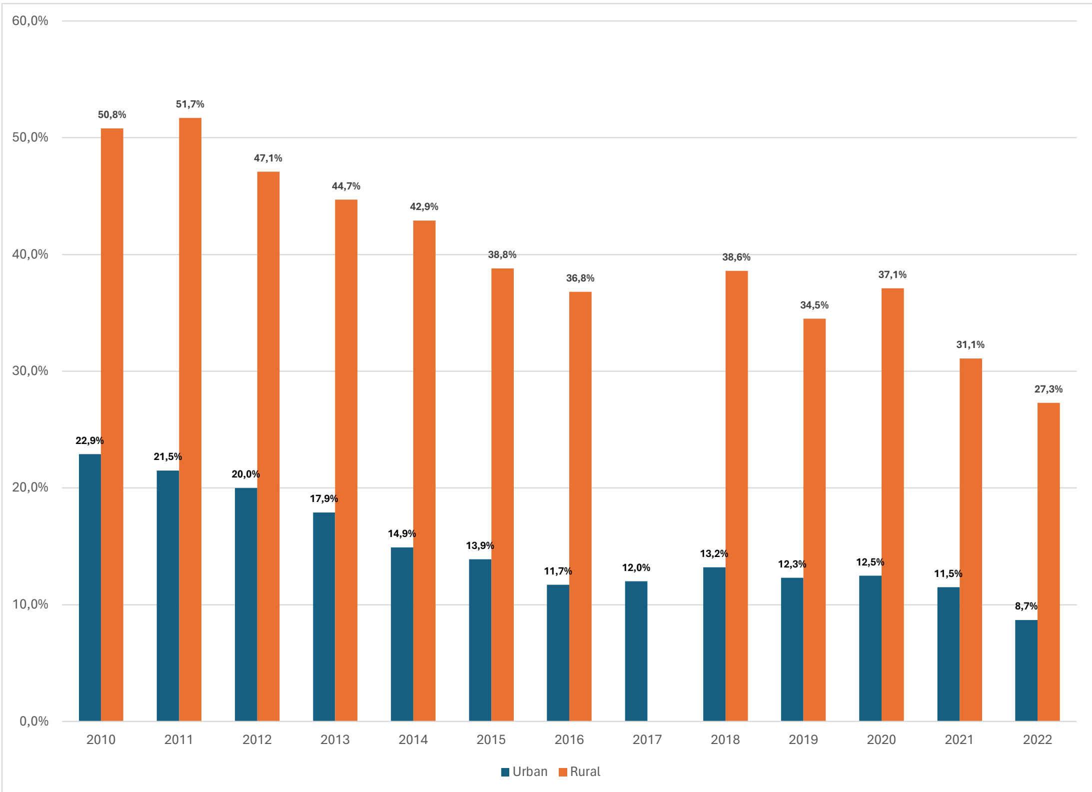


Figure 1. Multidimensional poverty 2010–2022 [2].

In Colombia, the principal source of employment in rural areas is the agricultural sector. This recognition stems from the country's robust natural resources and diverse climatic conditions, attributable to its location in the tropical zone and geographical features such as extensive plains and mountain formations. These geographical attributes contribute to the presence of varied altitudes and climatic zones. Considering the data reported [4], Figure 2 shows the historical series of agricultural employment participation at the national level in Colombia.

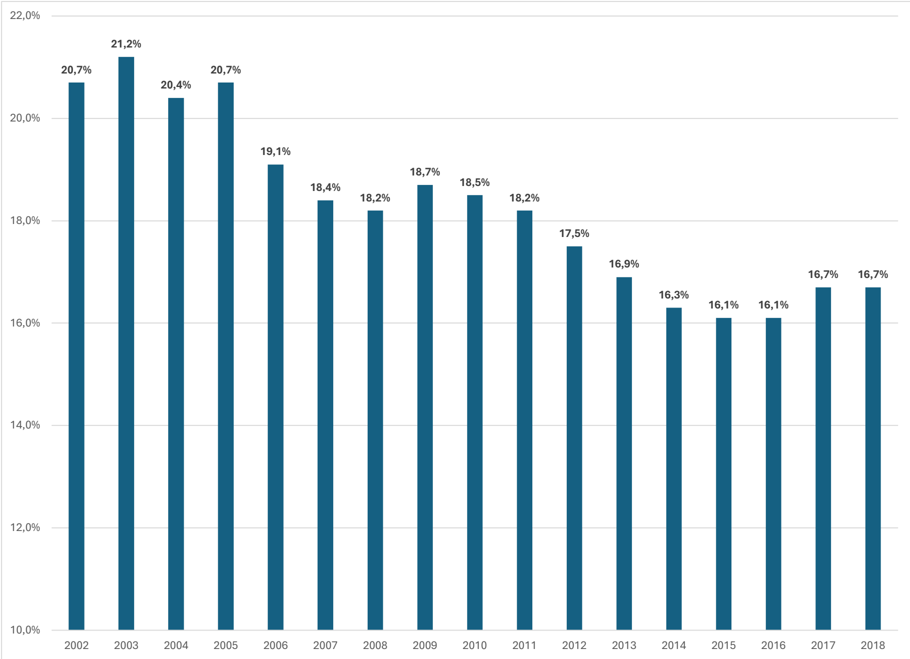


Figure 2. Agricultural employment participation at the national level [4].

As shown in Figure 2, the percentage of agricultural employment in Colombia has decreased significantly in the last decades. In 2002, it stood at 20.7%, and by 2018 it had decreased to 16.7%. Numerous studies have analyzed the reasons why agricultural employment has declined. First, land tenure has been a subject of investigation, revealing that concentrated ownership of arable land diminishes the generation of agricultural employment, whereas communal tenure fosters such employment [5–7]. Furthermore, advancements in agricultural production have resulted in reduced demand for farm labor [8–10]. The educational attainment of the population is a crucial factor; research indicates that a higher level of education among the children of rural families leads them away from agricultural activities, prompting them to pursue non-agricultural jobs in urban areas [11–13].

In rural areas, new non-agricultural employment opportunities have emerged, presenting more attractive wages in comparison to agricultural jobs [14–16]. Finally, job stability within the agricultural sector is lower, prompting individuals to opt for employment in other sectors that offer greater stability [15–17].

As seen, it is necessary to propose a model that allows determining the amount of agricultural employment in Colombia. Considering the related studies presented previously, the decision was to incorporate three variables into the model: level of education, wage income, and employment contract type.

1.1. Related works

In the work of Akopov et al. [18], using clustering approaches, the Parallel Real-Coded Genetic Algorithm (P-RCGA) for cluster-based optimization was applied to solve issues related to evacuation processes. The proposed P-RCGA is rooted in the dynamic interchange of optimal decisions among a global population by means of dispersed processes with distinct individual features. The authors employed the algorithm for an emergency behavior simulation produced by human agent-rescuers through the use of objective functions in order to optimize the evacuation process.

A fuzzy K-means clustering-based optimization method was developed in Zhang and Zhang' work [19]. The algorithm was utilized for the optimization of regional economic industrial structure considering the multisubject of regional industrial planning. The authors considered an industrial planning model based on regional economic differences by analyzing various industrial cities. In this order, three models were proposed to study the development law of the industrial frame: A diffusion evolution model, an unbalanced evolution model, and an alternative evolution model. By examining and categorizing the numerous aspects that have influenced the development of urban industrial structure, the optimization model of industrial structure was created.

In another study, researchers collected over 1000 diffuse reflectance Fourier transform (DRIFT) mid-infrared spectra of agricultural soils from the West African savanna zone and clustered the data using K-means and fuzzy K-means (FKM) algorithms. The objective was to explore the feasibility of centroid-based clustering algorithms for identifying substructures in spectral data. A two-cluster pattern emerged, dividing the dataset into northern and southern regions. The FKM algorithm successfully identified a transition zone between the two clusters, which was not detectable with K-means. This transition zone was explained by a gradual change in aeolian dust deposition, topography, and geology [20].

A study conducted in Jianyang, China, evaluated effective soil nutrient management. Researchers collected 100 georeferenced soil samples from a depth of 0–20 cm. The analysis revealed that coefficients of variation (CV) of soil properties ranged widely, from low (1.132%) to moderate (45.748%). Ordinary kriging and semivariogram analysis demonstrated varying patterns of spatial variability for the studied soil properties, with spatial dependence ranging from weak to strong. Management zones were delineated using principal component analysis (PCA) and fuzzy K-means clustering. The soil properties significantly differed among the management zones. Consequently, the methodology used for delineating management zones could be effectively applied for site-specific soil nutrient management, avoiding soil degradation, while maximizing crop production in the study area [21].

In Iran, researchers employed principal component analysis and fuzzy C-means clustering methods to delimit soil management zones for sustainable production, enhanced soil management, and increased economic benefits in commercial citrus plantations. They analyzed biological and soil attributes, along with physicochemical properties, for the delimitation of these management zones. Additionally, an economic analysis was conducted based on the management zone results, determining changes in each zone using a relative cost (RC) value [22].

The studies presented demonstrate the effectiveness of fuzzy clustering algorithms in enhancing data classification, leading to more efficient processes. Their application in the agricultural sector proves to be particularly useful and aligns well with the objectives of this study.

1.2. Approach and organization

This work aims to design a neuro-fuzzy system for agricultural employment in Colombia using real data and fuzzy clustering; thus, the purpose is to have a fuzzy model that fits the data set by preliminarily obtaining the sets of the fuzzy system using the clustering process. In this way, the aim is to obtain a model composed of rules that can later be used in the formulation of policies to improve the quality of life of rural workers.

The rest of the paper is organized as follows. Section 2 displays the methods employed: First the fuzzy C-means clustering and then Takagi-Sugeno fuzzy systems and the neuro-fuzzy systems. Later, in Section 3, the implementation and results are presented. Finally, in Sections 4 and 5, the discussion and conclusions are presented.

2. Methods

This section describes the techniques employed: The fuzzy C-means (FCM) clustering approach, and the Takagi-Sugeno type fuzzy systems for implementing the neuro-fuzzy agricultural model.

In the first stage, the FCM algorithm is used to determine the fuzzy sets associated with the inputs and the initial configuration of the fuzzy system. Subsequently, the parameters are adjusted with the neuro-fuzzy system training algorithm. In this work, standard versions of the fuzzy C-means algorithm and the neuro-fuzzy system training process are utilized.

2.1. Fuzzy C-means clustering

In the fuzzy C-means approach, each datum belongs to a group in a limited extension delineated by a membership degree [23–25]. The procedure considers the degree of belonging of a datum to a cluster where overlap between clusters may occur, which can be adjusted in the procedure to determine the clusters.

Using clustering, it is possible to build a fuzzy inference system (FIS) by generating membership functions to depict the linguistic concepts associated with each group. Varied clustering metrics can be employed to set the optimal fuzzy partition of X. The one most extensively employed is related to the least square error value depicted in Eq 2.1.

$$J_m = \sum_{k=1}^{n} \sum_{i=1}^{c} (\mu_{ik})^m \|x_k - r_i\|_A^2 \tag{2.1}$$

In Eq 2.1 the part $\|x_k - r_i\|_A^2$ determines the square distance between the centers of the groups $(r_1, r_2, \ldots, r_c)$ and the datum $(x_1, x_2, \ldots, x_n)$. In this order, $\|\cdot\|_A$ is the norm induced by the weight matrix A (positive definite). Taking A as the identity, the norm corresponds to the square of the Euclidean distance, as is calculated in Eq 2.2.

$$\|x_k - v_i\|_A^2 = (x_k - r_i)^T A (x_k - r_i) \tag{2.2}$$

Having the i-th cluster and the k-th datum in this cluster, the coefficient $(\mu_{ik})^m$ is the m-th power of the associated membership value. Taking $m > 1$, this value moderates the fuzzy overlapping [23–25].

Algorithm 1 displays the steps employed in the fuzzy C-means clustering process. First, the cluster centers are selected randomly, and the fuzzy membership matrix is initialized. Then, until the change of the centers is less than a determined value ε, the center vectors and the membership matrix are updated repeatedly.

Algorithm 1 Fuzzy C-means clustering algorithm.

Require: Number of clusters c, and fuzzification coefficient $m > 1$

1: Randomly select the clusters, centers r_i

2: Initialize the fuzzy membership matrix μ_{ik}

3: **repeat**

4: Calculate the center vectors r_i

5: Update the fuzzy membership matrix μ_{ik}

6: **until** the change of the centroids is less than a given value ε

2.2. Takagi-Sugeno fuzzy systems

A fuzzy inference system (FIS) allows for the inclusion of information based on linguistic rules in its structure. The Takagi-Sugeno structure is useful when developing a systematic approach to prompt fuzzy rules [26]. The structure of a Takagi-Sugeno model for a standard fuzzy rule shows the following structure:

$$\text{If } X_1 \text{ is } A_1 \text{ and } X_2 \text{ is } A_2, \text{ then } Y = f(X_1, X_2)$$

In this rule, A_1 and A_2 represent the fuzzy sets in the antecedent (inputs), while in the output, $Y = f(X_1, X_2)$ is a polynomial of X_1 and X_2. Consequently, the structure of the consequent as follows

- Zero-order Takagi-Sugeno model, $f = A_0$.
- First-order Takagi-Sugeno model, $f = A_0 + A_1X_1 + A_2X_2$.

Such a system is applicable when implementing the Adaptive Neuro-Fuzzy Inference System (ANFIS). The parameters of both the membership and output functions are adapted as they happen in a neural network.

The output FIS result $Y_s[k] = F(X[k])$ is calculated utilizing the input data $X[k]$. Then, the system parameters are adjusted using the error function 2.3, where the difference with the real output data $Y[k]$ is calculated.

$$J = \frac{1}{2}\left[Y[k] - F\left(X[k]\right)\right]^2 \tag{2.3}$$

Employing the derivatives of Eq 2.3, the adjustment of the fuzzy system parameters is performed until the value of the root mean square error (RMSE) given in Eq 2.4 is less than a given value ε. In Eq 2.4, N corresponds to the total number of data utilized for training. Algorithm 2 displays the steps employed for the neuro-fuzzy training process, where the output of the fuzzy system is calculated by using the data inputs, followed by the calculation of the error function and the respective derivatives. Then, the values of the fuzzy system parameters are updated.

$$RMSE = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left[Y[k] - F\left(X[k]\right)\right]^2} \tag{2.4}$$

Algorithm 2 Neuro-fuzzy training process.

Require: First configuration of the neuro-fuzzy system
1: Load training data
2: **repeat**
3: Using the data inputs, compute the output of the fuzzy system
4: Calculate the error function and the respective derivatives
5: Update fuzzy system parameters
6: **until** RMSE is less than a given value ε

3. Implementation and results

This section first presents the results of the clustering process and then the fuzzy logic systems obtained with the clustering process. Clustering with 2, 3, 4, and 5 groups is considered. It is important to consider that a high number of clusters increases the number of fuzzy sets, which can increase the complexity of the fuzzy system and thus also lower its interpretability.

The data for the dependent and independent variables were obtained from the National Quality of Life Survey conducted by the National Administrative Department of Statistics (Departamento Administrativo Nacional de Estadistica - DANE) in Colombia, covering the years 2010 to 2022. The percentage of the population with primary education was employed to determine the education level variable, while the average salary of agricultural employees was used for the salary income variable. In addition, the percentage of agricultural employees with verbal contracts was also considered. The data were entered into the computational model using the min-max normalization technique, as shown in Eq 3.1.

$$normalized\ value = \frac{value - minimum}{maximum - minimum} \tag{3.1}$$

The data employed are displayed in Figure 3 (with normalized values). The respective inputs and outputs considered are the following:

- Input 1 (X_1): rural elementary education.
- Input 2 (X_2): rural workers with verbal contracts.
- Input 3 (X_3): average income of rural worker.
- Output (Y): amount of agricultural employment.

In this way, the system consists of three inputs associated with the characteristics of rural workers (elementary education, verbal contracts, average income). The output is associated with the amount of agricultural employment. It should be noted that the variables are on a normalized scale from 0 to 1. For the clustering process only the inputs are utilized. Conversely, the training of the neuro-fuzzy system requires employing inputs and outputs.

The implementation was made in MATLAB version 2017a using the Fuzzy Logic toolbox. The function “fcm” is employed for the cluster process; meanwhile, the “genfis” function is utilized to create the neuro-fuzzy system via the clustering method. Finally, the neuro-fuzzy training is performed employing the “anfis” function.

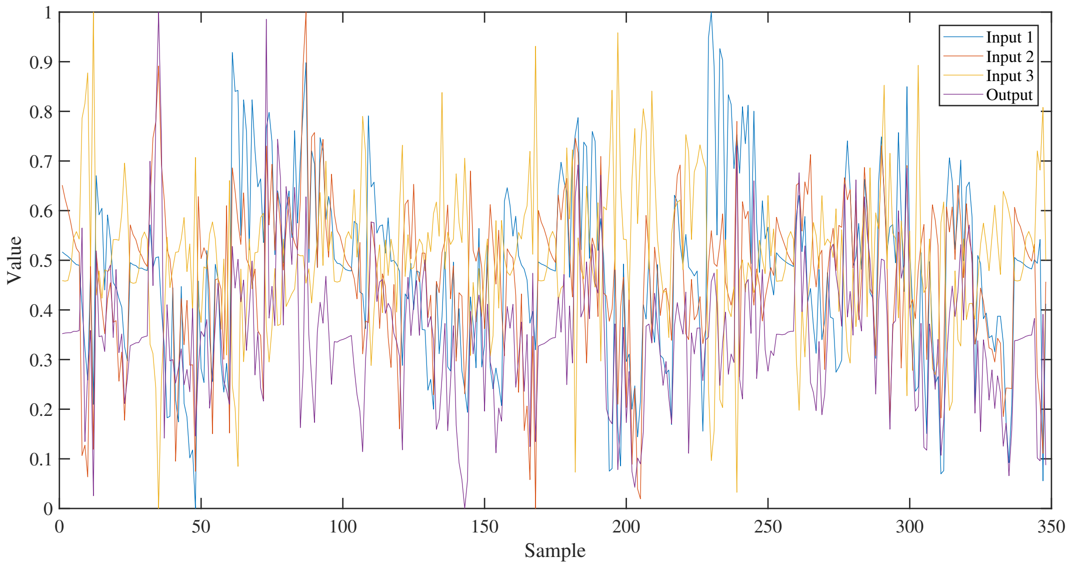


Figure 3. Data employed.

3.1. Clustering results

The formation of 2, 3, 4, and 5 groups is considered for the clustering process. Thus, the respective values of the objective function J (Eq 2.1) obtained for each case are as follows

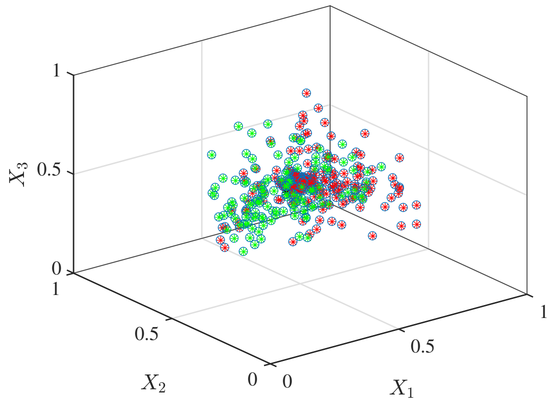
- Using 2 clusters: $J = 11.804343$.
- Using 3 clusters: $J = 7.151393$.
- Using 4 clusters: $J = 5.268694$.
- Using 5 clusters: $J = 4.119224$.

It should be noted that increasing the number of clusters decreases the value of the fitness function. It should be kept in mind that increasing the number of clusters can reduce the interpretability of the fuzzy logic system.

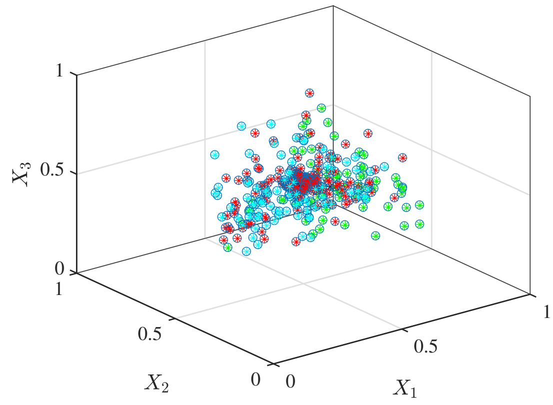
The results of the clusters can be seen in Figure 4a for 2 groups, in Figure 4b for 3 groups, in Figure 4c for 4 groups, and in Figure 4d for 5 groups. It should be noted that in all cases a uniform distribution of the elements in the cluster is achieved.

The clustering process makes it possible to identify whether there is an adequate organization of the data that allows a relationship between them and, in this way, build a fuzzy logic system based on the clusters found.

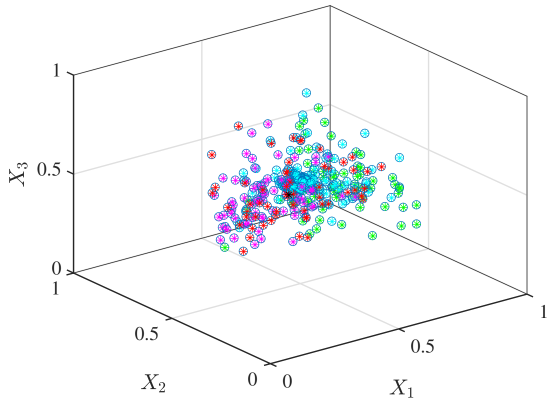
It should be noted that the majority of groups are presented for intermediate values, which can be reflected later in the construction of the fuzzy sets. The clusters found allow one to establish the fuzzy sets and also the inference rules of the fuzzy logic system.



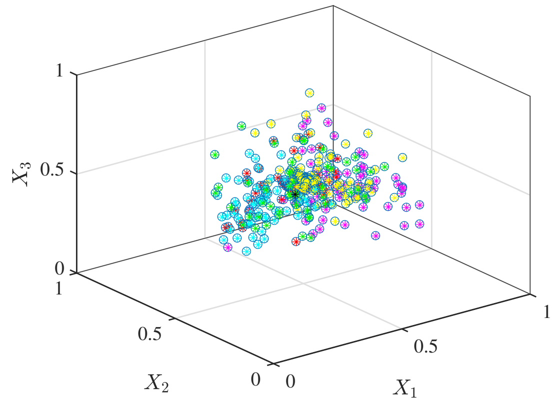
(a) Cluster formation for 2 groups.



(b) Cluster formation for 3 groups.



(c) Cluster formation for 4 groups.



(d) Cluster formation for 5 groups.

Figure 4. Cluster formation for 2, 3, 4, and 5 groups.

3.2. FIS results

After carrying out the clustering process and observing that the clusters allow adequate grouping of the data, the next step is to perform the training of the neuro-fuzzy system (Takagi-Sugeno).

In this order, the neuro-fuzzy systems are made considering 2, 3, 4, and 5 fuzzy sets in each input (associated with the clusters formed). The respective Root mean squared error (RMSE) values obtained after training for the different fuzzy systems considered are as follows

- Case of 2 fuzzy sets: $RMSE = 0.107196$.
- Case of 3 fuzzy sets: $RMSE = 0.100078$.
- Case of 4 fuzzy sets: $RMSE = 0.099986$.
- Case of 5 fuzzy sets: $RMSE = 0.098025$.

It should be noted that the value of the RMSE decreases as the number of fuzzy sets increases; however, having a high value of fuzzy sets can increase the complexity of the system, decreasing its interpretability.

Regarding the configuration using five membership functions, Figure 5 displays the structure of the fuzzy system (composed of five rules), with the respective inputs X_1, X_2, X_3, and the output Y.

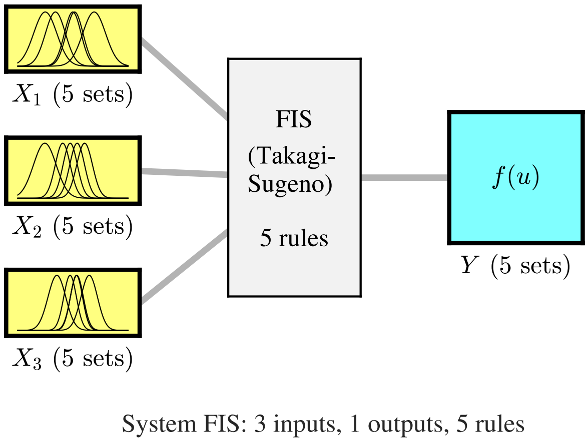
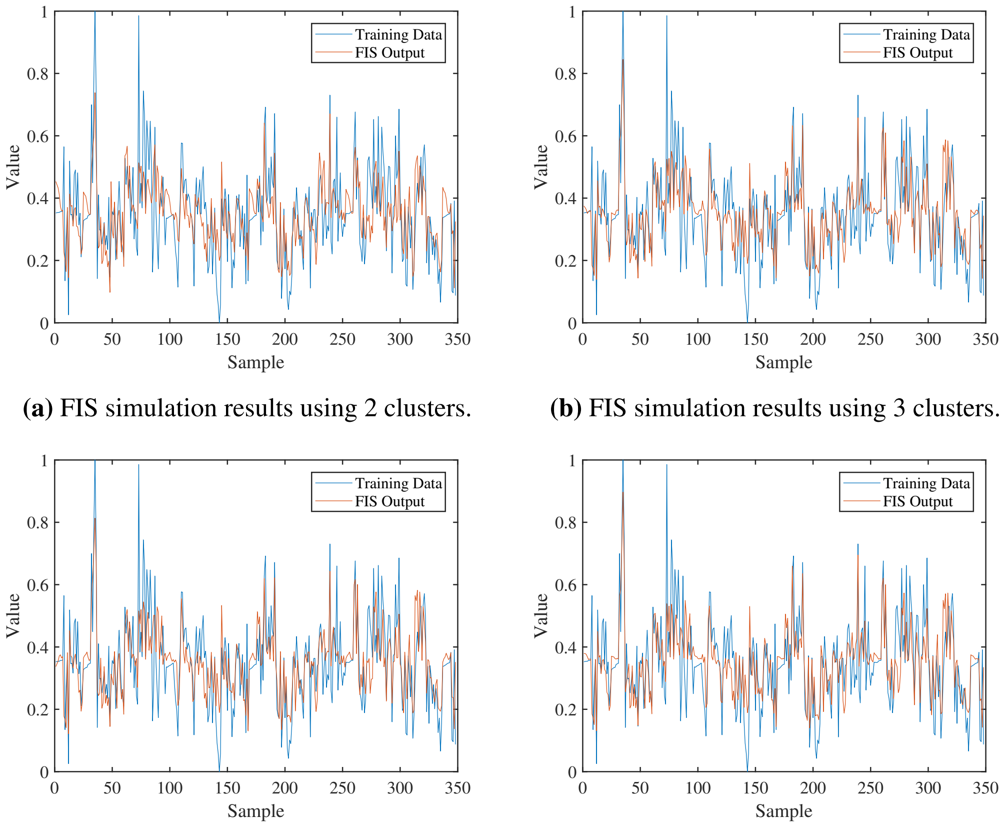


Figure 5. Fuzzy inference system.

The simulation results of the fuzzy inference systems can be seen in Figure 6a for 2 groups, in Figure 6b for 3 groups, in Figure 6c for 4 groups, and in Figure 6d for 5 groups. It should be noted that as the number of clusters increases, a better fit to the real data is obtained.



(a) FIS simulation results using 2 clusters.

(b) FIS simulation results using 3 clusters.

(c) FIS simulation results using 4 clusters.

(d) FIS simulation results using 5 clusters.

Figure 6. FIS simulation results for different configurations.

Regarding the complexity of the fuzzy systems the number of fuzzy sets is equal to the number of rules, that is, as the fuzzy sets increase, the complexity of the system also increases. In this case, 2, 3, 4, and 5 clusters (fuzzy sets) are used to have a moderate complexity of the system.

Taking the system with the best RMSE obtained (using 5 clusters), Figure 7 shows the fuzzy sets and the rules obtained. The set of rules is the following:

- If (X_1 is $C_{1,1}$) and (X_2 is $C_{2,1}$) and (X_3 is $C_{3,1}$), then (Y is f_1).
- If (X_1 is $C_{1,2}$) and (X_2 is $C_{2,2}$) and (X_3 is $C_{3,2}$), then (Y is f_2).
- If (X_1 is $C_{1,3}$) and (X_2 is $C_{2,3}$) and (X_3 is $C_{3,3}$), then (Y is f_3).
- If (X_1 is $C_{1,4}$) and (X_2 is $C_{2,4}$) and (X_3 is $C_{3,4}$), then (Y is f_4).
- If (X_1 is $C_{1,5}$) and (X_2 is $C_{2,5}$) and (X_3 is $C_{3,5}$), then (Y is f_5).

Figure 7 displays an as example case where inputs X_1, X_2, and X_3 are equal to 0.5 in a normalized scale. In this way, the respective output Y obtained is 0.365. As seen, there is a large number of fuzzy sets associated with average values of the inputs. This figure shows the relationship between the membership functions for each input and the constant output function, thus achieving interpretability of the concepts (linguistic labels) and rules that compose the fuzzy logic system. As an example of the inference process, in the case considered in Figure 7, the rules 2 and 3 present the greatest contribution to calculate the output value.

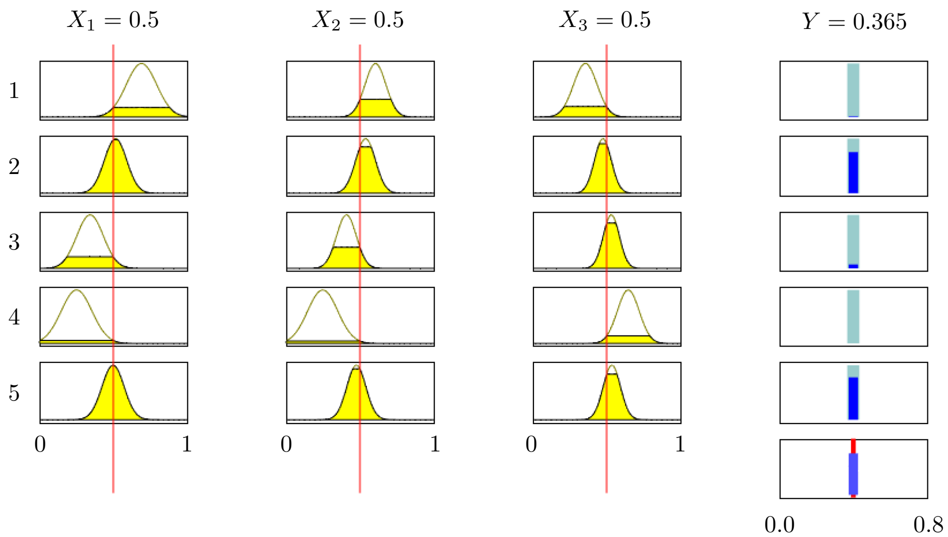
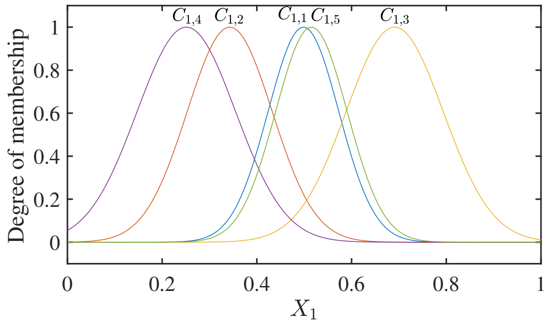
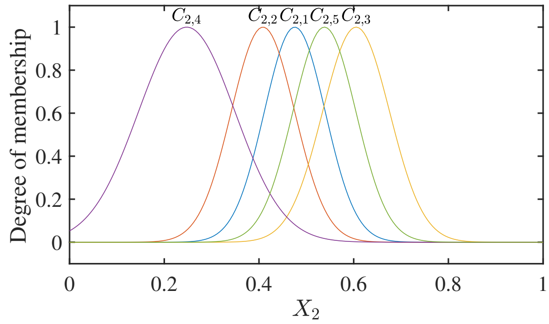


Figure 7. Fuzzy sets and rules.

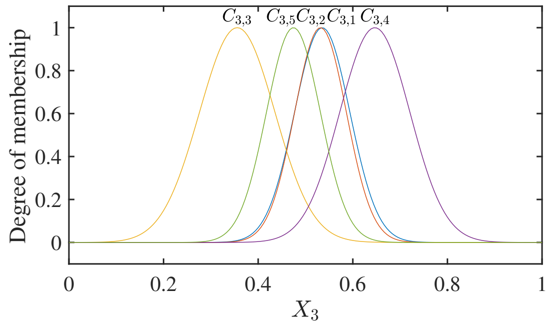
The respective fuzzy sets obtained are displayed in Figure 8: For X_1 in Figure 8a, for X_2 in Figure 8b, and for X_3 in Figure 8a. As shown, there is a large number of fuzzy sets associated with center values of the inputs. It is also worth noting that the organization of the fuzzy sets obtained for each input is different. This organization is given by the clusters found, which also allows for the fuzzy system rules, formulation.



(a) Associated fuzzy sets to input X_1.



(b) Associated fuzzy sets to input X_2.



(c) Associated fuzzy sets to input X_3.

Figure 8. Associated fuzzy sets to inputs X_1, X_2, and X_3.

Figure 8a displays the resulting fuzzy sets for input X_1 (rural elementary education). The sets, denoted from $C_{1,1}$ to $C_{1,5}$, use the first digit to represent the input (in this case, 1), and the second digit indicates the nominal number assigned to the fuzzy set. While there are five sets, $C_{1,3}$ and $C_{1,5}$ are nearly identical, allowing for the same linguistic label. The set with the highest values is $C_{1,3}$, which can be labeled as "high". Sets $C_{1,3}$ and $C_{1,5}$ both encompass values around 0.5, justifying the label "medium". Set $C_{1,2}$ can be assigned the label "medium-low", and finally, the set $C_{1.4}$, containing the lowest values, can be labeled as "low".

Figure 8b illustrates the distribution of fuzzy sets for input X_2 (rural workers with a verbal contract). It appears slightly smoother than the X_1 input; however, the sets $C_{2,1}$, $C_{2,2}$, $C_{2,3}$, and $C_{2,5}$ exhibit some overlapping. The set $C_{2,4}$ can be labeled as "low", with $C_{2,2}$ as "medium-low", $C_{2,1}$ as "medium", and $C_{2,5}$ as "medium-high", and the fuzzy set $C_{2,3}$ can be assigned the label "high".

Figure 8c displays the five fuzzy sets for input X_3 (average income of a rural worker). The sets labeled $C_{3,1}$ and $C_{3,2}$ exhibit nearly identical distributions, allowing for the assignment of the same linguistic label. The fuzzy set $C_{3,4}$ can be labeled as "high", with sets $C_{3,1}$ and $C_{3,2}$ as "medium-high", and $C_{3,5}$ as "medium-low", and finally, the set $C_{3,3}$ can be labeled as "low".

Finally, Figure 9 displays the ANFIS structure for the FIS obtained from the clustering. This figure shows the connections of the inputs, the fuzzy sets, and the functions associated with the output. In this order, the clusters formed can be seen in the form of the connection of the neurons, along with the relationship between the clusters and the inference rules. In summary, this figure shows the structure of the fuzzy system as a neural network and thus the observation of the nodes where the information is processed.

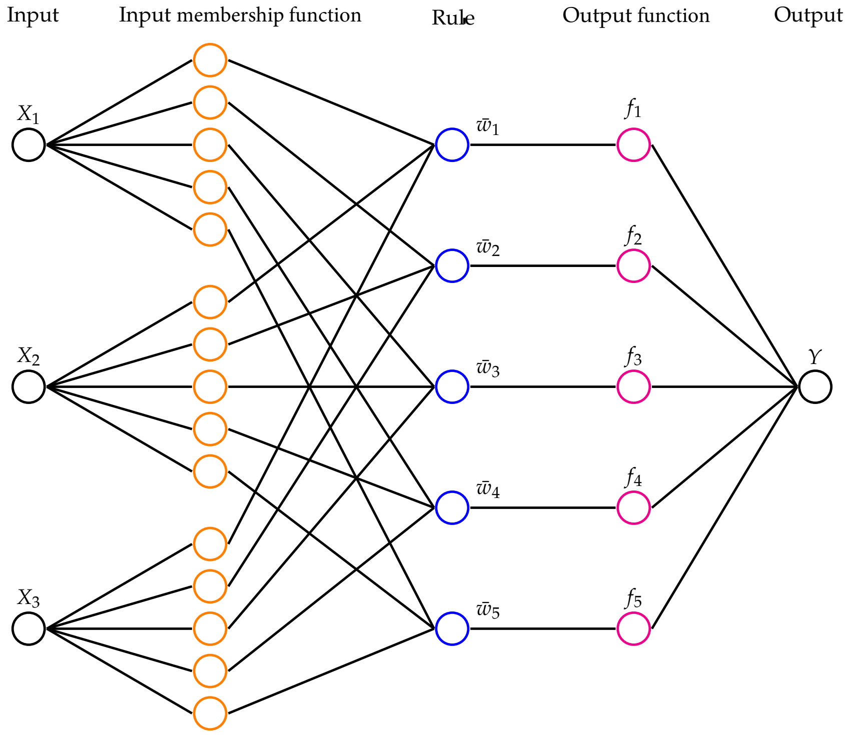


Figure 9. ANFIS structure.

3.3. Sensitivity analysis

In the fuzzy system, the concepts in the inputs are represented employing Gaussian membership functions given by Eq 3.2, where x is the value to evaluate, σ is the the standard deviation, and c the mean value (center). In this order, the Gaussian membership functions have a direct effect on the interpretability of the system, which is why a sensitivity analysis is performed by varying the value of σ, since this parameter is associated with the overlap of the membership functions.

$$\mu(x, \sigma, c) = e^{\frac{-(x-c)^2}{2\sigma^2}} \tag{3.2}$$

For the sensitivity analysis, we consider $\sigma = w\,\sigma_0$ where σ_0 is the previously obtained value (in the training process). The weighting factor w is considered with values 0.25, 0.5, 1.0, 1.5, and 2.0. The respective variation of parameter σ is made for all membership functions for each input separately. Table 1 contains the sensitivity analysis results, displaying the RMSE obtained for each case. In these results, the minimum RMSE is obtained with $\sigma = \sigma_0$, showing that the neuro-fuzzy system training process was successful. It is also observed that for all inputs (X_1, X_2, X_3) a greater error occurs for lower values of σ_0 (when the bell shape is narrower). This shows that the Gaussian membership functions representing the concepts in the fuzzy system must exhibit overlap in a way that the simulated data approximates the real data (see Figure 8 for the case of σ_0).

Table 1. Sensitivity analysis results.

Input	$0.25\,\sigma_0$	$0.5\,\sigma_0$	σ_0	$1.5\,\sigma_0$	$2.0\,\sigma_0$
X_1	0.145161	0.127389	0.098025	0.108899	0.115834
X_2	0.158072	0.125366	0.098025	0.103221	0.108482
X_3	0.150992	0.128374	0.098025	0.106198	0.117913

4. Discussion

The use of clustering to build the fuzzy logic system can be a suitable alternative since it allows having rules that relate the formed clusters. In this way, it can have a smaller number of rules compared to an approach where the combination of all the fuzzy sets is carried out to establish the fuzzy rules.

For the design of the fuzzy logic system, its interpretability must be considered in addition to achieving a suitable RMSE value. That is to say that both the fuzzy sets and the rules allow for a direct interpretation with the phenomenon to be modeled.

Based on the RMSE values obtained and the comparison between the simulated and real signal, it is observed that, as the number of fuzzy sets in the system inputs increases, the RMSE improves, and the approximation of the simulated signal to the real one becomes more accurate. This highlights the importance of this design parameter. However, a significant increase in the number of fuzzy sets also leads to an increase in the computational load and could hinder the interpretability of the system due to the excess of rules generated.

Since the interpretability of the fuzzy system is given by the shape of the membership functions, the sensitivity analysis carried out allows us to see the effect that the overlapping of the Gaussian functions has on the RMSE value. Regarding the results obtained, it is observed that the overlapping of the Gaussian functions must be maintained to achieve a data fit.

This tool proves beneficial for public policymakers since it allows one to visualize diverse rules presented as logical sentences. This visualization can provide valuable insights and serve as guidance for shaping public policies during the formulation process.

It is essential to clarify that the process of formulating public policy is inherently complex, as articulated by the garbage can model. This model posits that the constituents of the process (problems, the solutions, the participants, and the opportunities) are intermingled within a system characterized as organized anarchy [27]. Furthermore, policymakers in public policy are constrained by cognitive, temporal, and information limitations [28]. Therefore, a tool that helps improve the decision-making process becomes highly valuable.

Furthermore, it is essential to consider that this tool can be aligned with the policymaker's political agenda. The system enables the display of various conditions, represented as rules, empowering the formulators to discern which of these rules align with their political agendas.

5. Conclusions

The clustering process carried out preliminarily allows us to observe the possibility of having an adequate division of the input data. In most cases, adequate segmentation of the data is observed.

When training the fuzzy logic system using real data, it is observed that a suitable fit is achieved, which indicates that the formed clusters allow the generation of the rules of the fuzzy system adequately.

Increased clustering of fuzzy sets at the inputs improve the fit of the simulated signal to the real one. However, this may result in an increase in computational load and hinders the interpretability of the system.

The sensitivity analysis showed that the optimization process of the neuro-fuzzy system was satisfactory, since when performing the variation of the fuzzy sets, the lowest RMSE value is obtained with the values from the training process.

The fuzzy logic system obtained is the Takagi-Sugeno type; thus, in a subsequent work, extension to a Mamdani-type system can be considered to have fuzzy sets in the output (consequent) and in this way obtain a higher interpretability of the fuzzy logic system.

The implementation of neuro-fuzzy systems to model agricultural employment will allow public policymakers to generate guidelines that adjust to their political agendas with a lower degree of uncertainty.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The authors express gratitude to the Universidad Distrital Francisco José de Caldas.

Conflict of interest

The authors have no conflict of interest to declare.

References

1. ONU. Portada-desarrollo sostenible, 2015. Available from: https://www.un.org/sustainabledevelopment/es/.
2. DANE. Pobreza multidimensional, 2023. Available from: https://www.dane.gov.co/index.php/estadisticas-por-tema/pobreza-y-condiciones-de-vida/pobreza-multidimensional.
3. DANE. Boletín técnico (pobreza multidimensional en colombia, 2022. Available from: https://img.lalr.co/cms/2021/09/03041930/boletin-tec-pobreza-multidimensional-20.pdf.
4. DANE. Empleo y desempleo, 2024. Available from: https://www.dane.gov.co/index.php/estadisticas-por-tema/mercado-laboral/empleo-y-desempleo.
5. Jiménez WS, Gómez LEN, Díaz RG (2018) Cambio estructural de la vocación agrícola y pecuaria en el municipio de purificación, tolima, Colombia. *Libre Empresa* 15: 137–148. https://doi.org/10.18041/1657-2815/libreempresa.2018v15n2.5361
6. Rios LAM, Villegas JV, Suarez A (2020) Local perceptions about rural abandonment drivers in the colombian coffee region: Insights from the city of manizales. *Land Use Policy* 91: 104361. https://doi.org/10.1016/j.landusepol.2019.104361

7. Gottlieb C, Grobovšek J (2019) Communal land and agricultural productivity. *J Dev Econ* 138: 135–152. https://doi.org/10.1016/j.jdeveco.2018.11.001
8. Zhang YM, Diao XS (2020) The changing role of agriculture with economic structural change—the case of China. *China Econ Rev* 62: 101504. https://doi.org/10.1016/j.chieco.2020.101504
9. Rijnks RH, Crowley F, Doran J (2022) Regional variations in automation job risk and labour market thickness to agricultural employment. *J Rural Stud* 91: 10–23. https://doi.org/10.1016/j.jrurstud.2021.12.012
10. Edeme RK, Nkalu NC, Idenyi JC, et al. (2020) Infrastructural development, sustainable agricultural output and employment in ecowas countries. *Sustain Futures* 2: 100010. https://doi.org/10.1016/j.sftr.2020.100010
11. Diaz RT, Osorio DP, Hernández EM, et al. (2022) Socioeconomic determinants that influence the agricultural practices of small farm families in northern colombia. *J Saudi Soc Agric Sci* 21: 440–451. https://doi.org/10.1016/j.jssas.2021.12.001
12. Sofer M (2001) Pluriactivity in the moshav: Family farming in israel. *J Rural Stud* 17: 363–375. https://doi.org/10.1016/S0743-0167(01)00012-2
13. Castaneda A, Doan D, Newhouse D, et al. (2018) A new profile of the global poor. *World Dev* 101: 250–267. https://doi.org/10.1016/j.worlddev.2017.08.002
14. Xie Y, Jiang QB (2016) Land arrangements for rural-urban migrant workers in china: Findings from jiangsu province. *Land Use Policy* 50: 262–267. https://doi.org/10.1016/j.landusepol.2015.10.010
15. Silva RP (2023) Current state and transformations of rural employment in latin america. an analysis of the case of chile. *Chil J Agric Anim Sc* 39: 121–132. https://doi.org/10.29393/CHJAA39-10EARP10010
16. Perazzi JR, Merli GO (2019) Labor elasticity of growth by sector and department in colombia: The importance of the agricultural employment elasticity. *Agroalimentaria* 25: 19–34.
17. Sen B, Dorosh P, Ahmed M (2021) Moving out of agriculture in bangladesh: The role of farm, non-farm and mixed households. *World Dev* 144: 105479. https://doi.org/10.1016/j.worlddev.2021.105479
18. Akopov AS, Beklaryan LA, Beklaryan AL (2020) Cluster-based optimization of an evacuation process using a parallel bi-objective real-coded genetic algorithm. *Cybern Inf Technol* 20: 45–63. https://doi.org/10.2478/cait-2020-0027
19. Zhang XW, Zhang YY (2022) Optimization of regional economic industrial structure based on edge computing and fuzzy k-means clustering. *Wirel Commun Mob Com* 2022: 8775138. https://doi.org/10.1155/2022/8775138
20. Heil J, Häring V, Marschner B, et al. (2019) Advantages of fuzzy k-means over k-means clustering in the classification of diffuse reflectance soil spectra: A case study with west african soils. *Geoderma* 337: 11–21. https://doi.org/10.1016/j.geoderma.2018.09.004
21. Metwally MS, Shaddad SM, Liu MG, et al. (2019) Soil properties spatial variability and delineation of site-specific management zones based on soil fertility using fuzzy clustering in a hilly field in Jianyang, Sichuan, China. *Sustainability* 11: 7084. https://doi.org/10.3390/su11247084

22. Zeraatpisheh M, Bakhshandeh E, Emadi M, et al. (2020) Integration of pca and fuzzy clustering for delineation of soil management zones and cost-efficiency analysis in a citrus plantation. *Sustainability* 12: 1–17. https://doi.org/10.3390/su12145809
23. Novák V, Perfilieva I, Dvořák A (2016) *Fuzzy cluster analysis*, John Wiley & Sons, 6: 137–148. https://doi.org/10.1002/9781119193210.ch6
24. Ramamoorthy V (2019) *Fuzzy C-mean clustering using data mining*. BookRix.
25. Bejarano LA, Espitia HE, Montenegro CE (2022) Clustering analysis for the pareto optimal front in multi-objective optimization. *Computation* 10. https://doi.org/10.3390/computation10030037
26. Jang JSR, Sun CT, Mizutani E (1997) Neuro-fuzzy and soft computing-a computational approach to learning and machine intelligence [book review]. *IEEE T Automat Contr* 42: 1482–1484. https://doi.org/10.1109/TAC.1997.633847
27. Cohen MD, March JG, Olsen JP (1972) A garbage can model of organizational choice. *Admin Sci Quart* 17: 1–25. https://doi.org/10.2307/2392088
28. Simon HA (1979) Rational decision making in business organizations. *Am Econ Rev* 69: 493–513.

AIMS Press