Qeios

PEER-APPROVED

v1: 18 May 2025

Research Article

Spatial Analysis of Soil Fertility Using Geostatistical Techniques and Artificial Neural Networks

Preprinted: 20 October 2023
Peer-approved: 18 May 2025

Qeios, Vol. 7 (2025)
ISSN: 2632-3834

Angel Rafael Valera Valera[1], **Eladio Ramon Arias Rodríguez**[1]

1. Soil and Water Research and Extension Center (CIESA-UNERG), Universidad Nacional Experimental Romulo Gallegos, Venezuela, Bolivarian Republic of

Information on the spatial variation of soil fertility attributes is an essential input in precision agriculture for soil management decisions. In this study, soil fertility was evaluated through the spatial distribution of property maps and subsequent integration into fertility classes, as a fundamental basis for the implementation of fertilization plans and amendments adjusted to crop requirements. For the evaluation of fertility, a systematic surface sampling was carried out at 70 sites in the "Agronomy" production field of Romulo Gallegos University, The Castrero sector, Roscio municipality, Guárico state, Venezuela. Ten soil attributes were analyzed: pH, electrical conductivity, organic matter, phosphorus, potassium, calcium, magnesium, and the relative amounts of sand, silt, and clay. Soil property maps were produced by geostatistical analysis and ordinary kriging interpolation, and artificial intelligence techniques based on an artificial neural network classification system, with the FKCN (fuzzy Kohonen clustering network) algorithm, were applied to generate soil fertility classes. The reliability of the maps for each variable was obtained by cross-validation with a reliability of more than 90%. The integration of the maps produced a map composed of five categories. The final soil class model presented a reliability equivalent to 86%, indicating a high degree of homogeneity within the soil classes obtained. This approach overcomes the limitations of traditional methods by integrating multiple variables into a coherent model and is capable of generating information that can be used as a basis for the establishment of experimental plots for research purposes and the specific management of nutrients present in the soil resource of the area under consideration.

Corresponding author: Ángel Rafael Valera Valera, eladioariasrod1956@gmail.com

Introduction

Soil fertility is a critical factor in precision agriculture, as it determines the soil's ability to support crop growth and optimize agricultural productivity. This quality arises from the interaction of physical, chemical, and biological soil properties, whose spatial variability can significantly influence agronomic management[1]. Traditionally, fertility assessment has been based on laboratory analyses, which, although valuable, have limitations by not considering soil spatial heterogeneity. This can lead to generalized fertilization recommendations that do not reflect the actual needs of specific areas[2]. In turn, knowledge of the spatial variation of soil fertility in agricultural fields is a fundamental aspect for the definition of the establishment of homogeneous productive plots, for site-specific management purposes (Srinivasan *et al.,*[1]).

One source of information related to fertility is the soil analyses carried out by laboratories, which provide this service in various locations in the country. This source of data constitutes a contribution of analytical results of soil properties related to reaction (pH), salinity (electrical conductivity), granulometry (clay, sand, silt), organic matter, macroelements (Phosphorus, Potassium), secondary elements (Calcium, Magnesium), microelements (Zinc, Copper, Iron, and Manganese), and exchangeable acidity (Aluminum, Hydrogen)[3][4]. Each soil analysis report for fertility purposes is an integration of results with the purpose of developing an organic and inorganic fertilization plan and amendments adjusted to the soil status and crop requirements, complemented with the management of climatic factors or irrigation and agro-cultural activities.

In this context, the spatial analysis of soil fertility facilitates decision-making when applying agronomic practices in productive or experimental areas, allowing the appropriate supply of nutrients to the soil and minimizing the impact on the soil resource for the benefit of biodiversity. However, the manual representation of soil fertility classes requires the elaboration of individual maps for each of the variables and the subsequent superimposition of these maps to obtain homogeneous areas and similar patterns that facilitate management, which implies biases and low precision in the final result[5]. Therefore, the systematic organization of soil data in geographic areas or land units is an opportunity to evaluate the spatial distribution of topsoil, express the spatial variation of soil fertility through thematic maps, and give a higher added value to soil analysis for fertility purposes through digital mapping products of soil properties and classes with a higher degree of homogeneity[6][7][8][9].

Currently, the use of remote sensing (RS) and geographic information systems (GIS) technologies has revolutionized the assessment and monitoring of soil fertility, making it possible to obtain detailed spatial information and facilitating the identification of large-scale trends and patterns[10]. In this way, digital soil mapping emerges as an essential tool to characterize the spatial variability of fertility, allowing the delimitation of homogeneous zones from the fertility point of view, and the subsequent establishment of experimental plots for an appropriate implementation of site-specific management strategies. These tools, combined with geostatistical methods, have proven to be highly effective in mapping and tracking the variability of nutrients and other edaphic properties, contributing to a more sustainable and accurate management of soil resources.

Geostatistical methods play an important role in the prediction of soil properties, where the interpolation method called ordinary kriging stands out[3][4][11]. However, the individual representation of the variables that define soil fertility does not cover the interest and the need to visualize the behavior of soil fertility as a whole. Spatial analysis allows us to evaluate the variation of individual soil properties and the conformation of soil classes, in order to support decision-making on homogeneous areas as a basis for site-specific management and for the promotion of precision agriculture.

Recent research has also highlighted the importance of considering the spatial interaction of soil fertility, since factors such as topography and water dynamics can influence nutrient distribution and availability in different areas of the same field[12]. Spatial analysis, supported by statistical models and artificial intelligence algorithms, is thus consolidated as an essential tool for delineating homogeneous zones and optimizing agronomic management strategies. This information serves as a basis for users to have a complete idea about the soil nutrient status of a sector on a single map, and also contributes to decision-making regarding the most appropriate soil management[2][13].

For the generation of soil classes, there are spatial analysis techniques based on artificial intelligence, such as fuzzy logic and artificial neural networks (ANN)[14]. These techniques are well suited to the study of soil attributes, which vary gradually over space, where the representation of this gradual variation can result in obtaining useful information and reducing errors in the definition of appropriate boundaries of soil units[15]. Techniques such as ordinary kriging and artificial intelligence systems, such as fuzzy neural networks, have proven to be effective in modeling soil properties and classifying soils into homogeneous categories[16][13]. The combination of the potential of fuzzy sets and ANNs has developed a comprehensive unsupervised classification technique called the Fuzzy Kohonen Clustering Network (FKCN)[17][18], which combines a self-organizing map (SOM) algorithm[19] and the Fuzzy C-means (FCM) algorithm[20][21].

There are few research works carried out in the field of soil science that take into account the combination of individual properties to express them as soil fertility categories. In this regard, the application of fuzzy-neural networks has given a great impulse to digital soil mapping, both in the prediction of properties and in obtaining soil classes. In Venezuela, fuzzy neural networks have been applied in the area of landscape classification and soil attribute prediction[22], in geomorphological digital mapping[23], Valera *et al.*[24], Viloria *et al.*[25][26], Valera[27], Sevilla[28] and Viloria *et al*[29], in the prediction of local soil properties and classes[30][31], in the study of soil and banana crop yield relationships[32] and in the delimitation of fertility classes[33].

This work presents a study of spatial analysis of fertility classes through the prediction of soil chemical and physical properties obtained in laboratory analysis by means of geostatistical techniques, and their subsequent grouping by means of a fuzzy artificial neural network algorithm. To evaluate the spatial behavior of soil fertility classes, the "Agronomy" Production Field of the National University of the Central Plains "Romulo Gallegos", located in The Castrero sector, Saint John of the Morros parish, John Germain Roscio municipality, Guárico state (Venezuela), was considered. The main purpose of the research is the spatial prediction of soil fertility classes through

doi.org/10.32388/0JAFR0.3

the integration of artificial neural networks and geostatistical techniques, as a basis for the generation of basic information required for the development of trials and experimental tests, which allow a spatial vision of the fertility status, and a better interpretation of the results of the different treatments to be established, agronomic trials, nutrient management, as well as future field research and evaluations for experimental purposes to be developed in the studied sector.

Materials and Methods

Study Area

The study area where the digital soil mapping test was carried out is located in the "Agronomy" production field of the National Experimental University of the Central Plains "Romulo Gallegos", located in the Castrero sector, Saint John of the Morros parish, John Germain Roscio municipality, Guárico state (Figure 1). The study unit is framed within a colluvial-alluvial valley, with a slope of 3 to 5%, dominated by hills and mountains. The soils in this area, belonging to the Saint Elizabeth Formation (from a geological point of view), are of moderate pedogenetic development (Large soil group *Haplustepts*), and are of medium fertility. The predominant vegetation corresponds to yaraguá grass (*Hyparrhenia rufa*), mastranto (*Mentha suaveolens*) and, to a lesser extent, some scattered tree species of chaparro (*Curatella americana*). The area has been used in the past for sporadic planting of crops such as corn and beans for commercial purposes; however, it is necessary to use the land for research purposes for the development of trials and experiments, after determining management zones for specific sites.

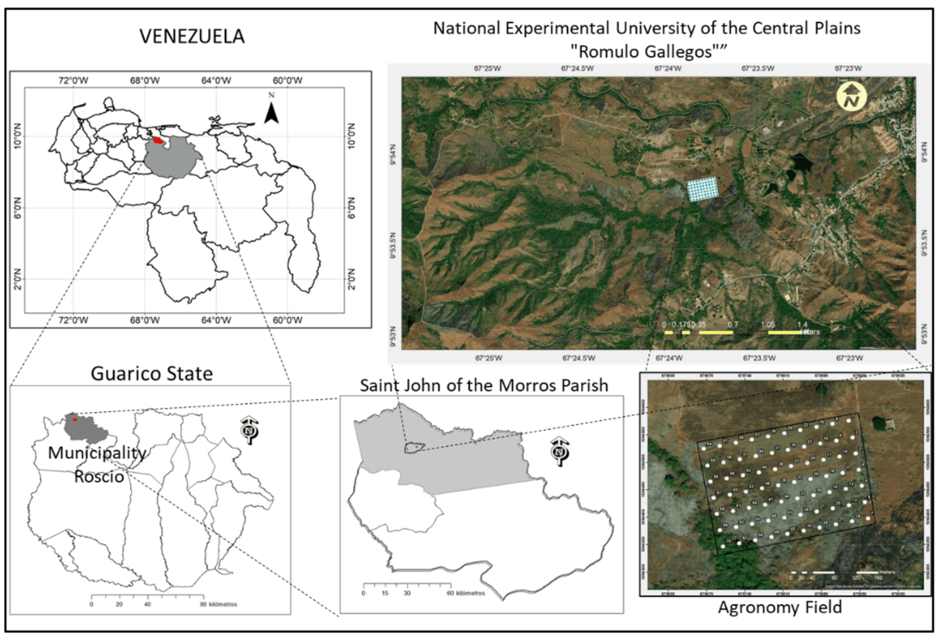

Figure 1. Relative location of the production field "Agronomy" in the basin of the river The Castrero, Guarico state, Venezuela.

Soil sampling

For the evaluation of the soils, a systematic sampling was carried out in the superficial horizon at a depth of 20 cm, in grids spaced at 30 m, for a total of 70 soil samples in an area of 6.15 ha (Figure 2). Each sampling point was georeferenced with the support of a global positioning system (GPS). The surface samples were diagnosed for fertility purposes, using the methodologies of the Soil Analysis Laboratory of the Soil and Water Research Center of the Romulo Gallegos University (CIESA-UNERG). Ten soil variables were analyzed: pH in water (1:2.5), electrical conductivity in water 1:5 (EC, dSm^{-1}), organic matter (OM, %), available phosphorus (P, $mgkg^{-1}$), available potassium (K, $cmol_{(+)}kg^{-1}$), calcium (Ca, $cmol_{(+)}kg^{-1}$) and available magnesium (Mg, $cmol_{(+)}kg^{-1}$), and the relative amounts of sand, silt, and clay (%).

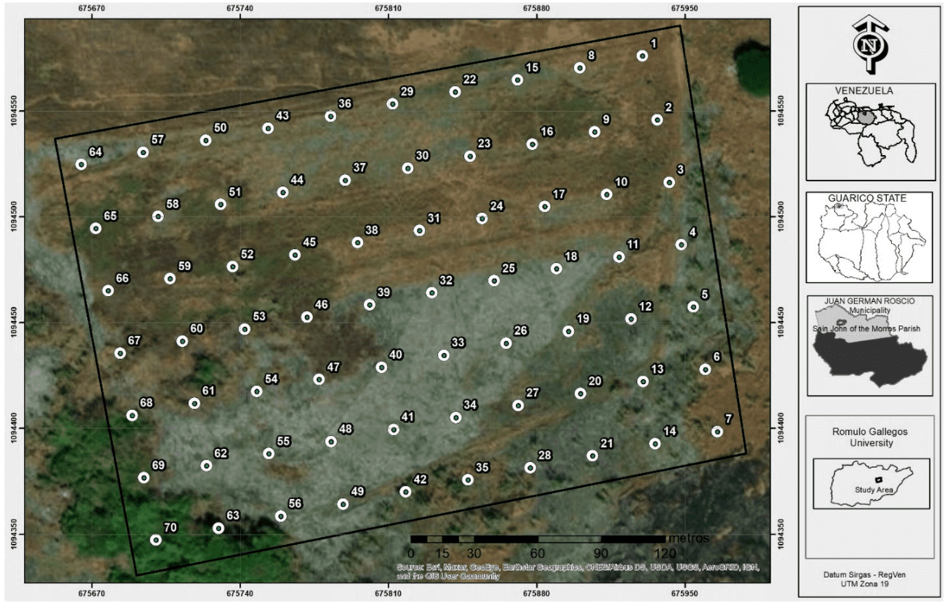

Figure 2. Distribution of soil sampling sites in the production field "Agronomy".

Statistical analysis

The data of the edaphic variables were subjected to an exploratory analysis (EDA) with the support of the statistical package SPSS® (IBM® Statistics, version 20), in order to determine the descriptive statistics, such as mean, median, variance, coefficient of variation, maximum and minimum values, and the asymmetry and kurtosis indices. Tukey's[34] methodology of external and internal fences was used in order to detect the presence of outliers. Additionally, the normality test of Kolmogorov-Smirnov was performed to evaluate the distribution of the data.

Soil Properties Interpolation

For the interpolation of soil properties, the ordinary geostatistical *kriging* method was used, which uses a semivariogram model to obtain the weights assigned to each reference point used in the estimation of the value of the regionalized variables that present spatial dependence. The semivariogram is defined by the semivariance function [γ ***(h)***], which is estimated with the following expression[35][36]:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{N(h)} \left[z\left(x_i\right) - z\left(x_{i+h}\right)\right]^2 \tag{1}$$

where N is the number of pairs of points separated by a given distance h; $z(x_i)$ is the value of the variable at a location x; $z(x_{i+h})$ is the value that the variable takes at another location located at a distance h from x[37]. The semivariogram contains the information concerning the regionalized variable, whose parameters are: the *nugget* variance (C_0), the structural variance (C_1), the threshold or plateau ($C_0 + C_1$) and the range (A_1), which indicates the distance within which there is spatial dependence[38][39]. For the estimation of the empirical semivariogram of soil properties, the necessary transformations were performed and possible trends in the data were removed. Then the adjustment to mathematical models was carried out with the geostatistical analysis extension of the ArcGIS software® (ArcMap v. 10.8). The fitted parameters were used to obtain optimal estimates of the soil variables at the unsampled sites, through interpolation using the ordinary *kriging* method[11].

Models of soil variables were generated from the total data, and the accuracy of the maps was obtained by cross-validation. Six indices were used in the evaluation: mean error (ME), standardized mean error (SME), root mean square error (RMSE), standardized root-mean square error (SRMSE), average standard error (ASE), and confidence level (%CL). The ME evaluates the systematic error and indicates the presence of under- or overestimation of the model, and the SME shows the deviation of the model obtained. The RMSE assesses the accuracy of the prediction and measures the amount of error between the measured and inferred data sets, i.e., it compares a predicted value and an observed or known value; whereas the SRMSE is more accurate the closer it is to the ideal value of unity (1). The ASE indicates the variability of predictions, whose estimates will be more appropriate if their values are closer to the RMSE.

Digital soil fertility class model

The neuro-fuzzy FKCN algorithm, implemented in a Java environment (Windows) by Viloria[25][26], was used to obtain the representative models of the soil fertility classes. The architecture of the FKCN neural network used in the analysis consists of three layers (Figure 3). The input layer contains the normalized values of ten (10) soil variables from the prediction models of these attributes. The distance layer includes the neurons equivalent to the preset number of digital soil classes, and the third layer computes the membership function of each cell to each of the soil classes, based on the distances computed in the previous layer and the preset values of the fuzzy coefficient (φ). In the distance layer, the separation d_{ij} existing between an input pattern X_j and the node weight ω_i is computed, with $i = 1, 2, \ldots, c$, where c represents the number of classes of the model to be estimated. Subsequently, the membership layer plots the distances d_{ij} into membership values U_{ij}, where U_{ij} represents the degree of membership of an input pattern X_j to a class c. In the learning process, the feedback from the membership functions layer to the distance layer occurs in order to adjust the centers of each class.

The soil variables were grouped in a data matrix for the application of the FKCN algorithm, which allowed the evaluation of pixel clustering with different numbers of classes (2 to 8) and different fuzzy coefficients (φ= 1.1 to 1.6). The fertility classes were assigned pedological significance through the interpretation of their spatial distribution, the descriptions of the class centers, and the matrices of similarity values (degree of belonging to each class) obtained by the FKCN algorithm, together with the information from the analysis of the soils in the area.

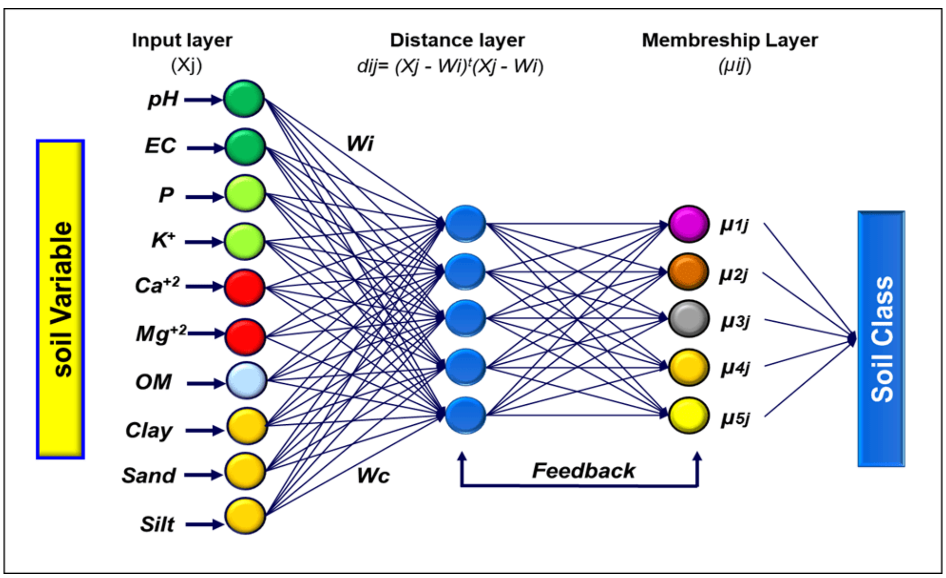

Figure 3. Structure of the fuzzy neural network used in soil class prediction.

Number of soil fertility classes

An inductive approach was used to obtain the best fuzzy class model, based on the procedure of Odeh *et al.*[40], which relates the *Fuzziness Performance Index* (FPI) to the number of classes. These parameters are obtained using the *Fuzzy Kohonen Clustering Networks* (FKCN) algorithm[17] of the *FKCN* program[25][26]. The selection of the optimal number of classes in FKCN was performed by repeated clustering for a range of numbers of classes. The FPI estimates the degree of fuzziness generated by each specific number of classes[40]. Mathematically, it is defined as:

$$FPI = 1 - [(cF - 1)/(c - 1)] \quad (2)$$

where ***c*** is the number of classes and ***F*** is the partition coefficient calculated as:

$$F = (1/n) \sum_{i=1}^{n} \sum_{k=1}^{c} (\mu_{ik})^2 \quad (3)$$

F is conceptually comparable to the ratio of the set of within-class variances to the between-class variance and is close to unity (1) for the most significant clustering. In the present study, the clustering of soil property maps in *raster* format was performed by previously setting the following parameters: a) number of classes (c= 3 to 10), b) fuzzy exponent φ = 1.10 to 1.60 with increments of 0.10; c) a maximum of 300 iterations, and d) stopping criterion (ε= 0.0001). The calculations used the *Mahalanobis* metric distance, which takes into account the correlation between some soil variables in the area evaluated.

Assessment of the predictive ability of soil fertility classes

To assess the predictive capacity of the classes obtained by fuzzy clustering, the final model was validated by means of a cross-validation process, using the discriminant functions of each class as multivariate statistics derived from the canonical discriminant analysis. In the cross-validation process, each case is classified using the discriminant functions derived from the rest of the cases.

Results and Discussion

Statistical analysis

Descriptive statistics indicated that the average values of the soils correspond to clay loam and clayey textural groups, with slightly to strongly acid reactions, low to medium phosphorus, and moderate to high potassium contents, high availability of calcium and magnesium, low to medium organic matter contents, and no salinity problems (Table 1).

Variable[1]	Min.	Max.	Ave.	Medium	Kurtosis	Asymmetry	SD	Var	CV (%)
pH (1:2.5)	4.97	6.40	5.70	5.66	−0.23	0.41	0.32	0.103	5.6
EC (dS m^{-1})	0.010	0.100	0.036	0.026	0.47	1.12	0.02	0.001	63.6
P (mg kg^{-1})	4.44	44.48	17.1	16.24	−0.20	0.57	9.83	96.55	57.6
K (cmol (+) kg^{-1})	0.31	1.44	0.71	0.67	0.68	0.91	0.26	0.069	37.3
Ca (cmol (+) kg-1)	0.90	2.40	1.59	1.55	0.63	0.53	0.29	0.082	18.0
Mg (cmol (+) kg^{-1})	0.17	1.41	0.88	0.90	0.43	−0.15	0.24	0.060	27.7
OM (%)	0.78	4.17	2.67	2.57	0.85	0.06	0.63	0.400	23.6
Clay (%)	14.00	70.00	41.1	40.48	0.14	0.13	11.23	126.0	27.3
Sand (%)	5.10	74.96	31.5	32.98	1.19	0.33	12.66	160.2	40.2
Silt (%)	5.04	52.40	27.3	29.00	-0.74	-0.17	11.45	131.1	41.9

Table 1. Descriptive statistics of the soil fertility variables of the experimental field.

[1] *Number of data: 70, SD: Standard deviation, CV: Coefficient of variation, EC: Electrical conductivity, P: Available phosphorus, K: Available potassium, Ca: Available calcium, Mg: Available magnesium, OM: Organic matter.*

Most of the variables show some similarity between the mean and the median, with the exception of the EC and K variables. At the same time, the greatest dispersion of the data is presented by the same variables together with the granulometry values, due to the expression of the standard deviation and variance; however, the coefficients of variation of the variables as a whole do not present problems in terms of the existence of extreme values in the data.

According to the coefficient of skewness or asymmetry, the variables pH, %sand, %clay, Ca, Mg, and %OM comply with the normal probability distribution function, and geostatistical methods can be applied to the data. However, for P, K, and EC, it was necessary to evaluate the data by transformations (normalization) for the subsequent application of some geostatistical method to the data. Regarding kurtosis, only the data for the K variables are concentrated with respect to the mean (small standard deviation), giving an elongated plot; while the data for pH, %silt, and P are scattered, presenting flattened or flattened plots.

The application of the test for external and internal fences indicated that the variables considered do not present outliers. Finally, with regard to the normality test, it was verified that only the variables K and MO come from normal populations, as the values of the statistical test are highly significant (p> 0.05). For the rest of the data, it was necessary to transform them.

Soil Properties Interpolation

The estimation of the empirical semivariogram of the soil variables was fitted to Gaussian, spherical, exponential, stable, and cylindrical mathematical models respectively (Figure 4), considering the isotropic behavior of the variables. The geostatistical parameters derived from fitting the semi-variograms to different theoretical models are expressed in Table 2, and the models for each variable are presented in Figure 5.

Variables	Model	C_0	C_1	A_1	C +C_{01}	RN (%)
pH (1:2.5)	Gaussian	0.00	0.11	56	0.11	0.0
EC (dS m^{-1})	Gaussian	0.00	0.00	59	0.00	0.0
P (mg kg^{-1})	Spherical	0.10	82.4	93	82.6	0.1
K (cmol (+) kg^{-1})	Circular	0.08	0.05	59	0.13	60.4
Ca (cmol (+) kg^{-1})	J-Bessel	0.01	0.04	59	0.05	30.7
Mg (cmol (+) kg-1)	Stable	0.00	0.06	59	0.06	0.0
OM (%)	Gaussian	0.03	0.41	59	0.40	6.8
Clay (%)	Gaussian	0.00	125.6	59	125.6	0.0
Sand (%)	Spherical	78.3	22.1	59	100.3	78.0
Silt (%)	J-Bessel	0.000	137.0	59	137.0	0.0

Table 2. Geostatistical parameters of the composite semivariogram of soil properties.

*C_0: Nugget variance, C_1: Structural variance, C_0 + C_1: Threshold, A_1: Range, RN: Relative Nugget (C_0/C_0 +C_1)*100), J-Bessel: Cylindrical symmetry function.*

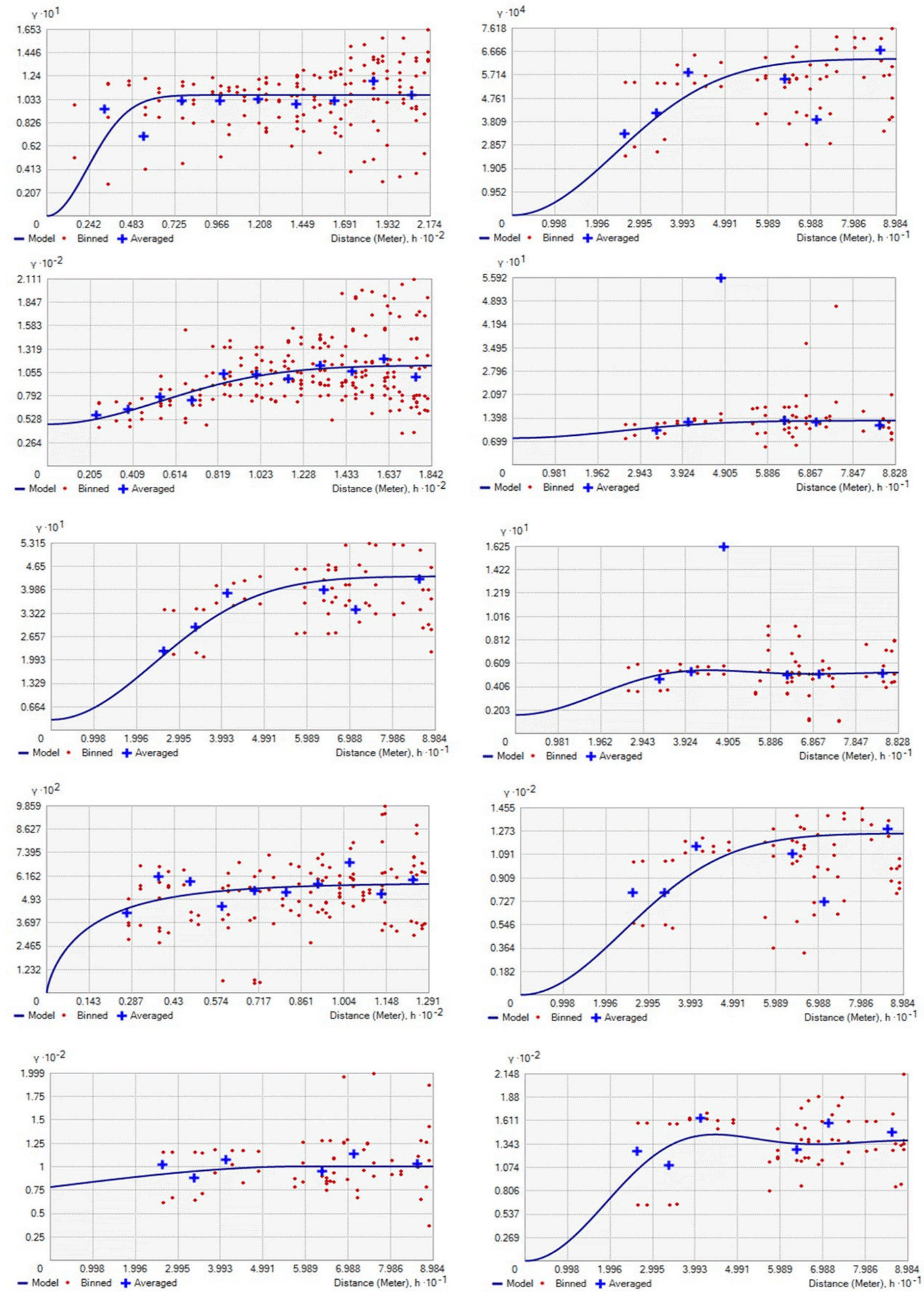

Figure 4. Semivariogram of soil variables in the production field "Agronomy".

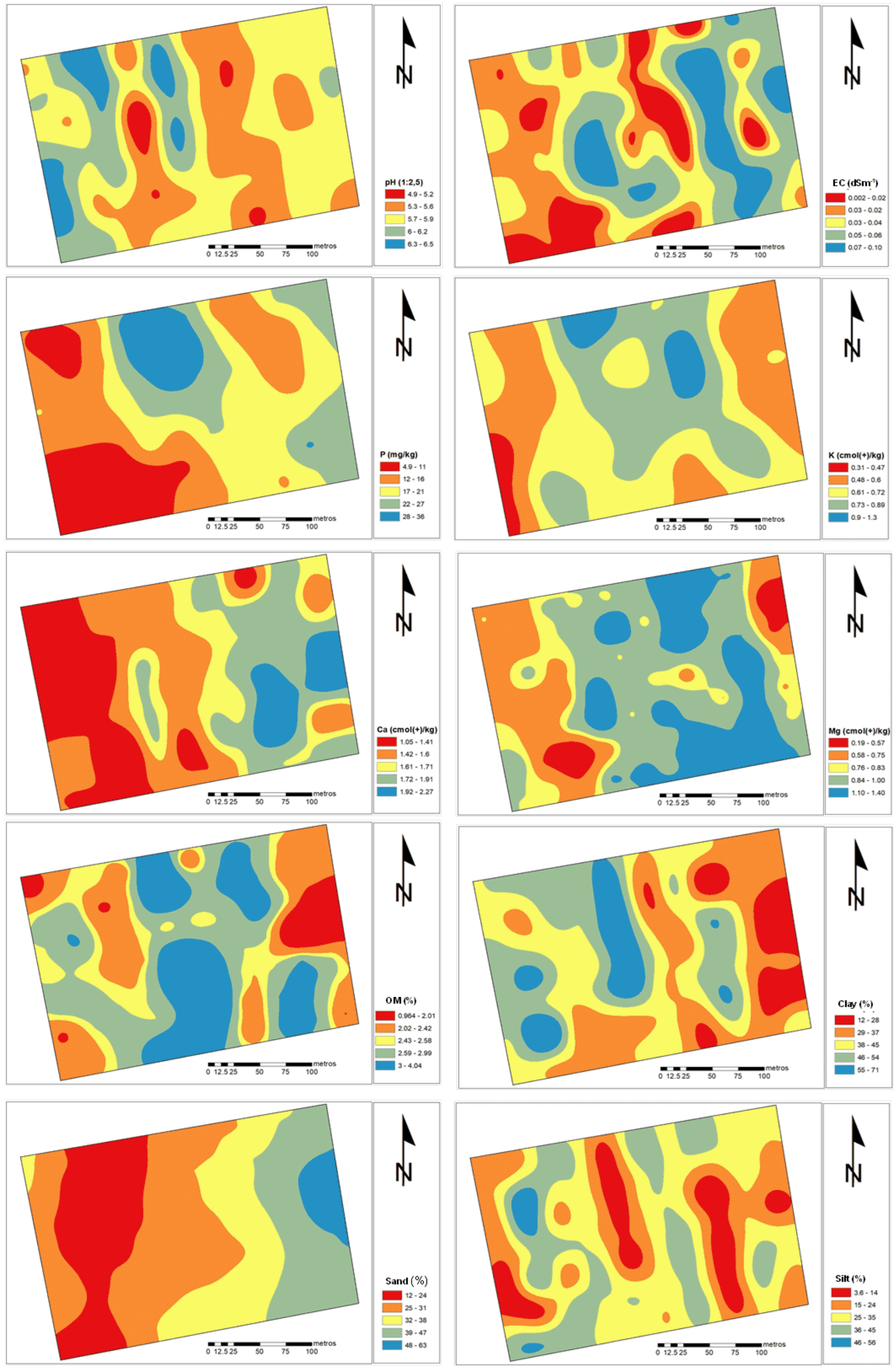

Figure 5. Model maps of soil variables in the production field "Agronomy".

The structural analysis of the semivariogram models indicated that the models of the variables mostly show strong spatial dependence (<25% random effect or relative *nugget*), although the variables K and Ca show moderate spatial dependence (relative variance between 25 and 75%), and % sand shows weak spatial dependence, with a relative *nugget* >75%. In general, all semivariograms show structure, with an increase in the total variance until reaching a maximum average distance of 59 m. That is, a spatial dependence range of 59 m stands out for all models, with the exception of the semivariogram of the available P variable, which presents a range 1.5 times the average.

 doi.org/10.32388/0JAFR0.3

The spatial variability observed in the study area, with a dependence range of 59 m, suggests that agronomic management should be adapted to scales smaller than 1 ha to ensure plot homogeneity. This is consistent with the recommendations of Srinivasan *et al.* [1], who emphasize the importance of considering local variability, which is an aspect of great importance in the establishment of experimental plots for research purposes and in the management of variable rates as a basis for precision agriculture.

Reliability Assessment of Prediction Models

The results of the validations of the soil variables are shown in Table 3, where the low values of the prediction errors, which are very close to zero for the ME, SME, and ASE indices, can be observed.

Variable	Regression Function	Índex					
		ME	SME	RMSE	SRMSE	ASE	CL (%)
pH (1:2.5)	0.4631 * x + 3.03326	0.00	0.00	0.25	1.08	0.23	99.6
EC (dS m^{-1})	0.4466 * x + 0.01720	0.00	0.01	0.02	1.30	0.02	100.0
P (mg kg^{-1})	0.3914 * x + 10.3986	-0.01	0.00	8.63	1.48	5.85	91.4
K^+ (cmol kg^{-1})	0.1555 * x + 0.58614	0.00	-0.01	0.25	0.96	0.26	99.7
Ca^{+2} (cmol kg^{-1})	0.4550 * x + 0.84930	0.00	0.00	0.23	0.98	0.23	99.7
Mg^{+2} (cmol kg^{-1})	0.2305 * x + 0.67800	0.00	0.01	0.21	0.96	0.22	99.7
OM (%)	0.3167 * x + 1.81928	0.01	0.02	0.55	0.98	0.56	99.2
Clay (%)	0.5668 * x + 18.1189	-0.02	0.00	9.08	1.31	7.19	86.1
Sand (%)	0.3824 * x + 19.4403	-0.41	-0.04	10.4	1.00	10.3	91.4
Silt (%)	0.1880 * x + 21.5820	0.20	0.02	10.8	0.93	11.6	85.1

Table 3. Prediction error of soil variables by cross-validations.

ME: Mean error, SME: standardized mean error, RMSE: root-mean square error, SRMSE: standardized root-mean square error, ASE: Average standard error, CL: Confidence level. EC: electrical conductivity, P: available phosphorus, K: available potassium, Ca: available calcium, Mg: available magnesium, OM: organic matter.

It is observed that the models that best fit the data used meet the requirements of small RMSE, small ASE close to RMSE, SRMSE close to 1, and a high percentage of reliability. According to the reliability of the models, most of them present values higher than 90%, except for the variables % Clay and % Silt, for which it is necessary to improve the density of the measurements. The greatest underestimation was presented by the physical variables, and the greatest uncertainty is given by the available phosphorus variable (far from the unit), which presented a greater variance and somewhat high variation coefficients. However, the particle size variables show ASE values very close to RMSE. For all the cases evaluated, the RMSE values are lower than the *standard deviation*, and are therefore adequate for the evaluation of the prediction models[41].

Generation of the digital soil fertility class model

Number of soil fertility classes

The representation of the variation of the fuzzy performance index (FPI) as a function of the number of classes for different coefficients is shown in Figure 6. The diagram shows that the most suitable number of soil classes was obtained with 5 classes, combined with a_φ of 1.2. The FPI value of 0.36 points to the intersection point at which there is a minimization of the degree of fuzziness, which determined the optimal number of classes, characterized by being less fuzzy and less internally disorganized for the set of variables related to soil fertility.

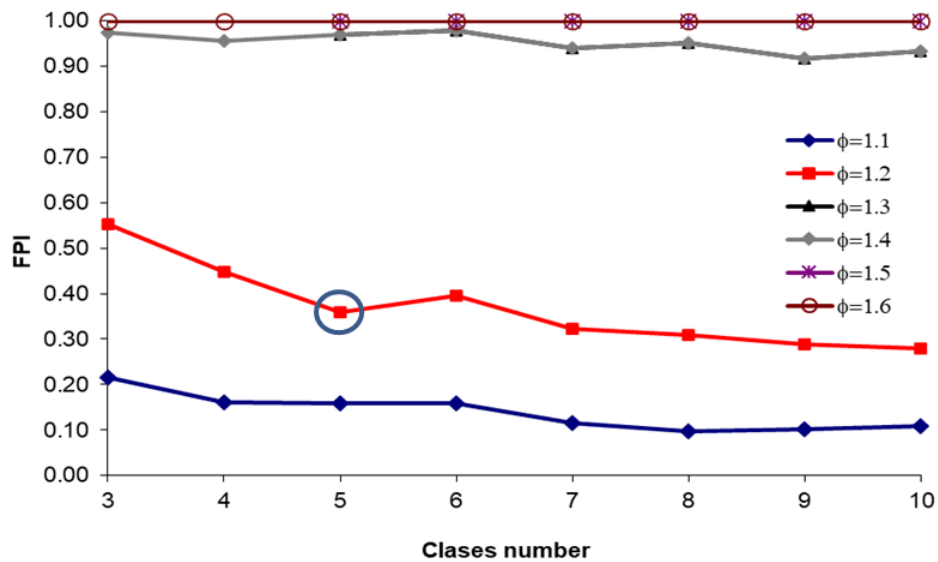

Figure 6. Variation of the fuzzy performance index (FPI) as a function of the number of soil classes.

The results of the values of the center of each fertility class (centroids) are shown in Table 4. This allowed the following significant aspects to be extracted: Class **1** includes soils of clayey textural classes and slightly acidic pH, and the lowest values of available phosphorus. Class **2 includes** soils of clay loam texture, with moderately acid reactions and average values for most of the available chemical elements that characterize it. Class **3** also includes soils with a clay loam texture and a moderately acid reaction, but with clay contents close to 40%, and with the highest available calcium and magnesium contents. Class **4** groups soils with the highest clay contents (>50%), and the highest levels of available phosphorus, assimilable potassium, and organic matter contents. Class **5 involves** clay loam soils with the lowest levels of assimilable potassium, available magnesium, and organic matter contents, but with the highest proportions of coarse-grained materials (sand).

Soil Variable	**Soil Fertility Classes**				
	1	**2**	**3**	**4**	**5**
pH water (1:2.5)	5.99	5.52	5.52	5.88	5.70
EC water (dS m^{-1})	0.02	0.03	0.06	0.03	0.05
P (mg kg^{-1})	11.0	17.0	18.0	25.0	22.0
K^+ (cmol kg^{-1})	0.59	0.71	0.75	0.93	0.57
Ca^{+2} (cmol kg^{-1})	1.35	1.58	1.80	1.54	1.77
Mg^{+2} (cmol kg^{-1})	0.73	0.99	1.07	0.90	0.68
OM (%)	2.41	2.87	2.73	3.04	2.09
Clay (%)	46.7	36.4	39.5	52.6	26.6
Sand (%)	26.0	31.3	37.5	24.5	47.3
Silt (%)	26.9	33.8	23.0	22.6	23.8

Table 4. Soil fertility classes center obtained with the FKCN algorithm.

EC: Electrical conductivity, P: Available phosphorus, K: Available potassium, Ca: Available calcium, Mg: Available magnesium, OM: Organic matter.

doi.org/10.32388/0JAFR0.3

The application of the FKCN algorithm also generated the membership degree values of each cell (pixel) to each of the soil fertility classes. The classification produced vectors of membership values for each model cell corresponding to each fertility class. These values were spatially represented, producing individual maps of class memberships, which reflect the spatial variation of membership degrees between 0 (dark colors) and 1 (light colors), through maps in *raster* format expressed in Figure 7.

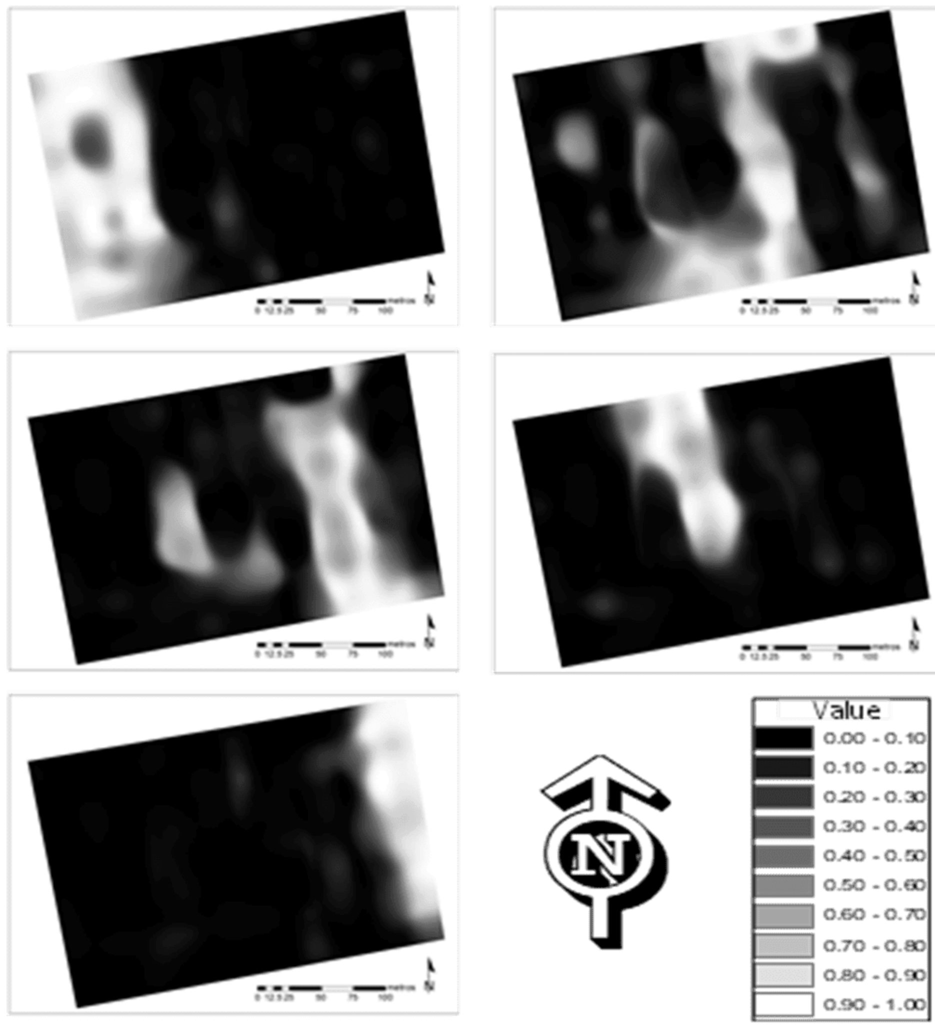

Figure 7. Maps of membership function values for each of the soil fertility classes.

The combination of the spatial distribution models of the membership values produced the integrated map of five soil fertility classes (Figure 8). To produce this map, the FKCN algorithm converted the neuro-fuzzy classes into discrete units, whereby each model cell was assigned to the class with the highest membership value. The final model corroborated the distribution of soil fertility classes, where spatial variation patterns allowed discriminating the dominance of textural classes with variations in soil reaction and availability of primary and secondary elements in the East-West sectors.

The final model also allowed visualization of the expression of the boundaries defined by the dominant fertility classes in the surface layer of the soils. These boundaries facilitate decision-making for soil management and for the development of productive plots. With regard to the surface area of the soil units: class 1 occupies 26.8% of the evaluated sector, class 2 occupies an area of 25.2%, class 3 represents 23.4% of the studied area, class 4 corresponds to 12.7% of the study area, and class 5 corresponds to 11.9% of the production field under consideration.

 doi.org/10.32388/0JAFR0.3

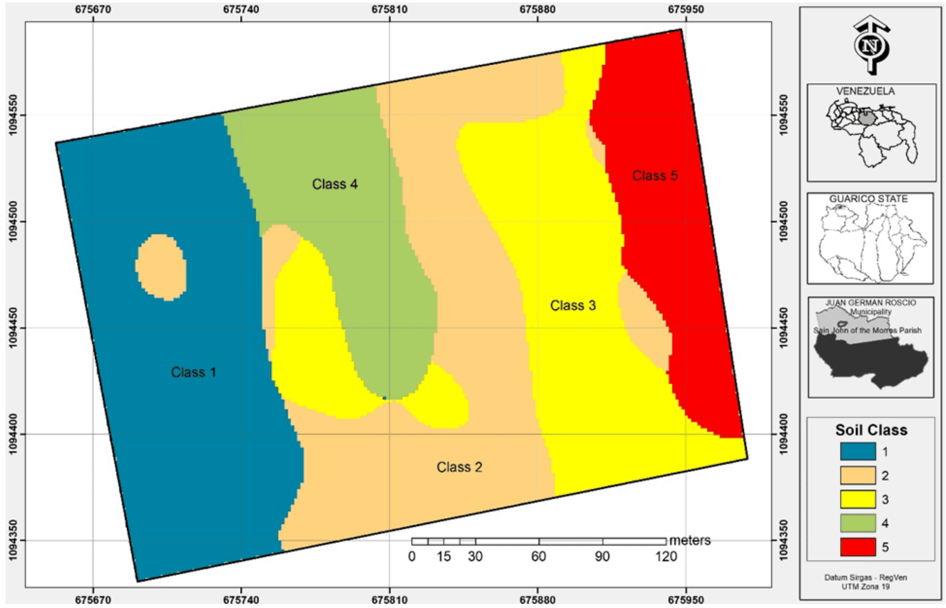

Figure 8. Soil fertility class distribution model of the production field "Agronomy".

Assessment of the predictive capacity of the digital soil fertility classes model

The results of the assessment of the predictive ability of the soil classes with multivariate statistics are reported in Table 5. The calculation of the accuracy of the model yielded values equivalent to 86%, with an uncertainty of less than 15%. In other words, the validation process of the soil fertility class model indicated that 86% of the cases were correctly classified by cross-validation, based on the ratio of correct reference points (60) to the total number of true points (70).

Class	Predicted membership group [a]				
	1	2	3	4	5
1	**100.0**	0.0	0.0	0.0	0.0
2	5.6	**83.3**	11.1	0.0	0.0
3	0.0	6.7	**93.3**	0.0	0.0
4	12.5	12.5	0.0	**75.0**	0.0
5	0.0	44.4	0.0	0.0	**55.6**

Table 5. Classification results (%) based on the sizes of the neuro-fuzzy soil fertility class.

[a] *Correctly classified 85.7% of the grouped cases validated by cross-validation.*

The highest degree of uncertainty is given by classes 5 and 2, where some sites in class 5 were classified as part of class 2, where confusions occur due to neighboring inclusions, as visualized in the final model (Figure 8). The presence of inclusions between classes highlights the need to increase the sampling density in future studies to reduce uncertainty. Therefore, it is suggested to explore the impact of future agricultural research on nutrient dynamics and to increase the sampling density to improve the accuracy of the models, especially for physical properties such as soil texture.

The reliability of the FKCN model is within the range reported in the literature for artificial intelligence techniques applied to soils. Studies such as Zhu et al.[16] and Valera and Orta[33] obtained accuracies of 76% and 80.1%, respectively, using fuzzy methods, while McKay et al.[42] reached only

doi.org/10.32388/0JAFR0.3

73.7% with traditional ANNs. This suggests that the integration of fuzzy logic and neural networks (FKCN) improves classification over purely deterministic approaches. On the other hand, supervised techniques such as Random Forest[2] and support vector machine (SVM)[1] have achieved similar accuracies (83-89%), but require labeled data, which limits their applicability in areas with sparse prior information. The advantage of FKCN lies in its unsupervised capability and the interpretability of fuzzy membership maps, key for decision-making in precision agriculture. The results of the FKCN approach demonstrated that it is an alternative for the generation of soil fertility classes.

Conclusions

The maps of the variables analyzed showed that there are gradual changes in the soil with respect to all attributes, which showed spatial dependence, which can affect the reliability of evaluations for research or production purposes.

The evaluated area is not internally homogeneous, possibly due to the influence of soil management and agronomic practices carried out in the area. This variability must be taken into account to avoid a differential effect on crops.

The establishment of experimental plots should not exceed the range of spatial dependence of fertility attributes, whose mode is 59 m, to include the variability of the evaluated soils. Therefore, the surface area for the establishment of plots that guarantee the homogeneity of the internal structure of the soils should not exceed 1.0 ha in order to allow its representativeness. This study provides a scientific basis for the implementation of site-specific management in the study area, recommending the establishment of research plots no larger than 1 ha to maintain the internal homogeneity of the soil.

The combination of ordinary kriging and the FKCN algorithm proved to be a robust tool for soil fertility classification, with a reliability of 86%. This approach overcomes the limitations of traditional methods by integrating multiple variables into a coherent model.

The information provided by the spatial analysis of individual soil properties and the map of neuro-fuzzy fertility classes is complementary and can be used as a basis for the establishment of experimental plots and the specific management of nutrients present in the soil resource of the area.

This work contributes to the advancement of digital soil mapping in Venezuela and highlights the potential of artificial intelligence techniques for sustainable soil management. The results can be extrapolated to areas of interest with similar edaphic conditions, offering a replicable methodological framework.

Statements and Declarations

Data Availability

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author Contributions

Conceptualization, A.R.V.V. and E.R.A.R.; methodology, A.R.V.V.; software, A.R.V.V.; validation, A.R.V.V.; formal analysis, A.R.V.V.; investigation, A.R.V.V. and E.R.A.R.; resources, A.R.V.V. and E.R.A.R.; data curation, A.R.V.V.; writing—original draft preparation, A.R.V.V.; writing—review and editing, A.R.V.V. and E.R.A.R.; visualization, A.R.V.V.; supervision, A.R.V.V.; project administration, A.R.V.V.; funding acquisition, A.R.V.V.

Acknowledgements

This research was supported by the Soil and Water Research and Extension Center of the National Experimental University of the Central Plains "Romulo Gallegos" (CIESA-UNERG).

References

1. a, b, c, d *Srinivasan R, Shashikumar BN, Singh SK (2022). "Mapping of Soil Nutrient Variability and Delineating Site-Specific Management Zones Using Fuzzy Clustering Analysis in Eastern Coastal Region, India." Journal of the Indian Society of Remote Sensing.* ***50****(3):533–547. doi:10.1007/s12524-021-01473-9.*
2. a, b, c *Padua S, Chattopadhyay T, Bandyopadhyay S, Ramchandran S, Jena RK, Ray P, Deb Roy P, Baruah U, Sah KD, Singh SK, Ray SK (2018). "A simplified soil nutrient information system: study from the No*

*rth East Region of India." Current Science. **114**(6):1241–1249.*

3. ^a,^ ^b^*Dankoub Z, Ayoubi S, Khademi H, Lu SG (2012). "Spatial Distribution of Magnetic Properties and Selected Heavy Metals in Calcareous Soils as Affected by Land Use in the Isfahan Region, Central Iran." Pedosphere. **22**(1):33–47. doi:10.1016/S1002-0160(11)60189-6.*
4. a, b*Ajami M, Ayoubi S, Khademi H (2020). "Spatial variability of rainfed wheat production under the influence of topography and soil properties in loess-derived soils, northern Iran." International Journal of Plant Production. **14**(1):1–12.*
5. ^*Kiran Y, Prasuna, Chaitanya L, B, Kethan B (2024). "Artificial Intelligence and Machine Learning Technique for Soil Classification System." International Journal of Research Publication and Reviews. **5**(1).*
6. ^*Ayoubi S, Zamani SM, Khormali F (2007). "Spatial variability of some soil properties for site specific farming in northern Iran." International Journal of Plant Production. **1**(2):225–236.*
7. ^*Zeraatpisheh M, Ayoubi S, Jafari A, Finke P (2017). "Comparing the efficiency of digital and conventional soil mapping to predict soil types in a semi-arid region in Iran." Geomorphology. **285**:186–204. doi:10.1016/j.geomorph.2017.02.015.*
8. ^*Azadmard B, Mosaddeghi MR, Ayoubi S, Chavoshi E, Raoof M (2018). "Spatial variability of near-saturated soil hydraulic properties in Moghan plain, North-Western Iran." Arabian Journal of Geosciences. **11**(16):452. doi:10.1007/s12517-018-3788-8.*
9. ^*Zeraatpisheh M, Ayoubi S, Sulieman M, Rodrigo-Comino J (2019). "Determining the spatial distribution of soil properties using the environmental covariates and multivariate statistical analysis: a case study in semi-arid regions of Iran." Journal of Arid Land. **11**(4):551–566. doi:10.1007/s40333-019-0059-9.*
10. ^*Singh S, Rai V, Upadhyay S, Singh S (2023). "Geo-spatial Tools for Assessing Soil Fertility: A Review." International Journal of Plant & Soil Science. **35**(18):1386–1394.*
11. a, b*Webster R, Oliver MA (1990). Statistical Methods in Soil and Land Resource Survey. Oxford, RU: Oxford University Press. 316 p.*
12. ^*Sekiya N, Watanabe K, Sakai M (2024). "Uncovering the Hidden Reality of Soil Fertility: Spatial Distribution of Soil Fertility Shaped by Topography Holds the Key to Sustainable Rice Production." Mie University.*
13. a, b*Shashikumar BN, Kumar S, George KJ, Singh AK (2022). "Soil variability mapping and delineation of site-specific management zones using fuzzy clustering analysis in a Mid-Himalayan Watershed, India." Environment, Development and Sustainability. **25**(8):8539–8559. doi:10.1007/s10668-022-02411-6.*
14. ^*Zolfaghari Z, Mosaddeghi MR, Ayoubi S (2015). "ANN-based pedotransfer and soil spatial prediction functions for predicting Atterberg consistency limits and indices from easily available properties at the watershed scale in western Iran." Soil Use and Management. **31**(1):142–154. doi:10.1111/sum.12167.*
15. ^*Burrough PA, Van Gaans PFM, MacMillan RA (2000). "High-resolution landform classification using fuzzy k-means." Fuzzy Sets and Systems. **113**:37–52.*
16. a, b*Zhu AX, Yang L, Li B, Qin C, English E, Burt JE, Zhou C (2008). "Purposive Sampling for Digital Soil Mapping for Areas with Limited Data." In: Hartemink AE, Mendonça-Santos ML, McBratney AB, editors. Digital Soil Mapping with Limited Data. New York: Springer-Verlag. pp. 233–245.*
17. a, b*Lin C, Lee C (1996). Neural fuzzy systems. New Jersey: Prentice Hall, Inc. 797 p.*
18. ^*Bezdek JC, Tsao EC, Pal NR (1992). "Fuzzy Kohonen Clustering Networks." Proc. IEEE Int. Conf. on Fuzzy Systems 1992 (San Diego). pp. 1035–1043.*
19. ^*Kohonen T (1982). "Analysis of a simple self-organizing process." Biological Cybernetics. **44**:135–140.*
20. ^*Bezdek JC (1981). Pattern Recognition with Fuzzy Objective Function Algorithms. New York: Plenum Press. 267 p. https://link.springer.com/book/10.1007/978-1-4757-0450-1.*
21. ^*Bezdek JC, Ehrlich R, Full W (1984). "FCM: the fuzzy c-means clustering algorithm." Computers and Geosciences. **10**:191–203.*
22. ^*Viloria A (2007). "Estimation of landscape classification models and prediction of soil attributes from satellite images and digital elevation models." [Special Degree Project]. Caracas, Venezuela: Universidad Central de Venezuela. 88 p.*
23. ^*Valera A, Viloria JA (2009). "Application of artificial intelligence techniques in the modeling of landscape units in the Güey River basin, Maracay - Aragua state." Memorias XVIII Congreso Venezolano de la Ciencia del Suelo; Santa Bárbara, Zulia, Venezuela. 7 p.*
24. ^*Valera A, Viloria JA, Viloria Á (2010). "Application of neuro-fuzzy networks in the geomorphometric classification of mountainous landscapes of Venezuela." In: Morales C, Cuervo J, Franco H, compilers. Resúmenes. XV Congreso Colombiano de la Ciencia del Suelo. Risaralda, Pereira, Colombia: SCCS. p. 97.*
25. a, b, c*Viloria A, Núñez H, Viloria J (2012). Terrain Classifier System by Fuzzy Kohonen. Version 1.0. Caracas: UCV. Facultad de Ciencias, Escuela de Computación, Centro de Ingeniería de Software y Sistemas, Laboratorio de Inteligencia Artificial.*

26. [a, b, c] *Viloria JA, Pineda MC, Viloria-Botello A, Núñez Y, Valera A (2012). "Prediction of soil surface stoniness with neuro-fuzzy networks in Venezuelan plains." XIX Congreso Latinoamericano de la Ciencia del Suelo. XXIII Congreso Argentino de la Ciencia del Suelo; 2012 Apr 16–20; Mar del Plata, Argentina. 6 p.*
27. [^] *Valera A (2012). Artificial Intelligence Technologies: Artificial neural networks and fuzzy set theory for geomorphometric analysis of mountain landscapes. Editorial Académica Española. 108 p.*
28. [^] *Sevilla V (2014). "Comparison of two digital mapping methods with a conventional agrological study in the Canoabo River Basin, Carabobo State." [Promotion Work]. Caracas, Venezuela: Universidad Central de Venezuela. 117 p.*
29. [^] *Viloria JA, Viloria-Botello A, Pineda MC, Valera A (2016). "Digital modelling of landscape and soil in a mountainous region: A neuro-fuzzy approach." Geomorphology. **253**:199–207.*
30. [^] *Valera A (2015). "Soil and landscape inventory with the support of digital mapping techniques in mountainous areas. Case Study: Caramacate River Basin, Aragua State." [Doctoral Thesis in Soil Sciences]. Maracay, Estado Aragua, Venezuela: Universidad Central de Venezuela, Postgrado en Ciencias del Suelo. 263 p. doi:10.13140/RG.2.1.1714.3920.*
31. [^] *Valera A (2018). Geomorphometry and Edaphometry. Digital Mapping of Landscapes and Soils with Artificial Intelligence Techniques. Mauritius: Editorial Académica Española. 317 p.*
32. [^] *Rey JC, Martínez G, Micale E, Fernández N, Namias E, Polanco MA, Valera A (2015). "Soil mapping using fuzzy logic and its relationship with banana (musa AAA) yield." XXII Congreso Venezolano de la Ciencia del Suelo; San Cristóbal, Táchira, Venezuela. 6 p.*
33. [a, b] *Valera AR, Orta F (2018). "Application of geostatistical techniques and artificial neural networks in the delimitation of soil fertility classes." UNERG Agrocientífica. **1**(1):1–19.*
34. [^] *Tukey J (1977). Exploratory Data Analysis. Reading, EUA: Addison-Wesley Pub.*
35. [^] *Upchurch D, Edmonds WJ (1991). "Statistical procedures for specific objectives." In: Spatial variabilities of soils and landforms. 2nd ed. Madison: SSSA. pp. 49–71.*
36. [^] *Ovalles F (1992). Methodology for determining the represented surface by samples taken for fertility purposes. Serie B. Maracay: FONAIAP-CENIAP-IIAG. 44 p.*
37. [^] *Ovalles F, Rey JC (1994). "Internal variability of fertility units in soils of the Lake Valencia depression." Agron. Trop. **44**:41–65.*
38. [^] *Burrough P (1986). Principles of geographical information systems land resources assessment. Oxford: Clarendon Press. 193 p.*
39. [^] *Grunwald S, Rivero RL, Ramesh K (2007). "Understanding spatial variability and its application to biogeochemistry analysis." In: Sarkar D, Datta R, Hannigan R, editors. Developments in Environmental Science. Vol 5. Elsevier Ltd. pp. 443–463.*
40. [a, b] *Odeh IOA, McBratney AB, Chittleborough DJ (1992). "Soil pattern recognition with fuzzy c-means: application to classification and soil landform interrelationships." Soil Sci. Soc. Am. J. **56**:505–516.*
41. [^] *Marchetti A, Piccini C, Francaviglia R, Santucci S, Chiuchiarelli I (2010). "Estimating Soil Organic Matter Content by Regression Kriging." In: McBratney AB, Harteming AE, editors. Digital Soil Mapping. Bridging Research, Environmental Application, and Operation. New York. Chapter 20.*
42. [^] *McKay J, Grunwald S, Shi X, Long RF (2010). "Evaluation of the transferability of a knowledge-based soil-landscape model." In: Boettinger J, Howell DW, Moore AC, Hartemink AE, Kienast-Brown S, editors. Digital Soil Mapping: Bridging Research, Production and Environmental Applications. Heidelberg: Springer. pp. 165–177.*

Declarations

Funding: This research was supported by the Soil and Water Research and Extension Centre of the National Experimental University of the Central Plains "Romulo Gallegos" (CIESA-UNERG).
Potential competing interests: No potential competing interests to declare.