UNERG AGRO-Científica
2024 5(1): 299-316
https://unerg.edu.ve/unerg-agro-cientifica/
ISSN: 2665-0061 (Online)

SOIL FERTILITY CLASSES WITH GEOSTATISTICAL TECHNIQUES AND FUZZY LOGIC IN EXPERIMENTAL FIELDS

Ángel Valera[1,2*], Eladio Arias[2], Ayuramy Martínez[3]

[1] Universidad Rómulo Gallegos, Centro de Investigación y Extensión en Suelos y Aguas (CIESA-UNERG), San Juan de los Morros, Estado Guárico, Venezuela, e-mail: angelvalera@unerg.edu.ve

[2] Universidad Rómulo Gallegos, Área de Ingeniería Agronómica, San Juan de los Morros, Estado Guárico, Venezuela, e-mail: eladioariasrod1956@gmail.com

[3] Instituto Nacional de Investigaciones Agrícolas, Calabozo, Estado Guárico, Venezuela, e-mail: amartinez@inia.gob.ve

[*] Autor correspondiente

Received: 05 – 03 - 2024; **Accepted**: 15 - 05 - 20243; **Published**: 28 - 06 - 2024

ABSTRACT

The establishment of experimental plots requires assessments of the continuous variation of soil properties and interpretation based on the spatial dependence of the most relevant variables. In order to predict the variation of soil fertility classes, two alternative spatial analysis techniques were combined. The first technique corresponds to the use of geostatistical analysis for the interpolation of individual soil properties of chemical and physical nature. The second technique consisted of the application of an unsupervised classification system based on fuzzy set theory using the FCM (fuzzy c-means) algorithm for the generation of a digital soil fertility class model. For this purpose, a surface sampling was carried out in 110 sites in the Experimental Field "El Rastro", El Rastro sector, Francisco de Miranda municipality, Guárico state (Venezuela). Ten soil variables were analyzed: pH, electrical conductivity, organic matter, available phosphorus, assimilable potassium, available calcium and magnesium, and the relative amounts of sand, silt and clay. The measured variables were interpolated at each sampling point using ordinary *kriging* and adjusted using theoretical semivariogram. An inductive method was used to obtain soil fertility classes, and a soil class model was obtained based on the integration of the variables. The reliability of the individual maps of each soil variable was cross-validated, an analysis of variance was applied to corroborate the predictive capacity of the variables, and multivariate statistics were used to evaluate the final model. The digital fertility map indicated that seven fertility classes predominate in the study area, with a reliability of more than 85%, indicating a high degree of homogeneity within the defined soil classes.

Keywords: Soil Fertility; Geoestatistics; Fuzzy Logic; FCM Algorithm.

CLASES DE FERTILIDAD DEL SUELO CON TÉCNICAS GEOESTADÍSTICAS Y LÓGICA DIFUSA CON FINES EXPERIMENTALES

RESUMEN

El establecimiento de parcelas experimentales requiere de evaluaciones de la variación continua de propiedades del suelo y la interpretación con base en la dependencia espacial de las variables más relevantes. Con la finalidad de predecir la variación de las clases de fertilidad del suelo se combinaron dos técnicas alternativas de análisis espacial. La primera técnica corresponde a la utilización del análisis geoestadístico para la interpolación de propiedades individuales del suelo de naturaleza química y física.

La segunda técnica consistió en la aplicación de un sistema de clasificación no supervisado basado en la teoría de conjuntos difusos mediante el algoritmo FCM (c-medias difuso), para la generación de un modelo digital de clases de fertilidad del suelo. Para tal fin, se realizó un muestreo superficial en 110 sitios en terrenos del Campo Experimental “El Rastro”, sector El Rastro, municipio Francisco de Miranda-estado Guárico (Venezuela). Se analizaron diez variables del suelo: pH, conductividad eléctrica, materia orgánica, fósforo disponible, potasio asimilable, calcio y magnesio disponible, y las cantidades relativas de arena, limo y arcilla. Las variables medidas fueron interpoladas en cada punto de muestreo utilizando *kriging* ordinario y ajustadas mediante semivariogramas teóricos. Se utilizó un método inductivo para la obtención de las clases de fertilidad del suelo, y se obtuvo un modelo de clases de suelo basado en la integración de las variables. La confiabilidad de los mapas individuales de cada variable del suelo se realizó mediante validación cruzada, para corroborar la capacidad predictiva de las variables se aplicó un análisis de varianza, y para la valoración del modelo final se empleó estadística multivariada. El mapa digital de fertilidad indicó que en el área de estudio predominan siete clases de fertilidad, las cuales presentaron una confiabilidad superior al 85%, lo que indicó un alto grado de homogeneidad dentro de las clases de suelo definidas.

Palabras clave: Fertilidad del Suelo; Geoestadística; Lógica Difusa; Algoritmo FCM

INTRODUCTION

Soil fertility is one of the most important soil qualities that can be greatly affected by its use and management, and is of great utility for the recognition of soil nutritional status and for fine-tuning crop recommendations through nutrient applications from organic and inorganic sources. 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; Valera y Arias, 2023). Therefore, knowledge of the spatial variation of soil fertility in experimental sites is very important for the definition and delimitation of the establishment of homogeneous plots, and avoiding overlaps between treatments and trials. The manual representation of soil fertility classes requires the elaboration of individual maps for each of the variables, and the subsequent superimposition of these for the definition of boundaries, which implies biases and low precision in the final result. The cartographic representation of soil fertility facilitates decision-making when establishing experimental plots and trials for research purposes.

Conventionally, fertility assessment is done through the analysis of soil test results, and the study or experimental areas are considered homogeneous, i.e. the spatial variation of attributes is not taken into account. Over time, some cartographic techniques have been implemented, which allow obtaining basic and reliable information on the spatial expression of soil properties. Within these techniques, geostatistical methods play an important role for the spatial prediction of soil properties, where the interpolation method called ordinary *kriging* stands out. However, the individual representation of the variables defining soil fertility does not cover the interest and the need to visualize the behavior as a whole, in a model of spatial variation of soil fertility classes.

Digital soil mapping (DSM) allows the integration of various models of spatial variation of individual soil properties to obtain soil classes, in order to support decision-making on area definition as a basis for site-specific management and for the promotion of precision agriculture. The application of DSM through the assessment of the spatial variation of soil fertility attempts to divide the soil continuum into classes, which exhibit a greater homogeneity of the combined influence of the variables considered in the soil analysis.

There is little research in the field of soil science that takes into account the combination of continuous variation of individual properties to express them as soil fertility categories. In this respect, the application of fuzzy set theory - as part of artificial intelligence technologies - has given a great boost to DSM both in predicting properties and in obtaining soil classes. In Venezuela, artificial intelligence technologies 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), Núñez (2011), Viloria *et al* (2012), Valera (2012) and Viloria *et al* (2016), in the prediction of soil properties and local soil classes (Valera, 2015; Valera, 2018) and in the study of soil and banana crop yield relationships (Rey *et al., 2015*).

This paper presents a DSM study for the definition of soil fertility classes, through the prediction of soil chemical and physical properties obtained in laboratory analyses, and their subsequent grouping. To evaluate the spatial behavior of soil fertility classes, the Experimental Field "El Rastro" from the National Experimental University of the Central Plains "Romulo Gallegos", located on the national road Guardatinajas El Rastro sector, El Rastro parish of the Francisco de Miranda autonomous municipality, Guárico state (Venezuela), was considered. The main purpose of the research was the spatial prediction of soil fertility classes through the theory of fuzzy or fuzzy sets 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 and agronomic trials, as well as field research and evaluations for experimental purposes to be developed in the sector studied.

MATERIALS AND METHODS

Study Area

The study area where the digital soil mapping test was carried out is located on the grounds of the Experimental Field "The Rastro" from the National Experimental University of the Central Plains "Romulo Gallegos", located in the sector El Rastro, national road Guardatinajas, parish El Rastro, Francisco de Miranda municipality, Guárico State, Venezuela (Figure 1). The study unit is framed in a sub-recessional plain, with a slope of 2 to 4%. The soils in this area were formed from Quaternary geological materials, with an incipient pedogenetic development, and are of low to moderate fertility.

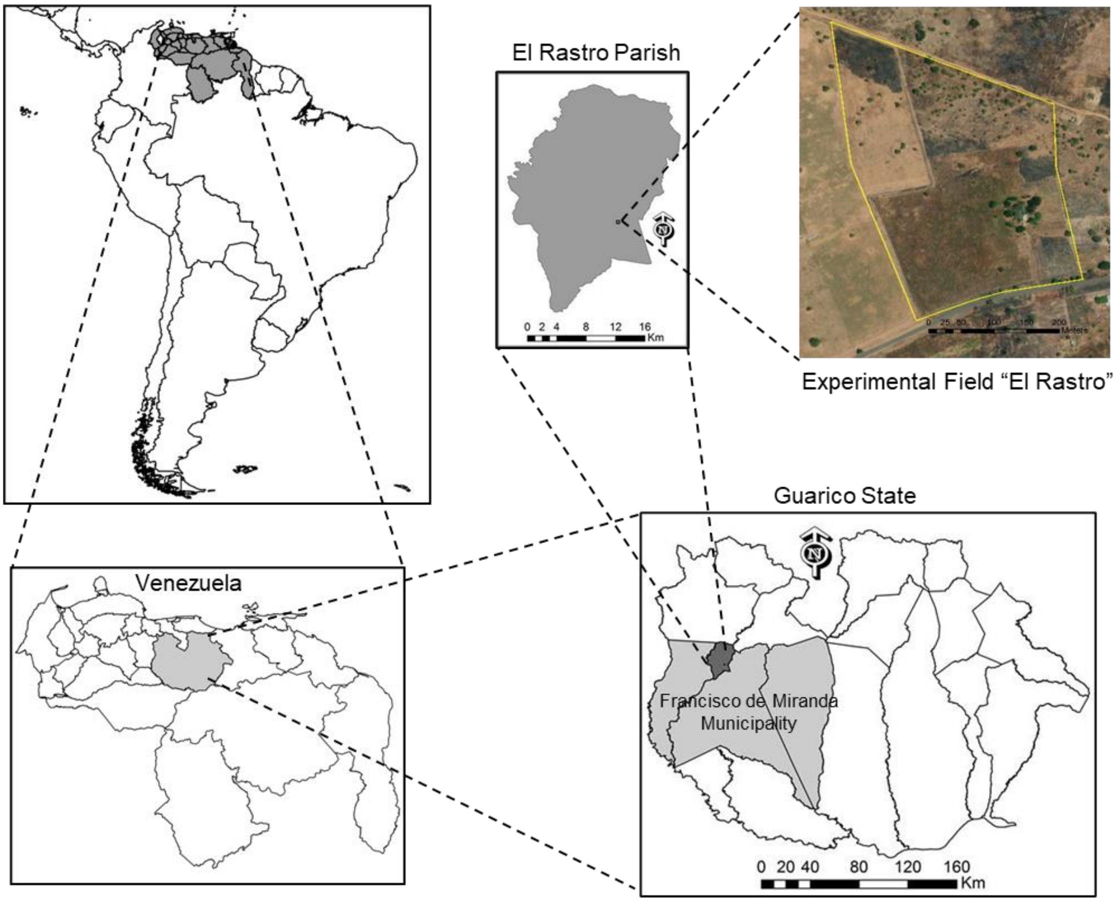

Figure 1. Relative location of the Experimental Field "El Rastro" in Miranda municipality, Guárico state, Venezuela.

Soil sampling

For the soil evaluation, a systematic sampling was carried out in the superficial horizon at 20 cm depth, in grids spaced at 20 m, for a total of 110 soil samples in an area of 4.85 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 Rómulo 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, $mgkg^{-1}$), calcium (Ca, $mgkg^{-1}$) and available magnesium (Mg, $mgkg^{-1}$), and the relative amounts of sand, silt and clay (%).

Statistical analysis

Soil property data were subjected to exploratory analysis (EDA) with the support of InfoStat software (Di Rienzo *et al*., 2019), in order to calculate descriptive statistics, such as: mean, median, variance, coefficient of variation, maximum and minimum values, and skewness and kurtosis indices. Tukey's (1977) outer and inner fences methodology 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.

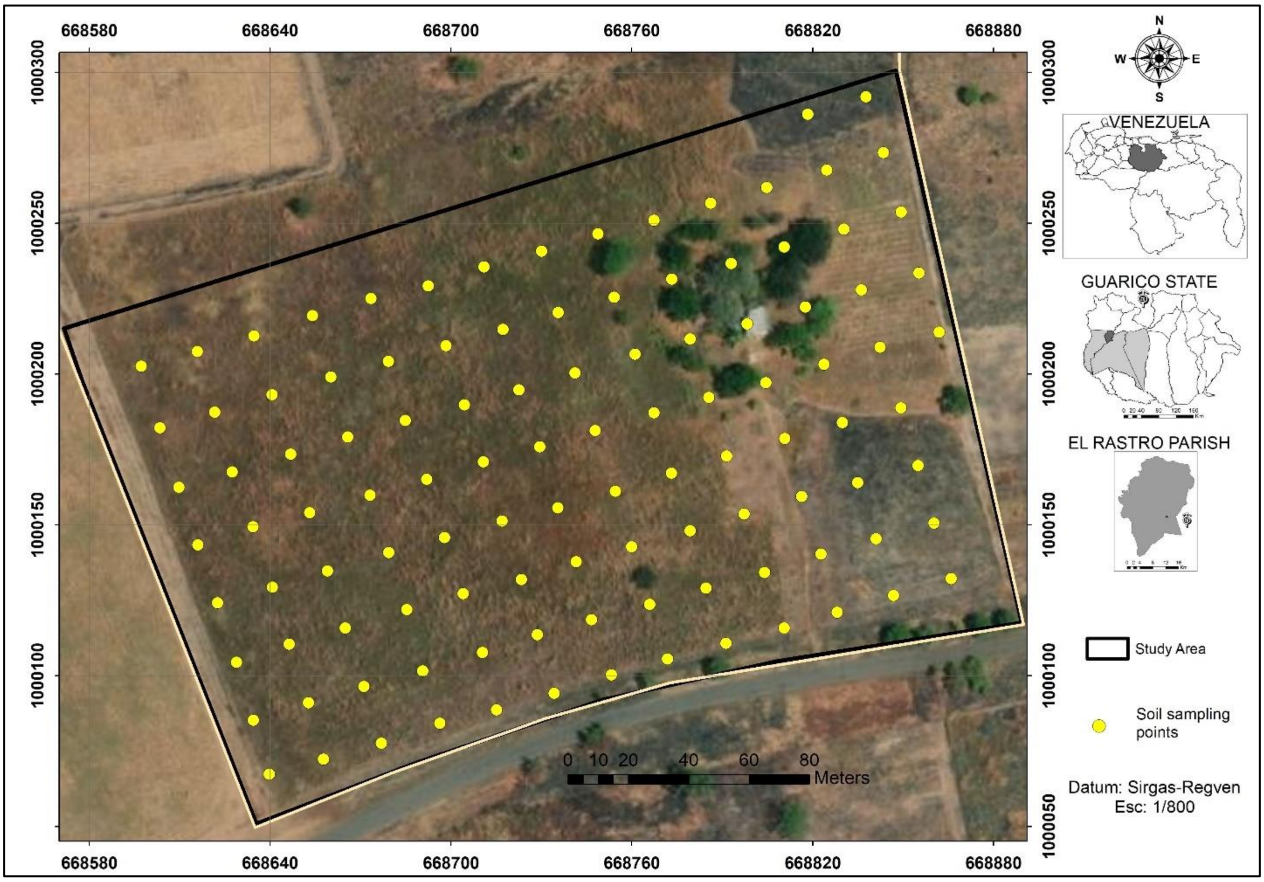

Figure 2. Distribution of soil sampling sites in the Experimental Field "El Rastro".

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 [γ ***(h)***], which is estimated with the following expression (Upchurch and Edmonds, 1991; Ovalles, 1992):

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

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 in a location $\boldsymbol{x}$; $\boldsymbol{z(x_{i+h})}$ is the value that the variable takes in 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 ($\boldsymbol{C_0 + C_1}$) and the range ($\boldsymbol{A_1}$), which indicates the distance within which spatial dependence exists

(Burrough, 1986; Grunwald *et al.,* 2007). The estimation of empirical semivariogram of soil properties and fitting to mathematical models was performed with the *Vesper* 1.6 program (Minasny *et al*., 2002). The fitted parameters were used to obtain optimal estimates of soil variables at the unsampled sites by interpolation using the ordinary *kriging* method (Webster and Oliver, 1990). Soil property models were generated from the total data and the accuracy of the maps was verified by cross-validations. Three indices were used in the evaluation: root mean square error (RMSE), mean error (ME) and mean absolute error (MAE). The RMSE assesses the accuracy of the prediction and measures the amount of error between two data sets, i.e. it compares a predicted value and an observed or known value; the ME assesses the systematic error and indicates the presence of under- or overestimation of the model; and the MAE ensures that the error result is strictly positive.

Digital soil fertility class model

The Fuzzy c-Means (FCM) algorithm, implemented in the FuzMe program by Minasny and McBratney (2002), was used to generate the digital soil class model. The algorithm optimally divides a dataset into a number of classes and computes the memberships or degrees of membership of each of the elements in each of the categories. The objective of the FCM algorithm (Bezdek, 1981; Bezdek *et al.,* 1984) is to minimize the weighted root mean square sum of the distances between the points $\boldsymbol{Z_k}$ and the center of the class $\boldsymbol{C_k}$, and the distances $\boldsymbol{d^2_{ik}}$, are weighted with the membership value ***i*,*k*.** Therefore, the objective function is:

$$J(Z;U,C) = \sum_{j=1}^{n} \sum_{k=1}^{c} (\mu_{ik})^{\phi} d_{ik}^{2} \tag{2}$$

where $\boldsymbol{Z}$ = {z_1, z_2 , ..., z_n} is the data to be classified, $\boldsymbol{U}$ = [μ_{ik}], is the fuzzy partition matrix of Z, $\boldsymbol{C}$ = [c_1, c_2 , ..., c_c] is the vector of centroids or patterns of the classes to be determined, $\boldsymbol{d^2_{ik}}$ is the squared distance between ***ik***, and $\phi \in [1,\infty)$ is a weighting exponent that determines the degree of fuzziness of the resulting classes. The membership function $\boldsymbol{\mu}$ from the *i-th* object to the *k-th* cluster in the ordinary fuzzy *k-means* algorithm employs the distance ***d*** used for similarity, and the fuzzy exponent (ϕ) to determine the amount of fuzziness:

$$\mu_{ik} = \left[(d_{ik})^2\right]^{-1/(\phi-1)} / \sum_{k=1}^{c} \left[(d_{ik})^2\right]^{-1/(\phi-1)} \tag{3}$$

Once the membership intensities have been determined, the centroids of the classes ($\boldsymbol{C_k}$) are calculated using the following equation:

$$c_k = \sum_{i=1}^{n} (\mu_{ik})^{\phi} x_i \,/\, \sum_{i=1}^{n} (\mu_{ik})^{\phi} \tag{4}$$

As for the initialization process, the FCM works by means of an iterative procedure that starts with a random distribution of the soil samples to be classified into ***k*** classes (De Gruijter and McBratney, 1988).

S*oil fertility classes*

In order to obtain the best fuzzy class model, an inductive approach was used, based on the procedure of Odeh *et al.* (1992), which relates the *Fuzziness Performance Index* (FPI) and the *modified partition entropy* (MPE) to the number of classes. These parameters are obtained using the *Fuzzy c-Means* (FCM) algorithm (Bezdek, 1981; Bezdek *et al*., 1984) of the *FuzMe* 3.5 program (Minasny and McBratney, 2002). The selection of the optimal number of classes in FCM was performed by repetition of the classification for a range of number of classes. For each clustering obtained, two classification parameters are generated, such as the FPI and the *modified partition entropy* (MPE). The FPI estimates the degree of fuzziness generated by each specific number of classes. Mathematically, it is defined as:

$$FPI = 1 - [(cF - 1)/(c - 1)] \tag{5}$$

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

$$F = (1/n)\sum_{i=1}^{n}\sum_{k=1}^{c}(\mu_{ik})^2 \tag{6}$$

F is conceptually comparable to the ratio of the set of within-class variances to the between-class variance and is close to 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: *i)* number of classes (c= 6 to 12), *ii)* fuzzy exponent (ϕ) = 1.1 to 1.6 with increments of 0.1*; iii)* a maximum of 300 iterations, and *iv)* stopping criterion (ε= 0.0001). The *Mahalanobis* metric distance was used in the calculations, which takes into account the correlation found between some soil properties of the studied area.

Assessment of the predictive ability of soil fertility classes

To evaluate the predictive capacity of the classes obtained by fuzzy clustering, a one-factor analysis of variance (S^2) was performed using the complement of relative variance (1-rv) (Beckett and Burrough, 1971), in order to verify the effect of edaphic properties on the differentiation of soil fertility classes in the studied sector. This index is analogous to the coefficient of determination and expresses the proportion of the variance that can be attributed to the classification. Under this criterion, for a classification to be worthwhile, the average intraclass variance should be less than the total variance (Webster and Oliver, 1990). Finally, the final model was validated with the original clustered cases, using the *Mahalanobis Distance* (D^2) as a multivariate descriptive statistic, derived from the canonical discriminant analysis.

RESULTS AND DISCUSSION

Statistical analysis

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

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

Variable[1]	**Min**	**Max**	**Mean**	**Median**	**K**	**As**	**SD**	**Variance**	**CV** (%)
pH (1:2.5)	4,2	6,0	5,1	5,1	-0,28	0,08	0,4	0,1	7,4
EC (dS m^{-1})	0,008	0,018	0,012	0,011	0,84	1,26	0,003	0,000	23,1
P (mg kg^{-1})	1,0	17,0	6,3	5,0	1,75	1,52	4,1	17,0	65,6
K (mg kg^{-1})	2,0	196,0	86,2	89,5	-0,82	0,15	51,0	2.601,0	59,1
Ca (mg kg^{-1})	80	420	227,5	220,0	0,55	0,80	76,5	5.855,8	33,6
Mg (mg kg^{-1})	3,0	441,0	115,9	77,5	1,37	1,31	105,9	11.204,6	91,3
OM (%)	0,4	2,5	1,1	1,2	1,48	0,55	0,4	0,2	34,4
Clay (%)	9,6	24,9	17,0	17,3	-0,30	-0,01	3,3	10,6	19,1
Sand (%)	59,8	73,1	67,9	67,8	0,12	0,08	2,7	7,3	4,0
Silt (%)	9,3	23,3	15,1	14,6	-0,11	0,57	2,9	8,4	19,1

[1]Number of data: 110, K: Kurtosis, As: Asymmetry, 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 variables P, K, Ca and Mg. At the same time, the greatest dispersion of the data is shown by the variables themselves, 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 of the data.

According to the skewness or asymmetry coefficient, the variables pH, K, % clay and % sand, comply with the normal probability distribution function, and geostatistical methods can be applied to the data. However, for %OM and Ca it is necessary to perform a data transformation (normalization) of square root type; and for P, Mg and EC, it is necessary to perform a logarithmic transformation for the subsequent application of a geostatistical method to the data.

The application of the test for external and internal fences indicated that the variables considered do not present outliers. As for the normality test, the variables K and OM, come from normal populations, since the values of the *Kolmogorov-Smirnov* Z test are highly significant (p> 0.05) (Table 2). However, for the rest of the data it was necessary to perform data transformation.

Table 2. Normality test of the soil data set of the Experimental Field "El Rastro".

Variable	Statistician	df	Sig.[1]
pH water (1:2.5)	0.095	110	0.016
EC water (dS m^{-1})	0.225	110	0.000
P (mg kg^{-1})	0.194	110	0.000
K (mg kg^{-1})	0.078	110	0.097
Ca (mg kg^{-1})	0.126	110	0.000
Mg (mg kg^{-1})	0.158	110	0.000
OM (%)	0.070	110	0.200
Clay (%)	0.100	110	0.009
Sand (%)	0.139	110	0.000
Silt (%)	0.155	110	0.000

[1]Significance level α = 0.05; df: degrees of freedom; n= 110.

Interpolation of soil properties

The estimation of the empirical semivariogram of the soil variables were fitted to Gaussian, spherical and exponential mathematical models, respectively (Figure 3), considering the isotropic behavior of the variables.

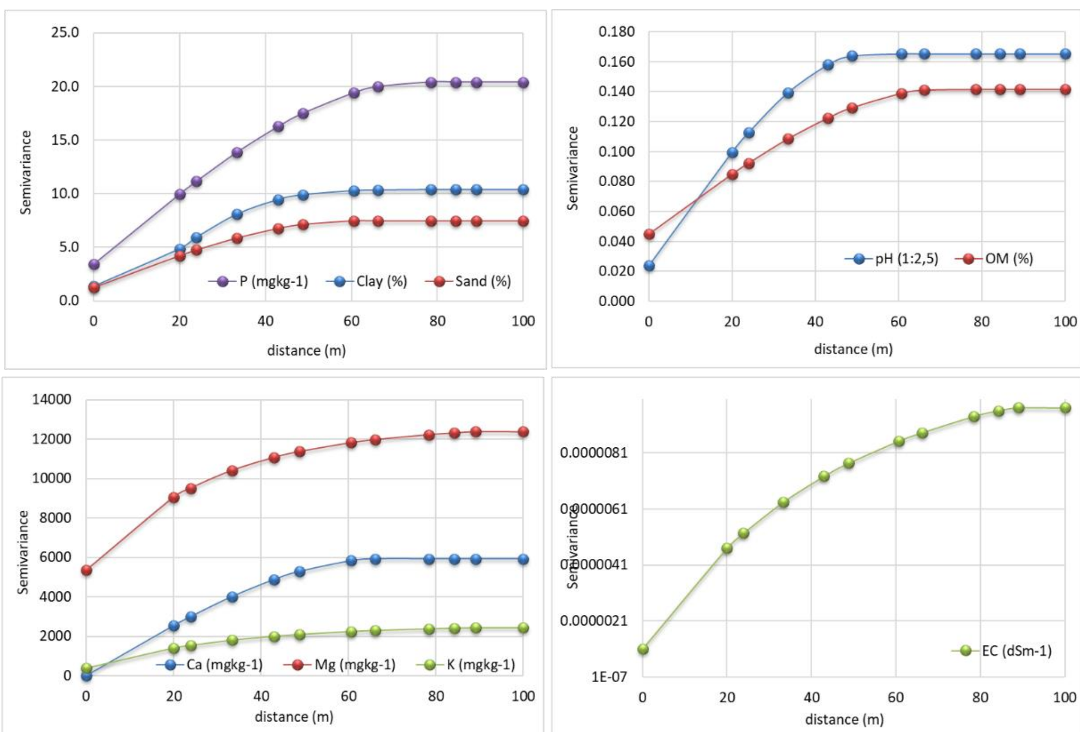

Figure 3. Semivariograms of soil variables in the Experimental Field "El Rastro".

The % silt, as obtained by difference of sand and clay contents, was not used in the subsequent analyses due to the high correlation with these variables, and to avoid obtaining a poorly conditioned matrix, which would interfere in the interpretation of the results. The geostatistical parameters derived from the adjustment of the semi-variograms to different theoretical models are shown in Table 3, and the models for each variable are presented in Figure 4.

Table 3. Geostatistical parameters of the composite semivariograms of soil properties.

Variables	Model	C_0	C_1	A_1	$C + C_1$	RMSE	AIC	RN (%)
pH (1:2.5)	Spherical	0.02	0.1	53.3	0.17	0.01	75	14.4
EC (dS m^{-1})	Spherical	1.1E-06	1.0E-05	45.1	0.00	9.4E-07	248	10.1
P (mg kg^{-1})	Spherical	3.45	17.0	76.3	20.40	0.68	21	16.9
K (mg kg^{-1})	Exponential	387	2.173	31.7	2.560	42.10	104	15.1
Ca (mg kg^{-1})	Spherical	0.00	5.934	68.0	5.934	148	129	0.0
Mg (mg kg^{-1})	Exponential	5.344	7.354	28.5	12.698	570	156	42.1
OM (%)	Spherical	0.04	0.1	70.4	0.14	0.01	66	31.7
Clay (%)	Gaussian	1.37	9.0	28.5	10.4	0.61	19	13.2
Sand (%)	Spherical	1.26	6.2	60.6	7.5	0.37	9	16.9

C_0*:* *Nugget* variance, C_1: Structural variance, $C_0 + C_1$: Threshold or sill, A_1: Range, AIC: Akaike information criterion, RMSE: Root mean square error, RN: Relative *nugget* ((C_1 /C_0 +C_1)*100).

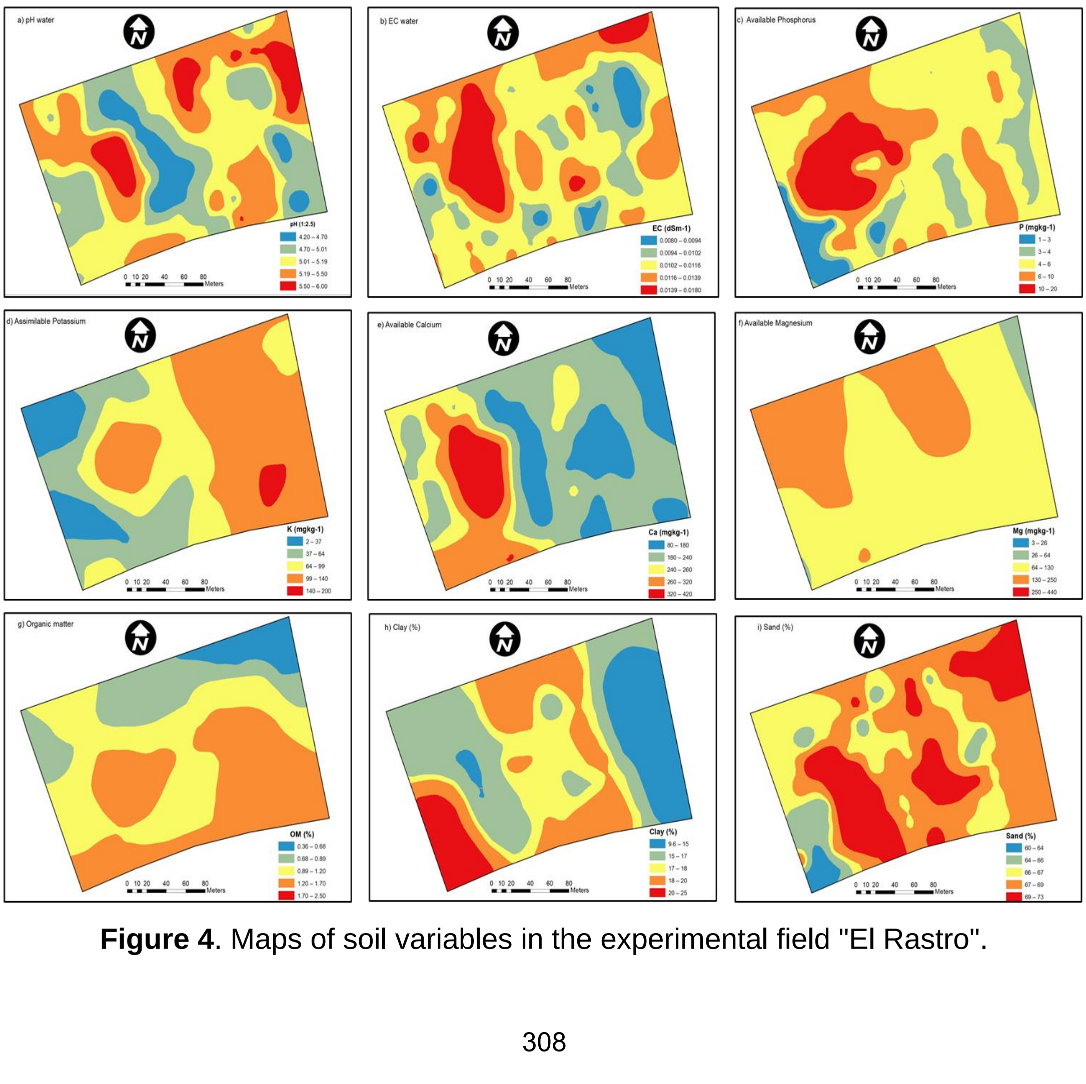

Figure 4. Maps of soil variables in the experimental field "El Rastro".

Assessing the reliability of prediction models

The results of the validations of the soil variables are shown in Table 4, where the low values of the prediction errors, which are very close to zero for the RMSE, ME and MAE indices, can be observed.

Table 4. Prediction error of soil variables by cross-validations.

Variables	Index		
	RMSE	**ME**	**MAE**
pH (1:2.5)	0.33	0.00	0.26
EC (dS m^{-1})	2.25E-03	-1.19E-05	1.84E-03
P (mg kg^{-1})	3.15	0.07	2.23
K (mg kg^{-1})	39.52	0.35	32.34
Ca (mg kg^{-1})	49.40	0.66	40.86
Mg (mg kg^{-1})	102.33	1.63	77.63
OM (%)	0.31	0.00	0.21
Clay (%)	2.335	0.003	1.853
Sand (%)	2.108	0.045	1.693

RMSE: Root means square error; *ME:* Mean error, *MAE*: Mean absolute error. EC: electrical conductivity, P: available phosphorus, K: assimilable potassium, Ca: available calcium, Mg: available magnesium, OM: organic matter.

The greatest uncertainty was found in the variables K, Ca and Mg, which had a greater variance and higher coefficients of variation than the rest of the attributes, and therefore somewhat higher RMSE and MAE values. This last index indicates that there is a slight overestimation in the values of K, Ca and Mg. For all the cases evaluated, the RMSE and MAE values are lower than the *standard deviation*, which means that they can be considered low, and are therefore suitable for the evaluation of prediction models (Marcheti *et al*., 2010).

Generation of the digital soil fertility class model

Soil fertility classes

The representation of the variation of the fuzzy performance index (FPI) as a function of the number of classes for different fuzzy coefficients is shown in Figure 5. The diagram shows that the most suitable number of soil classes was obtained with 7 classes, combined with aϕ of 1.2. The FPI value of 0.30 points to the point of intersection at which there is a minimization of the degree of fuzziness, which determined the optimal number of classes, characterized by being less diffuse and less internally disorganized for the set of variables related to soil fertility.

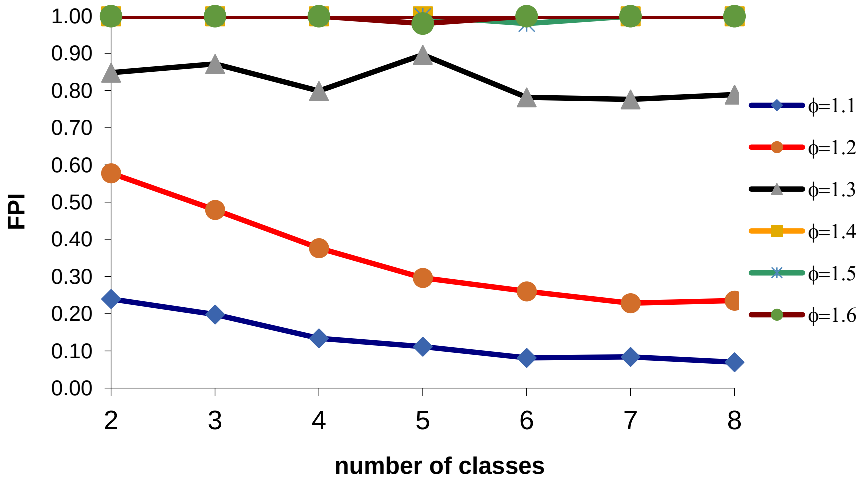

Figure 5. Variation of the fuzzy performance index (FPI) as a function of the number of classes.

The results of the centroid values for each fertility class (centroids) are shown in Table 5. This allowed the following statements to be drawn: **Class 1** groups soils with the lowest potassium and organic matter contents, although with moderately acidic pH. **Class 2** is characterized by soils with strongly acid pH and the highest phosphorus contents in the sector. **Class 3** includes soils with moderately acid reactions. Class 4 corresponds to soils with the highest levels of assimilable potassium and the lowest contents of calcium and clay fractions. **Class 5** involves soils with the highest organic matter content. **Class 6** includes soils characterized by high sand contents. **Class 7** groups soils with a sandy-clay loam texture with clay contents above 20%, the highest amounts of Calcium and the lowest contents of available Magnesium and Phosphorus.

Table 5. Centroids of soil fertility classes obtained with the FCM algorithm.

	Soil Fertility Class						
Soil Variable	**1**	**2**	**3**	**4**	**5**	**6**	**7**
pH water (1:2.5)	5.18	4.91	5.29	4.99	5.17	5.09	5.00
EC water (dS m^{-1})	0.012	0.013	0.012	0.011	0.011	0.011	0.011
P (mg kg^{-1})	6	10	5	5	6	6	4
K (mg kg^{-1})	37	72	102	118	114	71	52
Ca (mg kg^{-1})	246	227	206	188	192	243	265
Mg (mg kg^{-1})	172	127	126	83	111	118	72
OM (%)	0.94	0.94	0.76	1.2	1.35	1.24	1.17
Clay (%)	16.6	17.3	17.2	14	17.3	17.5	20.5
Sand (%)	66.7	67.6	68.1	68.3	67.6	69.2	65.3

EC: electrical conductivity (1:5), OM: organic matter.

The application of the FCM 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 6.

Figure 6. Maps of membership function values for each of the soil fertility classes.

The combination of the spatial distribution models of membership values produced the unified map of soil fertility class variation (Figure 7). To produce this map, the FCM algorithm transformed the fuzzy classes into discrete classes, so that each cell of the model was assigned to the class with the highest membership value. The final model corroborated the distribution of soil fertility classes, where the spatial variation patterns allowed discriminating the dominance of sandy loam soils in the north-eastern sectors, and a higher predominance of clays in the south-western region, which is related to the grain size distribution processes, where the finest particles accumulate in the lower regions of the sector. The final model also made it possible to visualize the expression of the boundaries defined by the dominant classes in the surface layer of the soils. These boundaries facilitate decision-making for the establishment of experimental plots, and enable options for possible explanations related to the study of soil fertility and the establishment of crops for experimental and research purposes in the study area.

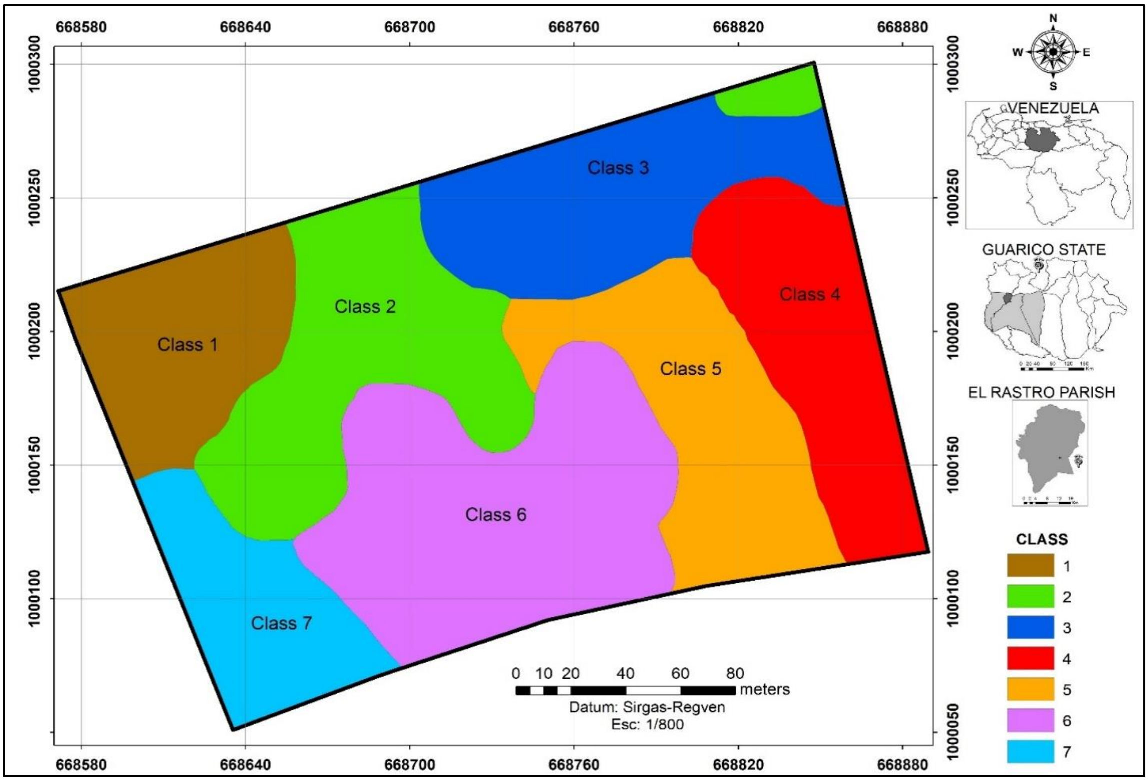

Figure 7. Soil fertility class distribution model of the Experimental Field "El Rastro".

With regard to the surface area of the soil units: class 1 occupies 10.6% of the sector under evaluation, class 2 occupies an area of 17.3%, class 3 represents 14.1 of the area under study, class 4 corresponds to 12.7 of the study area, class 5 corresponds to 13.9 of the experimental area, class 6 covers 22.5% of the experimental field under consideration, and class 7 occupies 9% of the area under study.

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

The analysis of variance for the one-factor classification allowed obtaining the variance of the soil variables by the effect of the fuzzy classes and the complement of the relative variance, which is shown in Table 6. The results indicate that in all the situations described the average intra-class variance (s^2_W) presents lower values than the total variance (s^2_T), which is an indication that the classifications performed are highly meritorious for the variables considered. According to the results of the relative variance complement (1-rv), the proportion of the variance that can be attributed to fuzzy classification is above 87% on average. This indicates that there is a high degree of homogeneity within the soil classes, which ensures that the predictions that can be made from these variables are fairly accurate.

Table 6. Sample mean, total variance, intraclass variance and complement of relative variance for soil properties.

Property	Average	S^2_T	S^2_W	1-rv
pH water (1:2.5)	5.10	0.607	0.124	0.796
EC water (dS m^{-1})	1.2E-02	3.9E-05	5.6E-06	0.859
P (mg kg^{-1})	6.3	165.98	8.79	0.947
K (mg kg^{-1})	86.2	17212	1858	0.892
Ca (mg kg^{-1})	227.5	21569	5246	0.757
Mg (mg kg^{-1})	115.9	43228	9917	0.771
OM (%)	1.13	0.87	0.12	0.868
Clay (%)	17.0	106.5	5.29	0.950
Sand (%)	67.9	45.9	5.38	0.883

S^2_T: Total variance, S^2_W: Intraclass variance, 1-rv: complement of the relative variance

Regarding the assessment of the predictive capacity of the soil classes with multivariate statistics, the results of the classification carried out are shown in Table 7.

Table 7. Results of the size-based classification of diffuse soil fertility classes.

Classes	Ranking (%)[1]	Error (%)
7	85.5	14.5

[1] Correctly classified according to the original grouped cases.

The Mahalanobis distance (D^2) for the original data yielded values above 85%, with an uncertainty of less than 15%. In other words, the validation process of the soil fertility class model indicated that 85.5% of the original cases were correctly classified. The highest degree of uncertainty is given by classes 3 and 6, with errors of 35.7 and 23.1% respectively, but in general, confusions occur between neighboring classes. The results of the validation of the FCM approach showed that it is an alternative for the generation of soil fertility classes. These results are slightly superior to those obtained by Zhu *et al.* (2008) and McKay *et al*. (2010) who applied a soil inference system for soil type prediction at subgroup and soil series level. These investigations expressed a reliability of 76 and 73.7%, respectively, for the data-constrained soil maps.

CONCLUSIONS

The soil property maps showed that there are gradual soil changes with respect to the attributes that showed spatial dependence, variations of which should be considered, as they may affect the reliability of studies, assessments or tests for research purposes.

The experimental plot at El Rastro appears to be homogeneous, according to the influence of soil formation factors, however, the unit is not internally homogeneous. This could have been influenced by the management of agronomic practices. This variability has to be taken into account to avoid a differential effect on the crops.

The evaluation of the digital fuzzy model indicated that the spatial prediction of soil fertility classes corresponds to what is expected in the studied sector, as the reliability was higher than 85%.

The combination of fuzzy set theory and geostatistical techniques provided an alternative that can contribute to improve decision making for the location of experimental plots and carry out local research of great importance, by generating predictions of soil properties and fertility classes with adequate accuracy, capable of capturing the continuous variation of soils in the studied sector.

ACKNOWLEDGEMENTS

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

REFERENCIAS

Beckett, P.H.T. and Burrough. P.A. (1971). The relation between cost and utility in soil survey. IV. Comparison of the utilities of soil maps produced by different survey procedures, and to different scales. J. Soil Sci. 22: 466-480.

Bezdek J.C. (1981). Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York. 256 p.

Bezdek J.C., R. Ehrlich and Full. W. 1984. FCM: the fuzzy c-means clustering algorithm. Computers and Geosciences. 10: 191-203.

Burrough, P. (1986). Principles of geographical information systems land resources assessment. Clarendon Press Oxford. 193 p.

Cambardella, C., T. Moorman, J. Novak, T. Parkin, D. Karlen, R. Turco, and Konopka, E. (1994). Field scale variability of soil properties in Central Iowa Soils. Soil Sci. Soc. Am. J. Vol. 58:1501-1511.

De Gruijter J.J. and McBratney A.B. (1988). A modified fuzzy k-means method for predictive classification. In: Bock H.H. (ed.). Classification and Related Methods of Data Analysis. Elsevier Science Publishers B.V., Amsetrdam.

Di Rienzo J.A., F. Casanoves, M.G. Balzarini, L. González, M. Tablada and Robledo, CW.

(2019). InfoStat versión 2015. Grupo InfoStat, FCA, Universidad Nacional de Córdoba, Argentina. URL http://www.infostat.com.ar.

Grunwald, S., R.L. Rivero and Ramesh, K. (2007). Understanding spatial variability and its application to biogeochemistry analysis. In: D. Sarkar, R. Datta and R. Hannigan (Ed.). Developments in Environmental Science, 5:443-463. Elsevier Ltd. ISSN: 1474-8177.

Marchetti, A.; C. Piccini, R. Francaviglia, S. Santucci and 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. and 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.

Minasny, B., A.B. Mcbratney and Whelan, M. (2002). VESPER. Version 1.6. Australian Centre for Precision Agriculture. McMilan Building. The University of Sidney, NSW 2006.

Núñez, Y. (2011). Modelo automatizado de unidades de paisaje a escala 1:50000 con un enfoque neuronal difuso. En la Cuenca del río Tucutunemo, Estado Aragua. Trabajo de Especialización de Geomática. Postgrado en Ciencia del Suelo. Facultad de Agronomía. UCV. 75 p.

Ovalles, F. (1992). Metodología para determinar la superficie representada por muestras tomadas con fines de fertilidad. FONAIAP-CENIAP-IIAG. Maracay. Serie B. 44 p.

Rey, J.C., G. Martínez, E. Micale, N. Fernández, E. Namias, M.A. Polanco y 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.

Tukey, J. (1977). Exploratory Data Analysis. Addison-Wesley Pub. Reading, EUA.

Upchurch, D. and 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.

Srinivasan, R., Shashikumar, B.N. and 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

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. y Arias, E. (2023). Cartografía digital de clases de fertilidad del suelo con técnicas de redes neuronales difusas. UNERG AGRO-Científica 4(2): 144-163

Valera, A. y 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., 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, and 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, J.A., M.C. Pineda, A. Viloria-Botello, Y. Núñez, y 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., and 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.