v1: 12 January 2024

Research Article

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

Peer-approved: 12 January 2024

Qeios, Vol. 6 (2024)
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 for precision agriculture and soil management decision-making. In this study, soil fertility assessment was carried out through the spatial distribution of thematic maps of individual properties and the subsequent integration into a digital mapping model of local fertility classes, as fundamental bases for the implementation of fertilization and amendment plans adjusted to soil status and crop requirements. For the evaluation of fertility, a systematic surface sampling was carried out at 70 sites in the "Agronomy" production field of the National University of the Central Plains "Romulo Gallegos", El Castrero sector, Juan German Roscio municipality, Guárico state, Venezuela. Ten soil variables were analyzed: pH (1:2.5), electrical conductivity (1:5), organic matter, available phosphorus, assimilable potassium, available calcium and magnesium, and the relative amounts of sand, silt, and clay. Soil property maps were produced by geostatistical analysis and interpolation by ordinary kriging, and artificial intelligence techniques based on an artificial neural network classification system were applied to generate soil fertility classes using the Fuzzy Kohonen Clustering Network (FKCN) algorithm by interpolating the values of the membership function for each of the classes. The reliability of the individual maps of each soil variable was obtained by cross-validation with a reliability level higher than 90%, with the exception of the variables % Clay and % Silt, which presented a reliability higher than 85%. The integration of the soil attribute maps and the combination of the values of belonging to each class produced a map integrated by five soil fertility categories. The final model of digital soil fertility classes presented a reliability equivalent to 86%, which indicated a high degree of homogeneity within the soil classes obtained for fertility purposes.

Corresponding author: Angel Rafael Valera Valera, angelvalera@unerg.edu.ve

Introduction

Soil fertility is an important quality resulting from the interaction between the physical, chemical, biological, and biochemical characteristics of the soil environment, which consists of the capacity to provide all the necessary conditions for plant growth and development. 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., (2022).

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, macro elements (phosphorus, potassium), secondary elements (calcium, magnesium), microelements (zinc, copper, iron, and manganese), and exchangeable acidity (aluminum, hydrogen). 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 by the management of climatic factors or irrigation and agricultural activities.

The spatial analysis of soil fertility facilitates decision-making when applying agronomic practices in productive spaces, allowing the appropriate supply of nutrients to the soil and minimizing the impact on the soil resource for the benefit of biodiversity (Shashikumar et. al, 2022). 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. Therefore, the systematic organization of soil data in geographical areas or land units is an opportunity to assess the spatial distribution of topsoil, to express the spatial variation of soil fertility through thematic maps, and to 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.

Within the spatial analysis techniques, geostatistical methods play an important role in the prediction of soil properties, where the interpolation method called ordinary kriging stands out (Webster y Oliver, 1990). However, the individual representation of the variables defining soil fertility does not cover the interest and the need to visualize the joint behaviors of soil fertility. Spatial analysis makes it possible to assess the variation of individual soil properties and the formation 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. This information serves as a basis for users to get a complete picture of the soil nutrient status of a sector on a single map and also contributes to decision-making on the most appropriate soil management (Padua et al., 2018; Shashikumar et al., 2022).

For the generation of soil classes, there are spatial analysis techniques based on artificial intelligence, such as fuzzy logic and artificial neural networks (ANN). 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 soil unit boundaries (Burrough et al. 2000). The combination of the potential of fuzzy sets and ANNs has developed a comprehensive unsupervised classification technique called the Fuzzy Kohonen Clustering Network (FKCN) (Lin and Lee, 1996; Bezdeck et al., 1992), which combines a self-organizing map (SOM) algorithm (Kohonen, 1982) and the Fuzzy C-means (FCM) algorithm (Bezdeck, 1981).

There are few research works in the field of soil science that take into account the combination of individual properties to express them as soil fertility categories. In this respect, 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 (Viloria, 2007), in geomorphological digital mapping (Valera and Viloria, 2009), Valera et al. (2010), Viloria et al. (2012), Valera (2012), Sevilla (2014), and Viloria et al. (2016), in the prediction of local soil properties and classes (Valera, 2015; Valera, 2018), in the study of soil and banana crop yield relationships (Rey et al., 2015), and in the delimitation of fertility classes (Valera and Orta, 2018).

This paper presents a study of spatial analysis of fertility classes through the prediction of chemical and physical properties of the soil obtained in laboratory analyses by means of geostatistical techniques, and their subsequent grouping by means of a fuzzy artificial neural network algorithm. To evaluate the spatial behaviors of soil fertility classes, the "Agronomy" production field of the National University of the Central Plains "Romulo Gallegos," located in the El Castrero sector, San Juan de los Morros parish, Juan German Roscio municipality, Guárico state (Venezuela), was considered.

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 University of the Central Plains "Romulo Gallegos," located in the El Castrero sector, San Juan de los Morros parish, Juan German Roscio municipality, Guarico state (Figure 1). The study unit is

doi.org/10.32388/0JAFR0.2

framed in an alluvial zone, with a slope of 3 to 5%. The soils in this area were formed from Quaternary geological materials, with a moderate pedogenetic development, and are of moderate fertility.

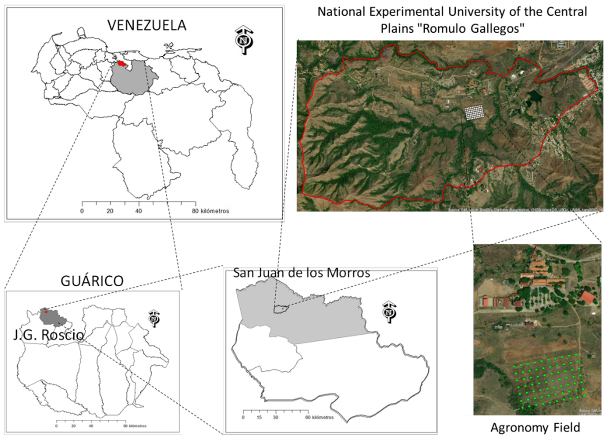

Figure 1. Relative location of the production field "Agronomy" in the basin of the river El Castrero, Guarico state, Venezuela.

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 Centre 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, $mgkg^{-1}$), assimilable 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 (%).

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 (1977) methodology of external and internal fences was used to detect the presence of outliers. Additionally, the normality test of Kolmogorov-Smirnov was performed to evaluate the distribution of the data.

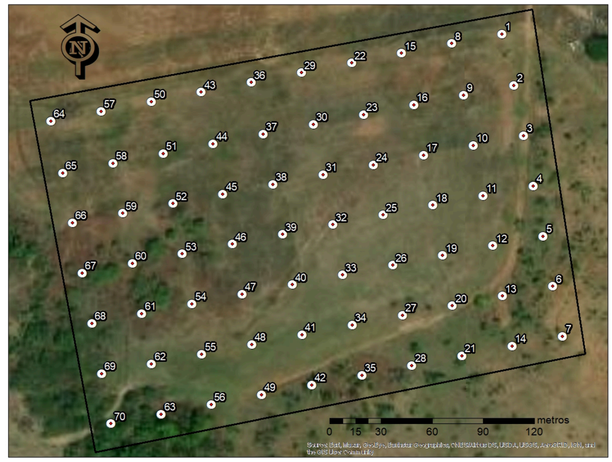

Figure 2. Distribution of soil sampling sites in the production field "Agronomy".

Interpolation of soil properties

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 [$\gamma(\boldsymbol{h})$], which is estimated with the following expression (Upchurch and Edmonds, 1991; Ovalles, 1992):

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

where $\boldsymbol{N}$ is the number of pairs of points separated by a given distance $\boldsymbol{h}$; $\boldsymbol{z(x_i)}$ is the value of the variable at a location $\boldsymbol{x}$; $\boldsymbol{z(x_{i+h})}$ is the value that the variable takes at another location located at a distance $\boldsymbol{h}$ from $\boldsymbol{x}$ (Ovalles and Rey, 1994). The semivariogram contains the information concerning the regionalized variable, whose parameters are: the nugget variance ($\boldsymbol{C_0}$), the structural variance ($\boldsymbol{C_1}$), the threshold or plateau ($\boldsymbol{C_0+C_1}$), and the range (A1), which indicates the distance within which there is spatial dependence (Burrough, 1986; Grunwald et al., 2007). 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

doi.org/10.32388/0JAFR0.2

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 (Webster and Oliver, 1990). 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), mean error standardized (MES), root-mean square (RMS), root-mean square standardized (RMSS), 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 SSE shows the deviation of the model obtained. The RMS 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 RMSS 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 RMS.

Digital soil fertility class model

The neuro-fuzzy FKCN algorithm, implemented in a Java environment (Windows) by Viloria (2012), 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 $\boldsymbol{d_{ij}}$ existing between an input pattern $\boldsymbol{X_j}$ and the node weight ω_i is computed, with $\boldsymbol{i}$ = 1, 2,..., $\boldsymbol{c}$, where $\boldsymbol{c}$ represents the number of classes of the model to be estimated. Subsequently, the membership layer plots the distances $\boldsymbol{d_{ij}}$ into membership values $\boldsymbol{U_{ij}}$, where $\boldsymbol{U_{ij}}$ represents the degree of membership of an input pattern $\boldsymbol{X_j}$ to a class $\boldsymbol{c}$. 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.

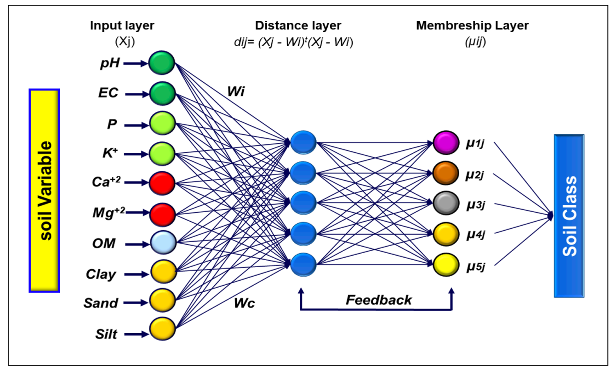

Figure 3. Structure of the fuzzy neural network used in soil class prediction.

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. (1992), which relates the Fuzziness Performance Index (FPI) to the number of classes. These parameters are obtained using the Fuzzy Kohonen Clustering Networks (FKCN) algorithm (Lin and Lee, 1996) of the FKCN program (Viloria et al., 2012). 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 (Odeh et al., 1992). Mathematically, it is defined as:

$$FPI = 1 - [(cF - 1)/(c - 1)] \tag{2}$$

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

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

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

doi.org/10.32388/0JAFR0.2

of classes ($c = 3$ to 10), b) fuzzy exponent $\phi = 1.10$ to 1.60 with increments of 0.10; c) a maximum of 300 iterations, and d) stopping criterion ($\varepsilon = 0.0001$). 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
MO (%)	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: Assimilable 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.

Interpolation of soil properties

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	0,000	0,110	56	0,11	0,0
EC (dS m^{-1})	Gaussian	0,000	0,000	59	0,00	0,0
P (mg kg^{-1})	Spherical	0,100	82,49	93	82,6	0,1
K (cmol (+) kg^{-1})	Circular	0,080	0,05	59	0,13	60,4
Ca (cmol (+) kg^{-1})	J-Bessel	0,016	0,036	59	0,05	30,7
Mg (cmol (+) kg-1)	Stable	0,000	0,058	59	0,06	0,0
OM (%)	Gaussian	0,030	0,410	59	0,40	6,8
Clay (%)	Gaussian	0,000	125,64	59	125,6	0,0
Sand (%)	Spherical	78,28	22,06	59	100,3	78,0
Silt (%)	J-Bessel	0,000	137,02	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.

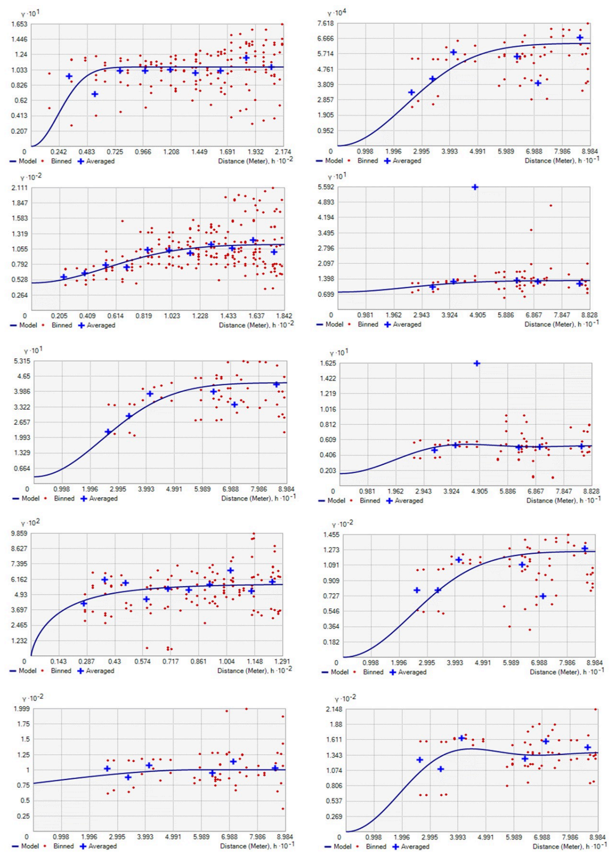

Figure 4. Semivariogram of soil variables in the production field "Agronomy".

doi.org/10.32388/0JAFR0.2

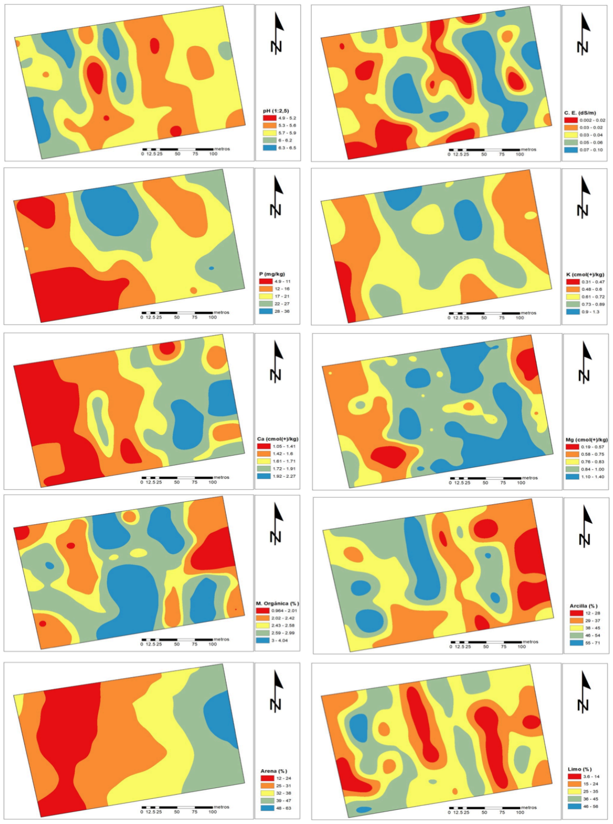

Figure 5. Model maps of soil variables in the production field "Agronomy".

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. In other words, a spatial dependence range of 59 m stands out for all models, with the exception of the semivariogram of the available P variable, which has a range 1.5 times the average.

Assessing the reliability 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, MES, and ASE indices, can be observed.

doi.org/10.32388/0JAFR0.2

Variable	Regression Function	Index					
		ME	MES	RMS	RMSE	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, *MES*: mean error standardized, *RMS*: root-mean square, *RMSS*: root-mean square standardized, ASE: Average standard error, CL: Confidence level. EC: electrical conductivity, P: available phosphorus, K: assimilable 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 RMS, small ASE close to RMS, RMSE 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 RMS. For all the cases evaluated, the RMS values are lower than the standard deviation and are therefore adequate for the evaluation of the prediction models (Marcheti et al., 2010).

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 aϕ 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.

doi.org/10.32388/0JAFR0.2

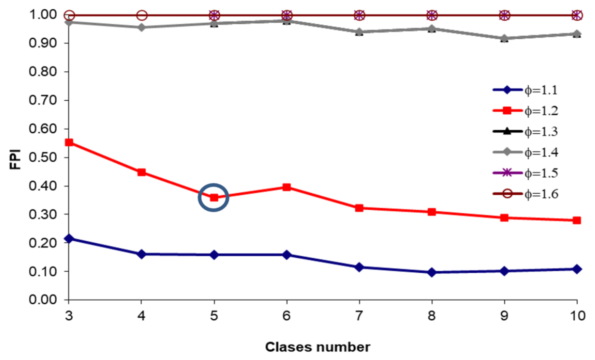

Figure 6. Variation of the fuzzy performance index (FPI) as a function of the number of soil classes.

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).

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.

doi.org/10.32388/0JAFR0.2

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	17	18	25	22
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
MO (%)	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: Assimilable potassium, Ca: Available calcium, Mg: Available magnesium, OM: Organic matter.

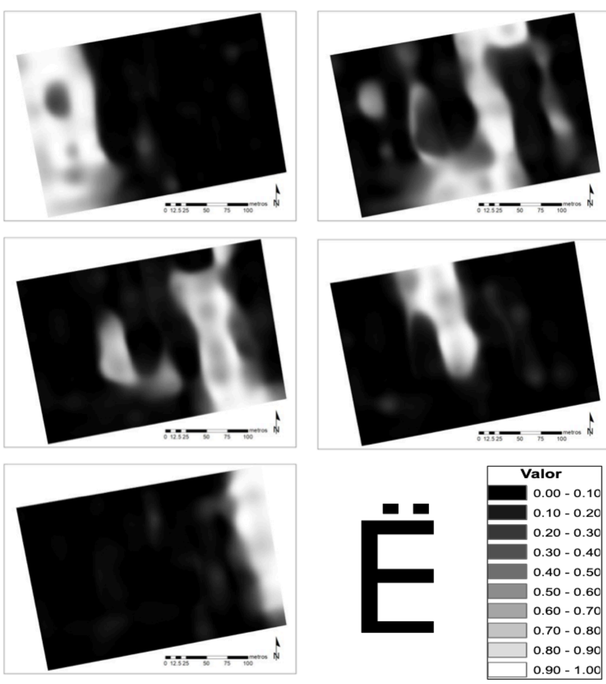

Figure 7. Maps of membership function values for each of the soil fertility classes.

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.

doi.org/10.32388/0JAFR0.2

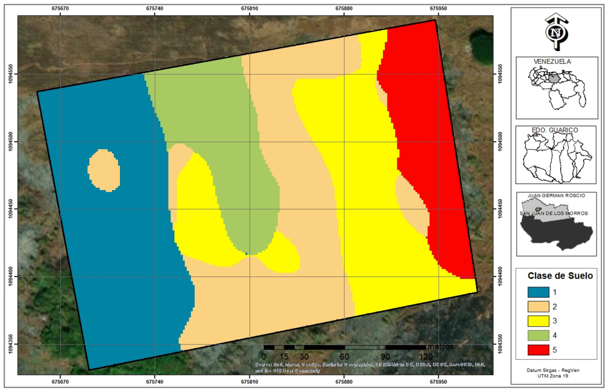

Figure 8. Soil fertility class distribution model of the production field "Agronomy".

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.

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).

 doi.org/10.32388/0JAFR0.2

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 results of the validation of the FKCN approach demonstrated that it is an alternative for the generation of soil fertility classes. These results are slightly higher than those obtained by various researchers, Zhu et al. (2008), McKay et al. (2010), and Valera and Orta (2018), whose research expressed a reliability of 76, 73.7, and 80.1% respectively for the soil maps obtained.

Conclusions

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

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

The establishment of productive plots should not exceed the range of spatial dependence of the fertility attributes, whose mode is 59 m, in order to include the variability of the assessed soils. Therefore, the area for the establishment of productive plots that guarantee the homogeneity of the internal structure of the soils should not be larger than 1.0 ha to allow the representativeness of the soil.

The evaluation of the digital neuro-fuzzy model indicated that the spatial prediction of soil fertility classes corresponds to what is expected in the studied sector, as the reliability was equivalent to 86%.

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 soil resource management in the area.

Acknowledgements

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).

References

- Bezdek J.C. & R. Ehrlich, Full. W. 1984. FCM: the fuzzy c-means clustering algorithm. Computers and Geosciences. 10: 191-203.
- Bezdek, J.C., E.C. Tsao, Pal, N.R. 1992. Fuzzy Kohonen Clustering Networks, in Proc. IEEE Int. Conf. on Fuzzy Systems 1992 (San Diego), pp. 1035-1043.
- Burrough, P. 1986. Principles of geographical information systems land resources assessment. Clarendon Press Oxford. 193 p.
- Burrough, P.A., P.F.M.Van Gaans, & MacMillan, R.A. 2000. High-resolution landform classification using fuzzy k-means. Fuzzy Sets and Systems 113: 37–52.
- Grunwald, S., R.L. Rivero & Ramesh, K. 2007. Understanding spatial variability and its application to biogeochemistry analysis. In: D. Sarkar, R. Datta

doi.org/10.32388/0JAFR0.2

and R. Hannigan (Ed.). Developments in Environmental Science, 5:443-463. Elsevier Ltd. ISSN: 1474-8177.

- Kohonen, T. 1982. Analysis of a simple self-organizing process. Biological Cybernetics 44: 135-140.
- Lin, C. & Lee, C. 1996. Neural fuzzy systems. Prentice Hall, Inc. New Jersey, USA. 797 p.
- Marchetti, A.; C. Piccini, R. Francaviglia, S. Santucci, & Chiuchiarelli, I. 2010. Estimating Soil Organic Matter Content by Regression Kriging. In: Digital Soil Mapping. Bridging Research, Environmental Application, and Operation. (Ed. A.B. McBratney and A.E. Harteming). Chapter 20. New York. 241 p. ISBN 978-90-481-8862-8.
- McKay, J., Grunwald, S., Shi, X. & Long, R.F. 2010. Evaluation of the transferability of a knowledge-based soil-landscape model. In: Boettinger J., D.W. Howell, A.C. Moore, A.E. Hartemink, & S. Kienast-Brown (eds.). Digital Soil Mapping: Bridging Research, Production and Environmental Applications. pp. 165-177. Springer, Heidelberg.
- Odeh, I.O.A.; A.B. McBratney; & D.J. Chittleborough. 1992. Soil pattern recognition with fuzzy c-means: application to classification and soil landform interrelationships. Soil Sci. Soc. Am. J. 56: 505- 516.
- Ovalles, F. 1992. Metodología para determinar la superficie representada por muestras tomadas con fines de fertilidad. FONAIAP-CENIAP-IIAG. Maracay. Serie B. 44 p.
- Padua, S., T. Chattopadhyay, S. Bandyopadhyay, S. Ramchandran, R.K. Jena, P. Ray, P. Deb Roy, U. Baruah, K D. Sah, S.K. Singh & S. K. Ray. 2018. A simplified soil nutrient information system: study from the North East Region of India. Current Science, 114 (6): 1241-1249.
- Rey, J.C., G. Martínez, E. Micale, N. Fernández, E. Namias, M.A. Polanco, & Valera, A. 2015. Mapeo de suelos por medio de lógica difusa y su relación con el rendimiento de banano (musa AAA). XXII Congreso Venezolano de la Ciencia del Suelo. San Cristóbal, Táchira. Venezuela. 6 p.
- Sevilla, V. 2014. Comparación de dos métodos de cartografía digital con un estudio agrológico convencional en la Cuenca del Rio Canoabo, Estado Carabobo. Trabajo de Ascenso para optar a la Categoría de Asistente. Universidad Central de Venezuela. Caracas, Venezuela. 117 p.
- Shashikumar, B.N., Kumar, S., George, K.J. Kumar & Singh, A. 2022. Soil variability mapping and delineation of site-specific management zones using fuzzy clustering analysis in a Mid-Himalayan Watershed, India. Environ Dev Sustain. https://doi.org/10.1007/s10668-022-02411-6
- Srinivasan, R., Shashikumar, B.N., & Singh, S.K. 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. https://doi.org/10.1007/s12524-021-01473-9
- Tukey, J. 1977. Exploratory Data Analysis. Addison-Wesley Pub. Reading, EUA.
- Upchurch, D., Edmonds, W.J. 1991. Statistical procedures for specific objectives. In: Spatial variabilities of soils and landforms. SSSA Special publication No. 28. 2ª Ed. SSSA. Madison. pp: 49-71.
- Valera A.R. & M. Orta F. 2018. Aplicación de técnicas geoestadísticas y redes neuronales artificiales en la delimitación de clases de fertilidad de suelos. UNERG Agrocientífica 1 (1): 1-19
- Valera, A. & Viloria, J.A. 2009. Aplicación de técnicas de inteligencia artificial en el modelado de unidades de paisaje en la cuenca del río Güey, Maracay - estado Aragua. Memorias XVIII Congreso Venezolano de la Ciencia del Suelo. Santa Bárbara, Zulia. Venezuela. 7 p.
- Valera, A. 2012. Tecnologías de Inteligencia Artificial: Redes neuronales artificiales y teoría de conjuntos difusos para el análisis geomorfométrico de paisajes de montaña. Editorial Académica Española. 108 p. ISBN: 978-3-8484-7612-1.
- Valera, A. 2015. Inventario de suelos y paisajes con apoyo de técnicas de cartografía digital en áreas montañosas. Caso Cuenca del Río Caramacate, Estado Aragua. Tesis de doctorado en Ciencias del Suelo. Universidad Central de Venezuela. Postgrado en Ciencias del Suelo. Maracay, Estado Aragua, Venezuela. 263 p. DOI: 10.13140/RG.2.1.1714.3920
- Valera, A. 2018. Geomorfometría y Edafometría. Cartografía Digital de Paisajes y Suelos con Técnicas de Inteligencia Artificial. Editorial Académica Española. Mauritius. ISBN: 978-620-2-12102-6. 317p.
- Valera, A., J.A. Viloria &Viloria, Á. 2010. Aplicación de redes neuro-difusas en la clasificación geomorfométrica de paisajes montañosos de Venezuela. En: Resúmenes. XV Congreso Colombiano de la Ciencia del Suelo. Morales, C., J. Cuervo y H. Franco (compiladores). SCCS. Risaralda, Pereira. Colombia. p.97.
- Viloria J.A, A. Viloria-Botello, M.C. Pineda, & Valera, A. 2016. Digital modelling of landscape and soil in a mountainous region: A neuro-fuzzy approach. Geomorphology Vol. 253:199-207.
- Viloria, A. 2007. Estimación de Modelos de clasificación de paisaje y predicción de atributos de

suelos a partir de imágenes satelitales y modelos digitales de elevación. Trabajo Especial de Grado. Universidad Central de Venezuela. Caracas, Venezuela. 88 p.

- Viloria, A., Núñez, H. & J. Viloria (2012). Sistema Clasificador del Terreno por Fuzzy Kohonen. Versión 1.0. UCV. Facultad de Ciencias. Escuela de Computación. Centro de Ingeniería de Software y Sistemas. Laboratorio de Inteligencia Artificial. Caracas, Venezuela.
- Viloria, J.A., M.C. Pineda, A. Viloria-Botello, Y. Núñez, & Valera, A. 2012. Predicción de pedregosidad superficial del suelo con redes neuro-difusas en llanos venezolanos. XIX Congreso Latinoamericano de la Ciencia del Suelo. XXIII Congreso Argentino de la Ciencia del Suelo. Mar del Plata, Argentina - 16 al 20 de abril de 2012. 6 p.
- Webster, R. & Oliver, M.A. 1990. Statistical Methods in Soil and Land Resource Survey. Oxford University Press. Oxford, RU. 316p.
- Zhu, A. X., Yang, L., Li, B., Qin, C., English, E., Burt, J.E. & Zhou, C. 2008. Purposive Sampling for Digital Soil Mapping for Areas with Limited Data. In: Hartemink, A.E.; Mendonça-Santos, M. L., A.B. McBratney, A. B., eds, Digital Soil Mapping with Limited Data, Springer-Verlag: New York, pp 233-245.

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.

doi.org/10.32388/0JAFR0.2