

APLICACIÓN DE TÉCNICAS GEOESTADÍSTICAS Y REDES NEURONALES ARTIFICIALES EN LA DELIMITACIÓN DE CLASES DE FERTILIDAD DE SUELOS

Ángel R. Valera Valera¹, Merly Orta Fernández²

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 de Sistemas, San Juan de los Morros, Estado Guárico, Venezuela, e-mail: morta@unerg.edu.ve

RESUMEN

La delimitación de áreas homogéneas con fines de manejo de suelos o para el establecimiento de parcelas experimentales requiere del conocimiento de la variación de las propiedades del suelo y la interpretación basada en la dependencia espacial de las variables más relevantes. Con la finalidad de predecir la variación de propiedades y las clases de fertilidad del suelo se combinaron dos técnicas alternativas de análisis espacial y cartografía digital de suelos. La primera técnica corresponde a la aplicación del análisis geoestadístico para la interpolación de propiedades individuales del suelo de naturaleza química y física. La segunda técnica consistió en la aplicación de tecnologías de inteligencia artificial basada en un sistema de clasificación por redes neuronales artificiales, mediante el algoritmo FKCN (red de agrupamiento borroso de Kohonen o *Fuzzy Kohonen Clustering Network*), para la generación de un modelo digital de clases de fertilidad del suelo. En tal sentido, 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 equivalente al 80%, lo que indicó un alto grado de homogeneidad dentro de las clases de suelo delimitadas.

PALABRAS CLAVE: FERTILIDAD DEL SUELO; GEOESTADÍSTICA; RED NEURONAL ARTIFICIAL; ALGORITMO FKCN.

ABSTRACT

The delimitation of homogeneous areas for soil management purposes or for the establishment of experimental plots requires knowledge of the variation of soil properties and interpretation based on the spatial dependence of the most relevant variables. In order to predict the variation of properties and soil fertility classes, two alternative techniques of spatial analysis and digital soil mapping were combined. The first technique corresponds to the application of geostatistical analysis for the interpolation of individual soil properties of a chemical and physical nature. The second technique consisted in the application of artificial intelligence technologies based on a system of classification by artificial neural networks, by means of the algorithm FKCN (Fuzzy Kohonen Clustering Network), for the generation of a digital model of classes of soil fertility. In this sense, a superficial sampling was carried out in 110 sites in the "El Rastro" Experimental Field, 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 by theoretical semivariograms. 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 carried out through cross validation, to verify the predictive capacity of the variables, an

analysis of variance was applied, and multivariate statistics were used to assess the final model. The digital map of fertility indicated that seven fertility classes predominate in the study area, which presented a reliability equivalent to 80%, which indicated a high degree of homogeneity within the delimited soil classes.

Keywords: SOIL FERTILITY; GEOSTATISTICS; ARTIFICIAL NEURAL NETWORK; ALGORITHM FKC

INTRODUCCIÓN

La fertilidad del suelo es una de las cualidades más importantes del terreno que puede ser afectada en gran medida por su uso y manejo, y es de gran utilidad para el reconocimiento del estado nutricional del suelo y para afinar las recomendaciones de cultivos mediante aplicaciones de nutrientes de fuentes orgánicas e inorgánicas. El conocimiento de la variación espacial de la fertilidad del suelo en los sitios experimentales y como parte de una agricultura de precisión es importante para la definición del establecimiento de parcelas homogéneas, y evitar solapamientos entre los tratamientos y ensayos. La representación manual de las clases de fertilidad del suelo amerita la elaboración de mapas individuales de cada una de las variables, y la posterior superposición de éstos para la definición de límites, lo que implica sesgos y baja precisión en el resultado final. La representación cartográfica de la fertilidad de suelo, facilita la toma de decisiones al momento del establecimiento de parcelas experimentales y ensayos con fines de investigación.

Convencionalmente, la evaluación de la fertilidad se realiza a través del análisis de los resultados de análisis de suelos y las áreas de estudio o experimentales se consideran homogéneas; es decir, no se toma en cuenta la variación espacial de los atributos. Con el tiempo se han implementado algunas técnicas cartográficas, que permiten obtener información básica y confiable de la expresión espacial de las propiedades del suelo. Dentro de estas técnicas, los métodos geoestadísticos juegan un papel importante para la predicción espacial de propiedades del suelo, donde destaca el método de interpolación denominado *kriging* ordinario. Sin embargo, la representación individual de las variables que definen la fertilidad de los suelos no cubre el interés y la necesidad de visualizar el comportamiento de manera conjunta, en un modelo de variación espacial de clases de fertilidad de suelo.

La cartografía digital de suelos (CDS) permite integrar diversos modelos de variación espacial de propiedades individuales del suelo para obtener clases de suelo, con la finalidad de apoyar la toma de decisiones sobre la definición de áreas como base para el manejo de sitio específico y para el impulso de la agricultura de precisión. La aplicación de la CDS a través de la evaluación de la variación espacial de la fertilidad del suelo intenta dividir el continuo suelo en clases, que exhiban una mayor homogeneidad de la combinación de la influencia de las variables consideradas en el análisis de suelos.

En la actualidad existen diversas opciones para apoyar la cartografía digital de ambientes complejos y dinámicos. Dentro de estas se encuentran las técnicas de análisis espacial basadas en inteligencia artificial, tales como la lógica difusa y las redes neuronales artificiales (RNA), las cuales son fundamentales para la clasificación digital

de unidades de suelos (Ehsani y Quiel, 2009). Estas técnicas se adaptan adecuadamente al estudio de los atributos del suelo, los cuales varían gradualmente sobre el espacio, donde la representación de esa variación gradual puede resultar en la obtención de información útil y en la disminución de errores en la definición de límites apropiados de unidades edafológicas (Burrough et al. 2000). De esta manera, las RNA intentan construir un modelo matemático que consta de un conjunto de funciones sencillas vinculadas por sus pesos, mientras que los métodos de clasificación difusa tienen la capacidad de modelar la variación continua del suelo de acuerdo con los valores de una función de pertenencia. La combinación de la potencialidad de los conjuntos difusos y las RNA ha desarrollado una técnica integral de clasificación no supervisada denominada red de agrupamiento difuso de Kohonen o *Fuzzy Kohonen Clustering Network* (FKCN) (Lin y Lee, 1996; Bezdeck et al., 1992), que conjuga un algoritmo de mapas autoorganizados o *self organizing map* (SOM) (Kohonen, 1982) y el algoritmo *Fuzzy C-means* (FCM) (Bezdeck, 1981).

Son escasos los trabajos de investigación realizados en el ámbito de la Ciencia del Suelo que toman en cuenta la combinación de la variación continua de propiedades individuales para expresarlas como categorías de fertilidad de suelo. Al respecto, la aplicación de las redes neuronales - difusas - como parte de las tecnologías de inteligencia artificial - ha dado un gran impulso a la CDS tanto en la predicción de propiedades como en la obtención de clases de suelo. En Venezuela, las tecnologías de inteligencia artificial se han aplicado en el área de clasificación de paisajes y predicción de atributos del suelo (Viloria, 2007), en cartografía digital geomorfológica (Valera y Viloria, 2009), Valera et al. (2010), Núñez (2011), Viloria et al (2012), Valera (2012), Sevilla (2014) y Viloria et al (2016), en la predicción de propiedades y clases locales de suelo (Valera, 2015; Valera, 2018) y en el estudio de las relaciones suelo y rendimiento del cultivo de banano (Rey et al., 2015).

En este trabajo se presenta un estudio de CDS para la delimitación de clases de fertilidad, a través de la predicción de propiedades químicas y físicas del suelo obtenidas en los análisis de laboratorio y su posterior agrupamiento. Para evaluar el comportamiento espacial de las clases de fertilidad de los suelos, se consideró el Campo Experimental "El Rastro" de la Universidad Nacional Experimental de los Llanos Centrales "Rómulo Gallegos", ubicado en la carretera nacional vía Guardatinajas sector El Rastro, parroquia El Rastro del municipio autónomo Francisco de Miranda, estado Guárico (Venezuela). El propósito fundamental de la investigación fue la predicción espacial de clases de fertilidad del suelo a través de redes neuronales artificiales y las técnicas geoestadísticas, como base para la generación de información básica requerida para el desarrollo de ensayos y pruebas experimentales, que permitan una visión espacial del estado de la fertilidad, y una mejor interpretación de los resultados de los diferentes tratamientos, ensayos agronómicos, manejo de suelos, así como futuras investigaciones de campo y evaluaciones con fines experimentales a desarrollar en el sector estudiado.

MATERIALES Y MÉTODOS

Área de Estudio

La zona de estudio donde se llevó a cabo el ensayo de cartografía digital del suelo se encuentra ubicada en terrenos del Campo Experimental “El Rastro” de la Universidad Nacional Experimental Rómulo Gallegos, ubicada en el Sector El Rastro, carretera nacional vía Guardatinajas, parroquia El Rastro del municipio Francisco de Miranda, Estado Guárico, Venezuela (Figura 1). La unidad de estudio se encuentra enmarcada en una planicie subreciente, con pendiente de 2 a 4%. Los suelos en esta área fueron formados de materiales geológicos del cuaternario, con un incipiente desarrollo pedogenético, y son de baja a moderada fertilidad.

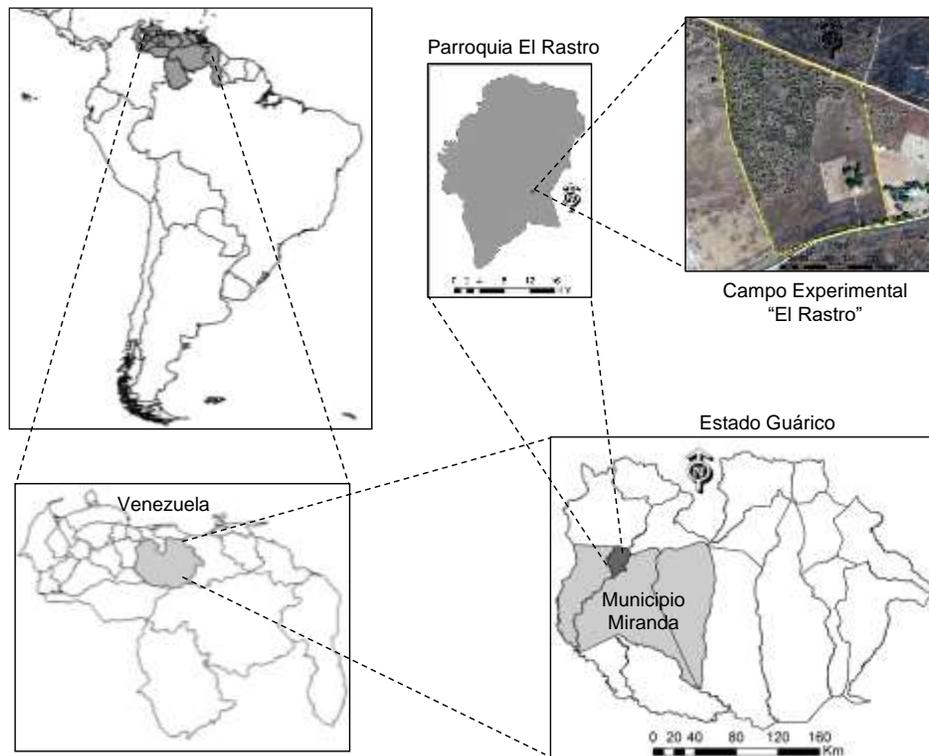


Figura 1. Ubicación relativa del Campo Experimental “El Rastro” en el municipio Miranda, estado Guárico, Venezuela.

Muestreo de suelos

Para la evaluación de los suelos se realizó un muestreo sistemático en el horizonte superficial a 20 cm de profundidad, en cuadrículas espaciadas a 20 m, para un total de 110 muestras de suelos en una superficie de 4,85 ha (Figura 2). Cada punto de muestreo fue georreferenciado con apoyo de un sistema de posicionamiento global (GPS). A las muestras superficiales se les realizó un diagnóstico con fines de fertilidad, utilizando las metodologías del Laboratorio de análisis de suelos del Centro de Investigación en Suelos y Aguas de la Universidad Rómulo Gallegos (CIESA-UNERG). Se analizaron diez variables de suelo: pH en agua (1:2,5), conductividad eléctrica en

agua 1:5 (CE, dSm^{-1}), materia orgánica (MO, %), fósforo disponible (P, mgkg^{-1}), potasio asimilable (K, mgkg^{-1}), calcio (Ca, mgkg^{-1}) y magnesio disponible (Mg, mgkg^{-1}), y las cantidades relativas de arena, limo y arcilla (%).

Análisis estadístico

Los datos de las variables edáficas se sometieron a un análisis exploratorio (AED) con apoyo del programa InfoStat (Di Rienzo *et al.*, 2015), con el propósito de determinar los estadísticos descriptivos, tales como: media, mediana, varianza, coeficiente de variación, valores máximos y mínimos, y los índices de asimetría y curtosis. Se utilizó la metodología de cercas externas e internas de Tukey (1977), con el fin de detectar la presencia de valores atípicos. Adicionalmente se realizó la prueba de normalidad de Kolmogorov-Smirnov, para evaluar la distribución de los datos.



Figura 2. Distribución de los sitios de muestreo de suelos en el Campo Experimental “El Rastro”.

Interpolación de las propiedades del suelo

Para la interpolación de las propiedades del suelo se utilizó el método geoestadístico *kriging* ordinario, el cual utiliza un modelo de semivariograma para la obtención de los pesos que se asignan a cada punto de referencia empleado en la estimación del valor de las variables regionalizadas que presentan dependencia espacial. El semivariograma es definido por la función de la semivarianza [$\gamma(h)$], la cual se estima con la siguiente expresión (Upchurch y Edmonds, 1991; Ovalles, 1992):

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

donde N es el número de pares de puntos separados por una determinada distancia h ; $z(x_i)$ es el valor de la variable en una localidad x ; $z(x_{i+h})$ es el valor que toma la variable en otra localidad ubicada a una distancia h de x (Ovalles y Rey, 1994). El semivariograma contiene la información referente a la variable regionalizada, cuyos parámetros son: la varianza *nugget* (C_0), la varianza estructural (C_1), el umbral (C_0+C_1) y el alcance (A_1), que indica la distancia dentro de la cual existe dependencia espacial [Burrough, 1986; Grunwald *et al.*, 2007]. La estimación de los semivariogramas empíricos de las propiedades del suelo, y el ajuste a modelos matemáticos se realizó con el programa *Vesper 1.6* (Minasny *et al.*, 2002). Con los parámetros ajustados se obtuvieron los estimados óptimos de las variables edáficas en los sitios no muestreados, a través de la interpolación con el método *kriging* ordinario (Webster y Oliver, 1990). Los modelos de las variables del suelo fueron generados con el total de los datos y la exactitud de los mapas se verificó mediante validaciones cruzadas. En la evaluación se utilizaron tres índices: el error cuadrático medio (ECM), el error medio (EM) y el error medio absoluto (EMA). El ECM evalúa la precisión de la predicción y mide la cantidad de error que hay entre dos conjuntos de datos, es decir, compara un valor predicho y un valor observado o conocido; el EM evalúa el error sistemático e indica la presencia de subestimación o sobrestimación del modelo, y el EMA garantiza que el resultado del error sea estrictamente positivo.

Modelo digital de clases de fertilidad del suelo

Para la obtención de los modelos representativos de las clases de fertilidad del suelo se utilizó el algoritmo neuro-difuso FKCN implementado en ambiente Java (Windows) por Vilorio (2007). La arquitectura de la red neuronal FKCN utilizada en el análisis consta de tres capas (Figura 3). La capa de entrada contiene los valores normalizados de nueve (9) variables edáficas procedentes de los modelos de predicción de variables del suelo. La capa de distancias incluye las neuronas equivalentes al número preestablecido de clases digitales de suelo y la tercera capa calcula la función de pertenencia de cada celda a cada una de las clases de suelo, basada en las distancias calculadas en la capa anterior y los valores preestablecidos del coeficiente difuso (ϕ). En la capa de distancia se calcula la separación d_{ij} existente entre un patrón de entrada X_j y el peso del nodo w_i , con $i = 1, 2, \dots, c$, donde c representa el número de clases del modelo a estimar. Posteriormente, la capa de membresía grafica las distancias d_{ij} en valores de membresía U_{ij} , donde U_{ij} representa el grado de pertenencia de un patrón de entrada X_j a una clase c . Se debe tener en cuenta que a lo largo del proceso de aprendizaje ocurre una retroalimentación de la capa de funciones de pertenencia a la capa de distancia, con la finalidad de ajustar los centros de clase. Las variables del suelo fueron agrupadas en una matriz de datos para la aplicación del sistema FKCN, lo que permitió evaluar el agrupamiento de píxeles con diversos números de clases (2 a 8) y diferentes coeficientes difusos ($\phi = 1,1$ a $1,6$). A las clases de fertilidad obtenidas se les asignó el significado edafológico a través de la interpretación de su distribución espacial, las descripciones de los centroides de clases, y las matrices de los valores de similitud

obtenidos mediante el algoritmo FKN, conjuntamente con la información de los análisis de los suelos de la zona.

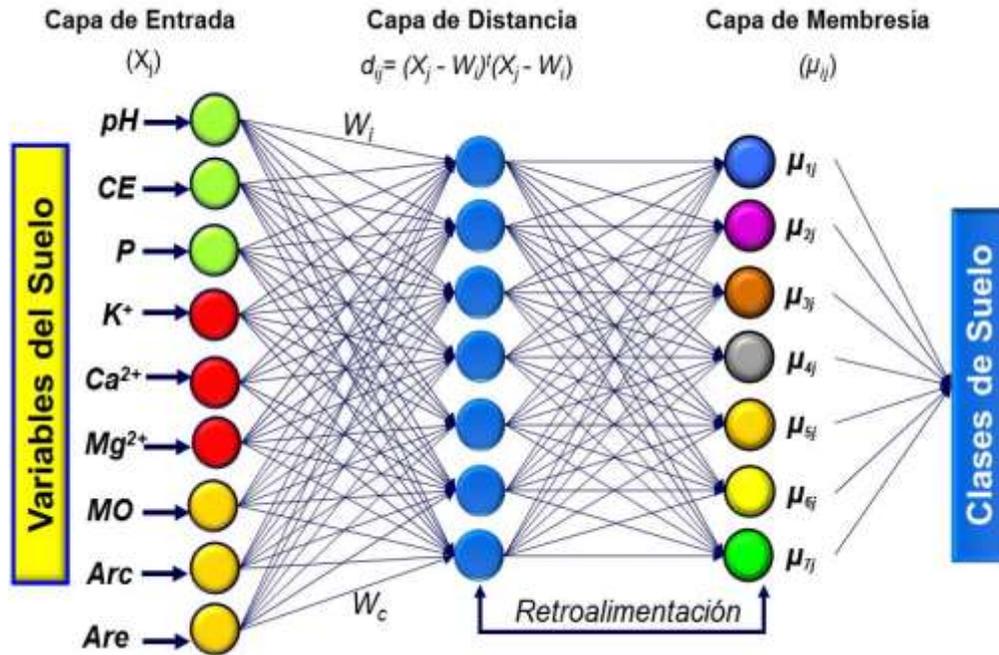


Figura 3. Estructura de la red neuronal difusa empleada en la predicción de clases de suelo.

Número de clases de fertilidad del suelo

Para la obtención del mejor modelo de clases difusas se empleó un enfoque inductivo, basado en el procedimiento de Odeh *et al.* (1992), el cual relaciona el índice de rendimiento difuso (FPI, *Fuzziness Performance Index*) y la entropía de partición modificada (MPE, *modified partition entropy*) con el número de clases. Estos parámetros se obtienen utilizando el algoritmo *Fuzzy c-Means* (FCM) (Bezdek 1981; Bezdek *et al.* 1984) del programa *Fuzme 3.5* (Minasny y McBratney 2002). La selección del número óptimo de clases en FCM se realizó por repetición de la clasificación para un rango de número de clases. En cada agrupamiento obtenido se generan dos parámetros de clasificación, tales como el FPI y la entropía de partición modificada (MPE, *modified partition entropy*). El FPI estima el grado de borrosidad generada por cada número específico de clases. Matemáticamente, se define como:

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

donde c es el número de clases y F es el coeficiente de partición calculado como:

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

F es conceptualmente comparable a la relación del conjunto de varianzas dentro de las clases y la varianza entre clases y es cercana a 1 para los agrupamientos más significativos. En el presente estudio, el agrupamiento de mapas de propiedades del

suelo en formato *raster* se realizó estableciendo previamente los siguientes parámetros: a) número de clases ($c= 6$ a 12), b) exponente difuso $\phi= 1,1$ a $1,6$ con incrementos de $0,1$; c) un máximo de 300 iteraciones, y d) criterio de detención ($\epsilon= 0,0001$). En los cálculos se utilizó la distancia métrica de *Mahalanobis*, la cual toma en consideración la correlación encontrada entre algunas variables del suelo de la zona estudiada.

Evaluación de la capacidad predictiva de las clases de fertilidad del suelo

Para evaluar la capacidad predictiva de las clases obtenidas mediante el agrupamiento difuso, se realizó un análisis de varianza de un factor, mediante el complemento de la varianza relativa ($1-r_v$) (Beckett y Burrough, 1971), a fin de verificar el efecto de las propiedades edáficas en la diferenciación de las clases de fertilidad de suelos en el sector estudiado. Este índice es análogo al coeficiente de determinación y expresa la proporción de la varianza que puede ser atribuida a la clasificación. Bajo este criterio, para que una clasificación sea meritoria, la varianza intraclases promedio debería ser menor que la varianza total (Webster y Oliver, 1990). Finalmente se realizó la validación del modelo final con los casos originales agrupados, empleando la *Distancia* de *Mahalanobis* (d^2) como estadístico descriptivo multivariado, derivado del análisis discriminante canónico.

RESULTADOS Y DISCUSIÓN

Análisis estadístico

Los estadísticos descriptivos indicaron que los valores promedios de los suelos corresponden a grupos texturales franco-arenosos, con reacciones que oscilan de moderada a fuertemente ácida, con bajos contenidos de fósforo y moderados a altos contenidos de Potasio, alta disponibilidad de Calcio y Magnesio, bajos a medianos contenidos de materia orgánica y no presentan problemas de salinidad (Cuadro 1).

Cuadro 1. Estadísticos descriptivos de las variables de fertilidad del suelo del Campo experimental.

Variable ¹	Mínimo	Máximo	Media	Mediana	Curtosis	Asimetría	DE	Varianza	CV (%)
pH (1:2,5)	4,2	6,0	5,1	5,1	-0,28	0,08	0,4	0,1	7,4
CE (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 ²⁺ (mg kg ⁻¹)	80	420	227,5	220,0	0,55	0,80	76,5	5.855,8	33,6
Mg ²⁺ (mg kg ⁻¹)	3,0	441,0	115,9	77,5	1,37	1,31	105,9	11.204,6	91,3
MO (%)	0,4	2,5	1,1	1,2	1,48	0,55	0,4	0,2	34,4
Arcilla (%)	9,6	24,9	17,0	17,3	-0,30	-0,01	3,3	10,6	19,1
Arena (%)	59,8	73,1	67,9	67,8	0,12	0,08	2,7	7,3	4,0
Limo (%)	9,3	23,3	15,1	14,6	-0,11	0,57	2,9	8,4	19,1

¹Número de datos: 110, DE: Desviación estándar, CV: Coeficiente de variación, CE: Conductividad eléctrica, P: Fósforo disponible, K: Potasio asimilable, Ca: Calcio disponible, Mg: Magnesio disponible, MO: Materia orgánica.

La mayoría de las variables presentan cierta similitud entre la media y la mediana, a excepción de las variables P, K, Ca y Mg. A su vez, la mayor dispersión de los datos la presentan las mismas variables, debido a la expresión de la desviación estándar y la varianza, sin embargo los coeficientes de variación de las variables en conjunto, no presentan problemas en cuanto a la existencia de valores extremos de los datos.

De acuerdo con el coeficiente de sesgo o asimetría las variables pH, K, % arcilla y % arena, cumplen con la función de distribución de probabilidad normal, se pueden aplicar métodos geoestadísticos a los datos. Sin embargo, para el %MO y Ca es necesario realizar una transformación de datos (normalización) de tipo raíz cuadrada; y para P, Mg y CE, es necesario hacer una transformación de tipo logarítmico para la posterior aplicación de algún método geoestadístico a los datos.

La aplicación de la prueba de cercas externas e internas indicó que las variables consideradas no presentan valores atípicos. En cuanto a la prueba de normalidad, las variables K y MO, proceden de poblaciones normales, ya que los valores de la prueba estadística son altamente significativos ($p > 0,05$) (Cuadro 2). Sin embargo, para el resto de los datos fue necesario realizar la transformación de datos.

Cuadro 2. Prueba de normalidad del conjunto de datos de suelo del Campo Experimental “El Rastro”.

Variable	Estadístico	gl	Sig. ¹
pH (1:2,5)	0,095	110	0,016
CE (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 ²⁺ (mg kg ⁻¹)	0,126	110	0,000
Mg ²⁺ (mg kg ⁻¹)	0,158	110	0,000
MO (%)	0,070	110	0,200
Arcilla (%)	0,100	110	0,009
Arena (%)	0,139	110	0,000
Limo (%)	0,155	110	0,000

¹Nivel significación $\alpha = 0,05$; gl: Grados de libertad; n= 110.

Interpolación de propiedades del suelo

La estimación de los semivariogramas empíricos de las variables del suelo, fueron ajustados a modelos matemáticos gaussianos, esféricos y exponenciales, respectivamente (Figura 4), considerando el comportamiento isotrópico de las variables. Debido a que el contenido de limo se obtiene por diferencia de las cantidades de arena y arcilla existen altas correlaciones entre las variables granulométricas, por lo que dicho atributo no se utilizó en los análisis posteriores, a fin de evitar la obtención de una matriz mal condicionada que pudiera interferir en la interpretación de los resultados. Los parámetros geoestadísticos derivados del ajuste de los semivariogramas a diferentes modelos teóricos, se expresan en el Cuadro 3, y los modelos de cada variable se presentan en la Figura 5.

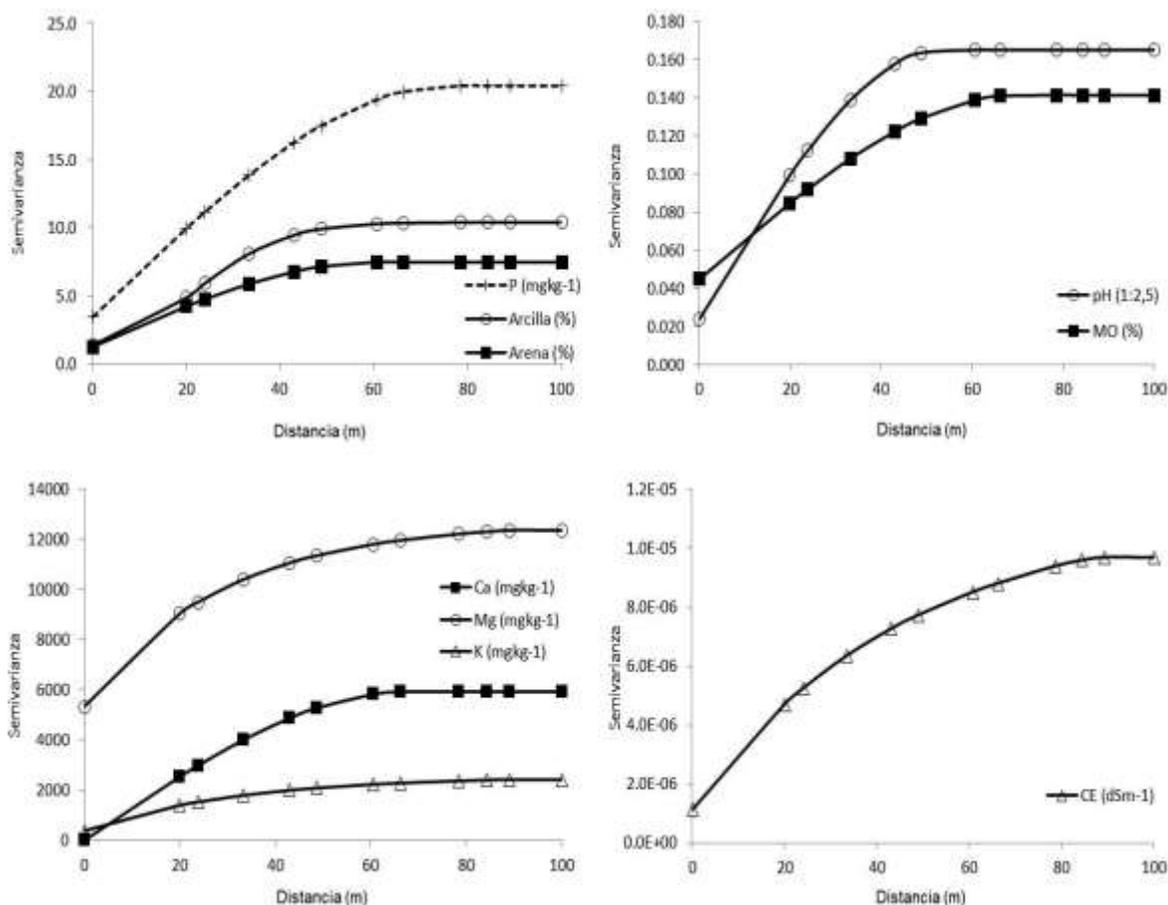


Figura 4. Semivariogramas de las variables de suelo en el Campo Experimental “El Rastro”.

Cuadro 3. Parámetros geoestadísticos de los semivariogramas compuestos de las propiedades del suelo.

Variables	Modelo	C ₀	C ₁	A ₁	C ₀ +C ₁	ERCM	CIA	NR (%)
pH (1:2,5)	Esférico	0,02	0,1	53,3	0,17	0,01	75	14,4
CE (dS m ⁻¹)	Esférico	1,1E-06	1,0E-05	45,1	0,00	9,4E-07	248	10,1
P (mg kg ⁻¹)	Esférico	3,45	17,0	76,3	20,40	0,68	21	16,9
K ⁺ (mg kg ⁻¹)	Exponencial	387	2.173	31,7	2.560	42,10	104	15,1
Ca ²⁺ (mg kg ⁻¹)	Esférico	0,00	5.934	68,0	5.934	148	129	0,0
Mg ²⁺ (mg kg ⁻¹)	Exponencial	5.344	7.354	28,5	12.698	570	156	42,1
MO (%)	Esférico	0,04	0,1	70,4	0,14	0,01	66	31,7
Arcilla (%)	Gaussiano	1,37	9,0	28,5	10,4	0,61	19	13,2
Arena (%)	Esférico	1,26	6,2	60,6	7,5	0,37	9	16,9

C₀: Varianza *nugget*, C₁: Varianza estructural, C₀ + C₁: Umbral, A₁: Alcance, CIA: Criterio de información de Akaike, ERCM: error de la raíz cuadrada de la media, NR: *Nugget* relativo ((C₀/C₀+C₁)*100).

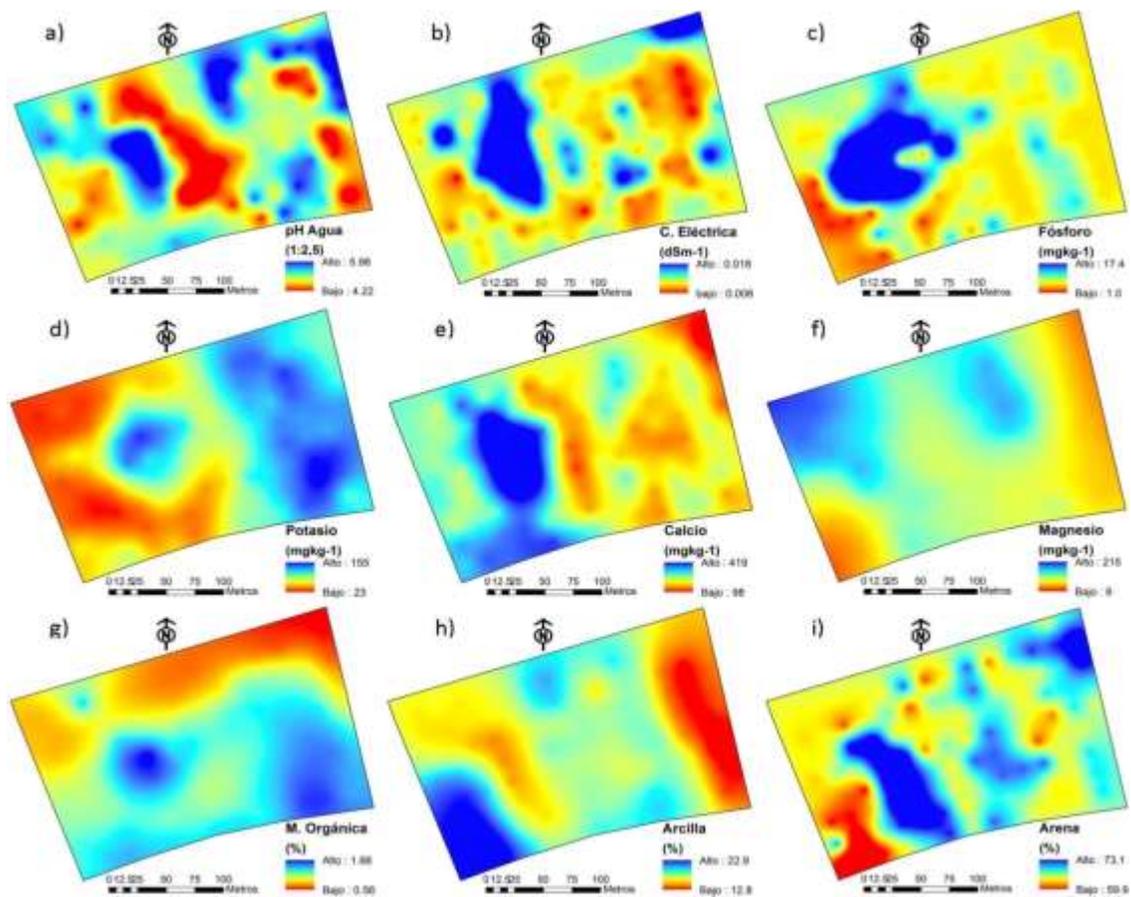


Figura 5. Mapas de modelos de variables del suelo en el Campo experimental "El Rastro"

Evaluación de la confiabilidad de los modelos de predicción

Los resultados de las validaciones de las variables del suelo se indican en el Cuadro 4, donde se observan los bajos valores de los errores de predicción, los cuales son muy cercanos a cero para los índices ECM, EM y EMA.

Cuadro 4. Error de predicción de las variables del suelo mediante validaciones cruzadas.

Variables	Índice		
	ECM	EM	EMA
pH agua (1:2,5)	0,33	0,00	0,26
CE agua (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 ⁺² (mg kg ⁻¹)	49,40	0,66	40,86
Mg ⁺² (mg kg ⁻¹)	102,33	1,63	77,63
MO (%)	0,31	0,00	0,21
Arcilla (%)	2,335	0,003	1,853
Arena (%)	2,108	0,045	1,693

ECM: error cuadrático medio; EM: error medio, EMA: Error medio absoluto. CE: Conductividad eléctrica, P: Fósforo disponible, K: Potasio asimilable, Ca: Calcio disponible, Mg: Magnesio disponible, MO: Materia orgánica.

La mayor incertidumbre la presentaron las variables K, Ca y Mg, las cuales presentaron una mayor varianza y coeficientes de variación más altos que el resto de los atributos y por tanto valores de ECM y EMA algo elevados. Este último índice indica que existe una ligera sobreestimación en los valores de K, Ca y Mg. Para todos los casos evaluados los valores de ECM y EMA son inferiores a la *desviación estándar*, lo que significa que pueden ser considerados bajos y por tanto son adecuados para la evaluación de modelos de predicción (Marcheti *et al.*, 2010).

Generación del modelo digital de clases de fertilidad del suelo

Número de clases de fertilidad del suelo

La representación de la variación del índice de rendimiento difuso (FPI) en función del número de clases para diferentes coeficientes difusos, se señala en la Figura 6.

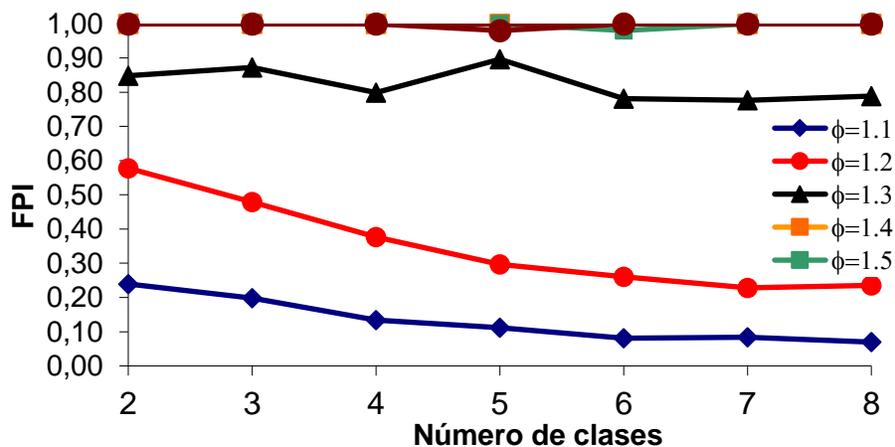


Figura 6. Variación del índice de rendimiento difuso (FPI) en función del número de clases de suelo.

En el esquema se visualiza que el número de clases de suelo más adecuado se obtuvo con 7 clases, combinado con un ϕ de 1.2. El valor del FPI de 0,30 señala el punto de intersección en la cual existe una minimización del grado de borrosidad, que determinó el número óptimo de clases, caracterizadas por ser menos difusas y menos desorganizadas internamente para el conjunto de variables relacionadas con la fertilidad del suelo.

Los resultados de los valores de los centros de cada clase de fertilidad (centroides) se indican en el Cuadro 5. Esto permitió extraer los siguientes argumentos: La clase 1 incluye suelos caracterizados por los elevados contenidos de arena, aunque también poseen los pH más elevados, y por consiguiente los mayores valores de Calcio y Fósforo disponible. La clase 2 incluye los suelos con reacciones moderadamente ácidas y con valores promedios en la mayoría de los componentes químicos que la caracterizan. La clase 3 agrupa los suelos de textura franco arcillo-arenosa con contenidos de arcilla superiores al 20%, además presentan los contenidos más bajos de Magnesio y de Fósforo disponible. La clase 4 agrupa suelos con los contenidos más elevados de Magnesio y los más bajos niveles de Potasio asimilable, lo cual implica una relación con desbalance nutricional. La clase 5 involucra a los suelos con menor

contenido de materia orgánica y los más bajos contenidos de Calcio disponible. La clase **6** se caracteriza por agrupar los suelos con los pH más ácidos del sector (reacción fuertemente ácida), y a clase **7** corresponde a los suelos con los mayores niveles de Potasio asimilable y materia orgánica, y los menores contenidos de arcilla.

Cuadro 5. Centroides de las clases de fertilidad del suelo obtenidas con el algoritmo FKCN.

Variable del Suelo	Clase de Fertilidad del Suelo						
	1	2	3	4	5	6	7
pH agua (1:2,5)	5.40	5.11	4.99	5.17	5.38	4.91	4.98
CE agua (dS m ⁻¹)	0.016	0.011	0.011	0.012	0.012	0.012	0.011
P (mg kg ⁻¹)	14	5	4	6	5	7	5
K ⁺ (mg kg ⁻¹)	100	89	50	35	107	65	123
Ca ⁺² (mg kg ⁻¹)	360	219	267	248	174	212	188
Mg ⁺² (mg kg ⁻¹)	134	119	71	175	99	129	91
MO (%)	1.29	1.28	1.19	0.94	0.68	0.88	1.31
Arcilla (%)	15.7	17.7	20.7	16.5	16.1	17.9	14.7
Arena (%)	70.5	68.5	65.2	66.6	69.0	67.1	67.7

CE: Conductividad eléctrica, P: Fósforo disponible, K: Potasio asimilable, Ca: Calcio disponible, Mg: Magnesio disponible, MO: Materia orgánica.

La aplicación del algoritmo FKCN también generó los valores de grados de pertenencia de cada celda (píxel) a cada una de las clases de fertilidad del suelo. La clasificación produjo vectores de valores de membresía para cada celda del modelo correspondiente a cada clase de fertilidad. Estos valores fueron representados espacialmente produciendo mapas individuales de los miembros de las clases, las cuales reflejan la variación espacial de los grados de pertenencia entre 0 (colores oscuros) y 1 (colores claros), a través de mapas en formato *raster* expresados en la Figura 7.

La combinación de los modelos de distribución espacial de los valores de membresía produjo el mapa unificado de la variación de clases de fertilidad de los suelos (Figura 8). Para producir este mapa el algoritmo FKCN transformó las clases neurodifusas a unidades discretas, por lo que cada celda del modelo fue asignada a la clase con el valor más alto de membresía. El modelo final corroboró la distribución de las clases de fertilidad de los suelos, donde los patrones de variación espacial permitieron discriminar la dominancia de suelos franco-arenosos en los sectores Nor-este, y una mayor predominancia de arcillas en la región Sur-oeste, lo cual está relacionado con los procesos de distribución granulométrica, donde las partículas más finas se acumulan en las regiones más bajas del sector. El modelo final también permitió visualizar la expresión de los límites definidos por las clases dominantes en la capa superficial de los suelos. Estos límites facilitan la toma de decisiones para el manejo de suelos y para el establecimiento de parcelas experimentales, y posibilitan las opciones para las posibles explicaciones relacionadas con el estudio de la fertilidad de los suelos y el establecimiento de cultivos con fines experimentales y de investigación en el área de estudio.

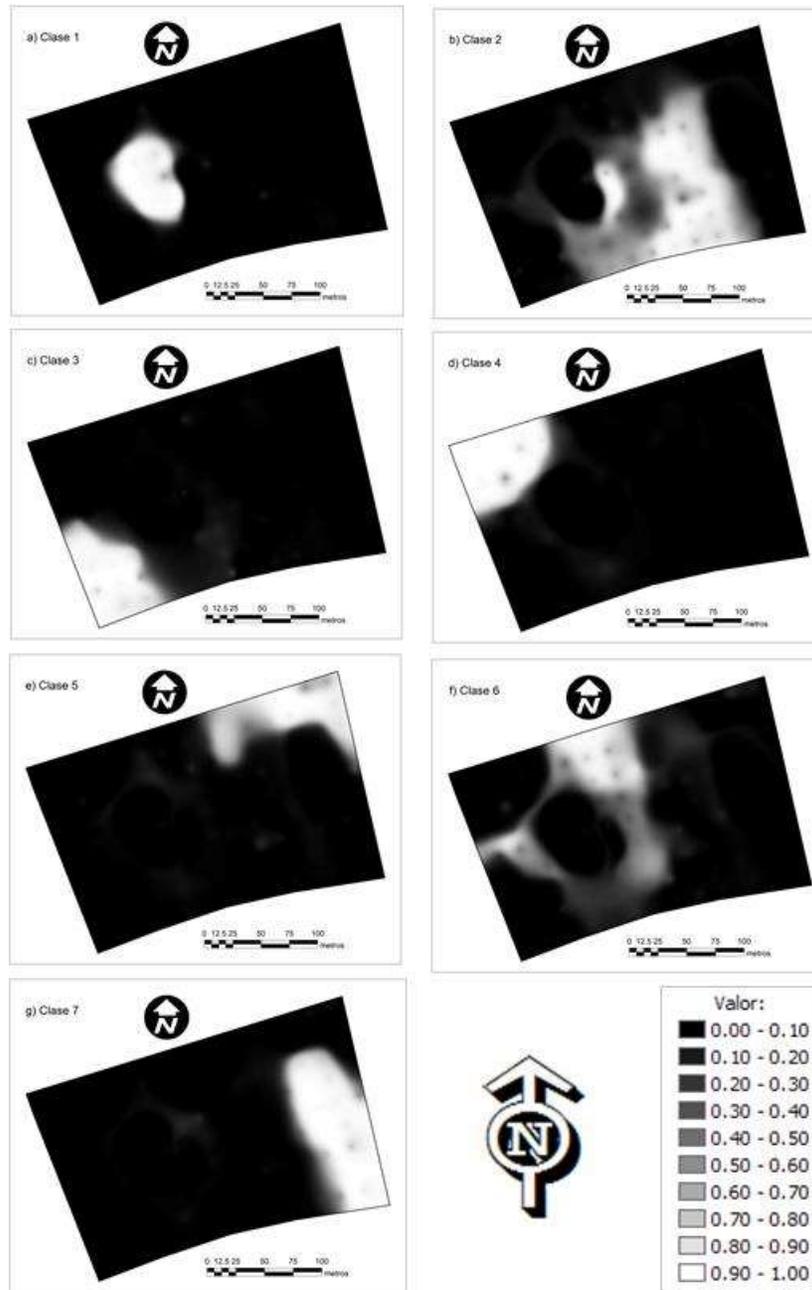


Figura 7. Mapas de valores de la función de pertenencia a cada una de las clases de fertilidad del suelo.

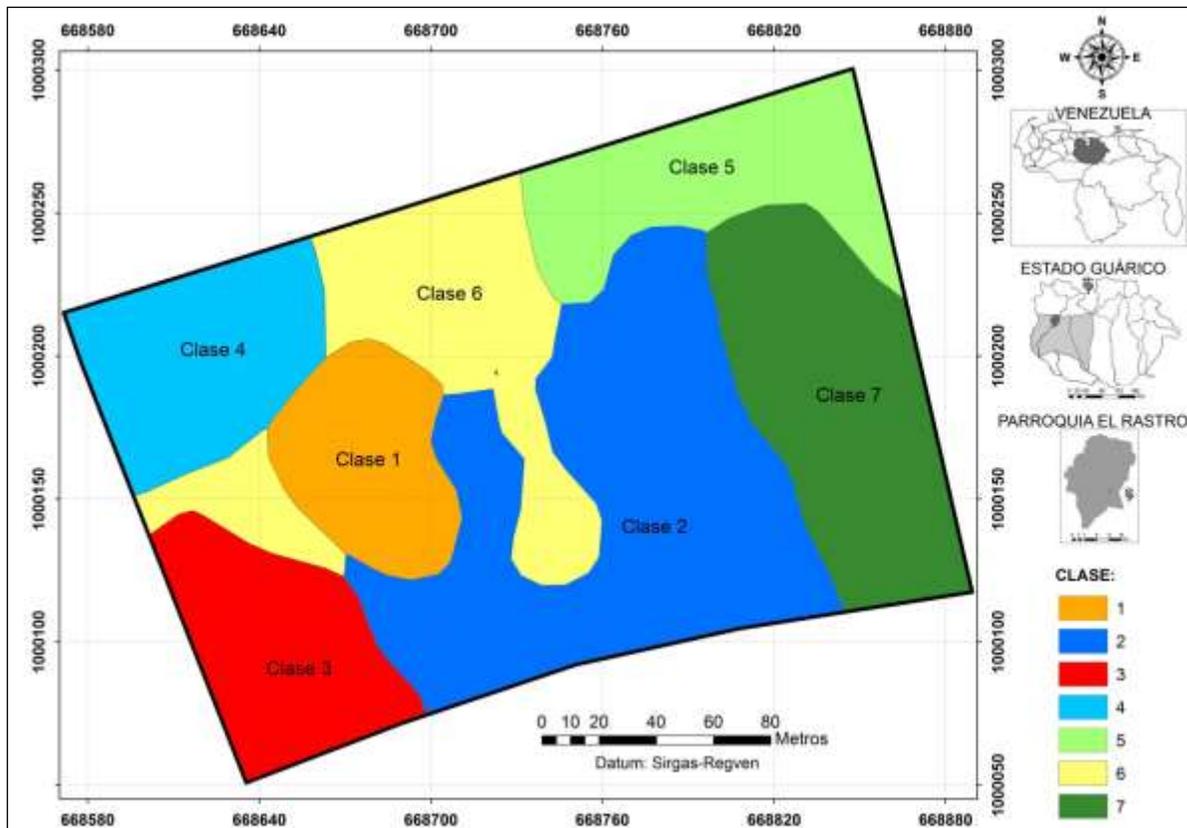


Figura 8. Modelo de distribución de clases de fertilidad del suelo del Campo Experimental "El Rastro".

En cuanto a la superficie de las unidades de suelo: la clase 1 ocupa el 8,1% del sector evaluado, la clase 2 ocupa una superficie de 28,6%, la clase 3 representan el 10,1 de la zona estudiada, la clase 4 corresponde al 10,6 de la zona de estudio, la clase 5 se corresponde con el 11,2% del área experimental, la clase 6 abarca el 15,3% del campo experimental en consideración, y la clase 7 ocupa el 16,1% de la zona estudiada.

Evaluación de la capacidad predictiva del modelo digital de clases de fertilidad del suelo

El análisis de varianza para la clasificación de un factor permitió la obtención de la varianza de las variables del suelo por efecto de las clases difusas y el complemento de la varianza relativa, lo cual se observa en el Cuadro 6. Los resultados indican que en todas las situaciones descritas la varianza intraclases (S^2_w) promedio presenta valores menores que la varianza total (S^2_T), lo cual es un indicativo de que las clasificaciones realizadas son altamente meritorias para las variables consideradas. De acuerdo con los resultados del complemento de la varianza relativa ($1-VR$), la proporción de la varianza que puede ser atribuida a la clasificación difusa está por encima del 87% en promedio. Esto indica que existe un alto grado de homogeneidad dentro de las clases de suelo, lo que garantiza que las predicciones que puedan ser realizadas a partir de éstas variables sean bastante precisas.

Cuadro 6. Promedio de la muestra, varianza total, varianza intraclase y complemento de la varianza relativa para las propiedades del suelo.

Propiedad	Promedio	S ² _T	S ² _w	1-VR
pH agua (1:2,5)	5,10	0,896	0,106	0,884
CE agua (dS m ⁻¹)	1,2E-02	5,8E-05	4,4E-06	0,933
P (mg kg ⁻¹)	6,3	158,7	9,23	0,777
K ⁺ (mg kg ⁻¹)	86,2	16.399	1.908	0,927
Ca ⁺² (mg kg ⁻¹)	227,5	51.212	3.413	0,962
Mg ⁺² (mg kg ⁻¹)	115,9	44.181	9.858	0,940
MO (%)	1,13	1,253	0,091	0,944
Arcilla (%)	17,0	118,5	4,54	0,884
Arena (%)	67,9	67,2	4,06	0,933

S²_T: Varianza total, S²_w: varianza intraclase, 1-VR: Complemento de la varianza relativa

En cuanto a la valoración de la capacidad predictiva de las clases de suelo con estadísticos multivariados, los resultados de la clasificación realizada se señalan en el Cuadro 7.

Cuadro 7, Resultados de la clasificación basada en los tamaños de las clases de fertilidad de suelo neurodifusas.

Clases	Clasificación (%) ¹	Error (%)
7	80,1	19,9

¹ Clasificados correctamente según los casos agrupados originales

La distancia de Mahalanobis (D²) para los datos originales, arrojó valores superiores al 80%, con una incertidumbre inferior al 20%. Es decir, el proceso de validación del modelo de clases de fertilidad del suelo indicó que el 80,1% de los casos originales fueron clasificados correctamente. El mayor grado de incertidumbre está dado por las clases 2 y 6, con errores de 20% cada una, cuyas confusiones ocurren debido a que son clases vecinas, tal como se visualiza en el modelo final (Figura 8). Los resultados de la validación del enfoque FKCN demostraron que es una alternativa para la generación de clases de fertilidad del suelo. Estos resultados son ligeramente superiores a los obtenidos por Zhu *et al.* (2008) y McKay *et al.* (2010) quienes aplicaron un sistema de inferencia de suelos para la predicción de tipos de suelo a nivel de subgrupos y serie de suelos. Estas investigaciones expresaron una confiabilidad de 76 y 73,7% respectivamente, para los mapas de suelo obtenidos con limitaciones de datos.

CONCLUSIONES

Los mapas de las variables analizadas demostraron que existen cambios graduales del suelo con respecto a los atributos que mostraron dependencia espacial, cuyas variaciones deben ser consideradas porque pueden afectar la confiabilidad de los estudios, evaluaciones o ensayos con fines de investigación.

El área evaluada no es internamente homogénea, debido posiblemente a la influencia

del manejo de suelo y las prácticas agronómicas realizadas en la zona. Esta variabilidad hay que tenerla presente para evitar un efecto diferencial en los cultivos.

La evaluación del modelo digital neurodifuso indicó que la predicción espacial de clases de fertilidad de suelo se corresponde con lo esperado en el sector estudiado, ya que la confiabilidad fue equivalente al 80%.

La combinación de las redes neuronales difusas y las técnicas geoestadísticas proporcionaron una alternativa que puede contribuir a mejorar la toma de decisiones para la ubicación de parcelas experimentales y realizar investigaciones locales de gran importancia, al generar predicciones de propiedades y clases de fertilidad del suelo con adecuada exactitud, capaz de capturar la variación continua de los suelos en el sector estudiado.

AGRADECIMIENTOS

Esta investigación fue apoyada por el Centro de Investigación y Extensión en Suelos y Aguas de la Universidad Nacional Experimental de los Llanos Centrales "Rómulo Gallegos" (CIESA-UNERG).

REFERENCIAS

- Beckett, P.H.T., 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, 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.
- Cambardella, C., T. Moorman, J. Novak, T. Parkin, D. Karlen, R. Turco, 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., 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., Amsterdam.
- Di Rienzo J.A., F. Casanoves, M.G. Balzarini, L. González, M. Tablada, Robledo, CW. 2015. *InfoStat versión 2015*. Grupo InfoStat, FCA, Universidad Nacional de Córdoba, Argentina. URL <http://www.infostat.com.ar>.
- Ehsani, A.H., Quiel, F. 2009. A semi-automatic method for analysis of landscape elements using shuttle radar topography mission and landsat ETM+ data. *Computers & Geosciences* 35: 373-389.

- Grunwald, S., R.L. Rivero, 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.
- 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. Hartemink). 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.
- Minasny, B., A.B. Mcbratney, 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, 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 Río Canoabo, Estado Carabobo. Trabajo de Ascenso para optar a la Categoría de Asistente. Universidad Central de Venezuela. Caracas, Venezuela. 117 p.
- 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. 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. Viloría; Viloría, 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.
- Valera, A., Viloría, 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.
- Viloría J.A, A. Viloría-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.
- Viloría, 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.
- Viloría, J.A., M.C. Pineda, A. Viloría-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.