

---

## PREDICCIÓN ESPACIAL DEL CARBONO ORGÁNICO EN SUELOS AGRÍCOLAS DE LA REGIÓN ORIENTAL DEL ESTADO GUÁRICO, VENEZUELA

Ángel Rafael Valera Valera<sup>1,2\*</sup>, Cándido Sumoza Agraz<sup>1,2</sup>, Williams Tovar Romero<sup>1,2</sup>

<sup>1</sup> Área de Ingeniería Agronómica, Universidad Rómulo Gallegos, San Juan de los Morros, Estado Guárico, Venezuela, <https://orcid.org/0000-0001-5500-1332e>, e-mail: [angelvalera@unerg.edu.ve](mailto:angelvalera@unerg.edu.ve)

<sup>2</sup> Centro de Investigación y Extensión en Suelos y Aguas (CIESA-UNERG), San Juan de los Morros, Estado Guárico, Venezuela, [sumagraz@hotmail.es](mailto:sumagraz@hotmail.es); [wuilltov@gmail.com](mailto:wuilltov@gmail.com)

\* Autor de correspondencia

**Recibido:** 14 - 05 - 2022; **Aceptado:** 02 - 06 - 2023; **Publicado:** 30 - 06 - 2023

---

### RESUMEN

La materia orgánica del suelo juega un papel importante en los sistemas agrícolas, y, es uno de los principales indicadores de la calidad, salud y productividad del suelo. La información sobre la distribución de la materia orgánica es clave no solo para el manejo de la fertilidad del suelo sino también para estimar la reserva de carbono orgánico del suelo (COS). Además, el conocimiento de la distribución espacial de las reservas de carbono orgánico contribuye al manejo sostenible del uso de la tierra y al funcionamiento de los ecosistemas, y es muy importante como indicador de la calidad del suelo. Con la finalidad de conocer la variación espacial del carbono orgánico en suelos agrícolas, en este trabajo se utilizaron 2.179 muestras de suelo compuestas georreferenciadas, derivadas de muestreos realizados a 20 cm de profundidad, en distintas unidades de producción en la región Oriental del estado Guárico, Venezuela. Para la estimación de la reserva de COS, en este estudio se utilizó la cartografía digital del suelo (DSM) para predecir y evaluar la distribución espacial de COS utilizando métodos geoestadísticos avanzados como Kriging Bayesiano Empírico (KBE) y regresión kriging Bayesiano Empírico (RKBE), y el algoritmo de aprendizaje automático denominado bosques aleatorios (RF, *random forest*). En el proceso de interpolación por el método KEB se empleó la totalidad de los datos de carbono orgánico, con RKEB se utilizaron los datos de suelo y variables relacionadas con los factores formadores del suelo. Los resultados indicaron que los tres métodos son capaces de estimar la distribución espacial de la reserva de COS, con variaciones que oscilaron entre 7 y 62 t ha<sup>-1</sup> para el método de interpolación KBE, entre 2 y 67 t ha<sup>-1</sup> para RKBE, y entre 14 a 56 t ha<sup>-1</sup> para RF. El método RF mostró un menor error cuadrático medio (ECM) y el más bajo error estándar promedio (EEP). Además, la generación del modelo RF arrojó un R<sup>2</sup>=0,95 y la validación con el 10% de los datos arrojó un coeficiente de acuerdo de 0,83 lo cual corresponde con una alta consistencia del modelo. Puede concluirse que RF proporciona una mejor estimación y variabilidad espacial del COS; sin embargo, una mayor selección y elección de variables auxiliares y una mejor selección de los puntos de muestreo del suelo podrían mejorar la precisión de la predicción de RF.

**Palabras clave:** Cartografía digital, Suelos, Geoestadística, Carbono orgánico, Bosques aleatorios, Kriging Bayesiano Empírico

---

### SPATIAL PREDICTION OF ORGANIC CARBON IN AGRICULTURAL SOILS IN THE EASTERN REGION OF GUÁRICO STATE, VENEZUELA

#### ABSTRACT

Soil organic matter plays an important role in agricultural systems, and is one of the main indicators of

soil quality, health and productivity. Information on the distribution of organic matter is key not only for soil fertility management but also for estimating the soil organic carbon stock (SOC). Furthermore, knowledge of the spatial distribution of organic carbon stocks contributes to sustainable land use management and ecosystem functioning, and is very important as an indicator of soil quality. In order to know the spatial variation of organic carbon in agricultural soils, 2,179 geo-referenced composite soil samples derived from sampling at 20 cm depth in different production units in the Eastern region of Guárico state, Venezuela, were used in this work. For the estimation of the SOC stock, digital soil mapping (DSM) was used in this study to predict and evaluate the spatial distribution of SOC using advanced geostatistical methods such as Empirical Bayesian Kriging (EBK) and Empirical Bayesian Kriging regression (EBKR), and the machine learning algorithm called random forest (RF). The interpolation process for the EBK method used the whole organic carbon data, with EBKR the soil data and variables related to soil-forming factors were used. The results indicated that all three methods are able to estimate the spatial distribution of the COS stock, with variations ranging from 7 to 62 t ha<sup>-1</sup> for the EBK interpolation method, from 2 to 67 t ha<sup>-1</sup> for EBKR, and from 14 to 56 t ha<sup>-1</sup> for RF. The RF method showed the lowest mean square error (MSE) and the lowest average standard error (ASE). Furthermore, the RF model generation yielded an R<sup>2</sup>=0.95 and the validation with 10% of the data yielded an agreement coefficient of 0.83 which corresponds to a high model consistency. It can be concluded that RF provides a better estimation and spatial variability of COS; however, a better selection and choice of auxiliary variables and a better selection of soil sampling points could improve the prediction accuracy of RF.

**Keywords:** Digital mapping, Soils, Fertility, Geoestatics, Organic carbon, Kriging Empirical Bayesian

---

## INTRODUCCIÓN

La materia orgánica del suelo juega un papel importante en los sistemas agrícolas, y, es uno de los principales indicadores de la calidad y productividad del suelo. La información sobre la distribución de la materia orgánica es fundamental tanto para el manejo de la fertilidad del suelo, como para estimar la reserva de carbono orgánico del suelo. Por otra parte, el aumento de la materia orgánica puede mitigar el cambio climático mediante el almacenamiento de carbono en suelo, lo cual influye tanto en el secuestro de COS como en la salud del suelo. El COS es el mayor reservorio de carbono terrestre (entre el 50 y el 80 % del carbono terrestre total) y un medio importante para mitigar el cambio climático (IPCC, 2016), determina la capacidad de fuente o sumidero de carbono de un paisaje e influye en las propiedades físicas, químicas y biológicas del suelo (Odebiri *et al.*, 2020), y además es la base principal de diversos estudios e investigaciones relacionadas con agricultura, ecología, cambio climático y desarrollo sostenible. En general, los suelos representan una importante reserva de carbono, y la estimación de la existencia de COS es muy relevante, si se tiene en cuenta que el COS es el factor central en la fertilidad del suelo y la mitigación del cambio climático. Es decir, el CO es un indicador esencial de la calidad del suelo y determina directamente la fertilidad del suelo y la productividad de las plantas, ya que juega un papel importante en el suministro de nutrientes al suelo y en la formación de una mejor estructura del suelo (Farooq *et al.*, 2022).

El estudio de la reserva de COS es muy importante porque la variación en el rendimiento de los cultivos dentro de un campo determinado es reflejada por la variación del COS, ya que, para lograr un manejo adecuado de los nutrientes del suelo para la obtención de un rendimiento uniforme de los cultivos, es necesario conocer la cantidad de los nutrientes del suelo dentro de un campo determinado, y la cantidad de carbono orgánico que se encuentra presente (Bashir *et al.*, 2021). Es

decir, un mejor monitoreo del carbono del suelo en las tierras agrícolas, es útil para el secuestro de carbono y para el manejo sostenible del suelo. Sin embargo, la perturbación antropogénica severa en las tierras de cultivo, principalmente en terrenos de poca pendiente, crea incertidumbre en la obtención de información precisa del suelo con datos limitados de muestras de suelo. En los agroecosistemas, la distribución espacial de las propiedades del suelo se ve afectada por procesos ecológicos naturales influenciados por muchos factores, incluidos el clima, el tipo de suelo, la topografía y el uso de la tierra (Liu *et al.*, 2015). Por lo tanto, se convierte en un desafío modelar el COS con precisión a escala de parcelas, y sobre un área más amplia que se extienda por varios kilómetros sin tener en cuenta estos factores.

Antes del surgimiento de las tecnologías geoespaciales, la distribución espacial de las propiedades del suelo, incluido el COS, se evaluaba a partir de estudios de suelo convencionales y análisis de laboratorio de muestras de suelo recolectadas utilizando estadísticas clásicas; un enfoque que es tedioso, lento y costoso. El método tradicional de levantamiento de suelos no podía proporcionar información detallada sobre la variación del suelo requerida para muchas aplicaciones ambientales, ya que se requieren enfoques alternativos. Como alternativa, se desarrolló la técnica de cartografía digital de suelos (DSM, *digital soil mapping*) y se convirtió en un enfoque de la ciencia del suelo. Bajo el marco de la DSM, se han desarrollado varios métodos de predicción para estimar la distribución espacial de las propiedades del suelo, y en especial la distribución de COS. La inferencia espacial de propiedades del suelo es un procedimiento o conjunto de procedimientos para la aplicación de un modelo de relaciones suelo-paisaje (e.g. scorpan) para predecir propiedades o clases de suelo, a partir de variables auxiliares y puntos de muestreo de suelos (Dobos *et al.*, 2006). Esta es esencialmente la importancia del mapeo cuantitativo de suelos, lo que permite la obtención de información precisa y actualizada de un sector de interés, a fin de garantizar la aplicación de un manejo de sitio específico de nutrientes para adaptarse a las condiciones variables espacialmente. Así que, para el manejo de los suelos agrícolas de manera eficiente y mantener su productividad, es esencial lograr estimaciones precisas del contenido de materia orgánica y de su variación espacio-temporal.

A través de los avances tecnológicos, existe una aplicación integral de métodos para el estudio de los suelos, mediante la utilización de datos ambientales auxiliares derivados del procesamiento de imágenes de satélite y modelos digitales de elevación. Entre estos se encuentran los métodos geoestadísticos, y los algoritmos de aprendizaje automático, como la regresión lineal múltiple, las redes neuronales artificiales, la máquina de vectores de soporte, los árboles de decisión, la regresión cubista y los bosques aleatorios. Los datos ambientales auxiliares, como los modelos digitales de elevación (MDE), la teledetección, los datos climáticos y la geología, se han combinado a través de modelos predictivos para estimar las propiedades del suelo. Se ha utilizado una gran cantidad de conjuntos de datos de MDE existentes para extraer atributos del terreno (e.g. altitud, pendiente, orientación del terreno, índice topográfico de humedad, etc.) como auxiliares para la predicción de diversas propiedades del suelo. Los datos de teledetección o sensores remotos, por otro lado, también han servido como fuente de datos importantes para el estudio tanto cualitativo como cuantitativo de las propiedades del suelo, incluido el COS, ya que proporcionan un enfoque preciso, reproducible y espontáneo para cuantificar la

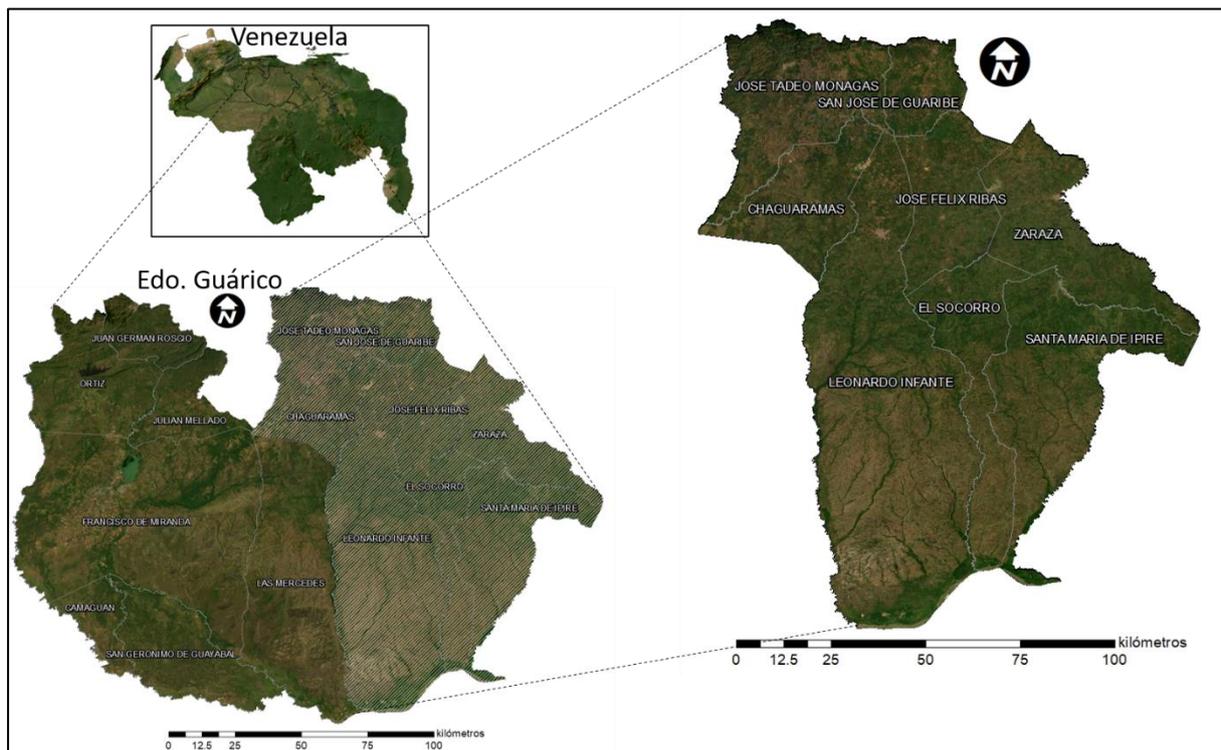
variabilidad del COS.

En esta investigación se utilizaron métodos geoestadísticos avanzados y una técnica de aprendizaje automático para la estimación del carbono orgánico en suelos agrícolas de la región oriental del estado Guárico (Venezuela), con apoyo de datos de laboratorio. Además, se utilizaron los datos auxiliares obtenidos de modelos digitales de elevación e imágenes de satélite, más la información geológica y climática de la zona en consideración.

## MATERIALES Y MÉTODOS

### Área de Estudio

El área de estudio corresponde a la parte oriental del estado Guárico, cuyos datos provienen de siete (7) municipios, donde destacan: Ribas, Infante, Zaraza, Santa María de Ipire, El Socorro, Chaguaramas y Monagas (Figura 1). La superficie de estos municipios abarca un total de 29.419 km<sup>2</sup>, equivalente al 44,3% del territorio del estado Guárico. La zona de interés se encuentra ubicada entre las coordenadas Este: 1.086.027 m - 1.299.509 y Norte: 850.807 m - 1.111.148 m (Proyección Cónica conforme de Lambert, Datum: WGS 1984).



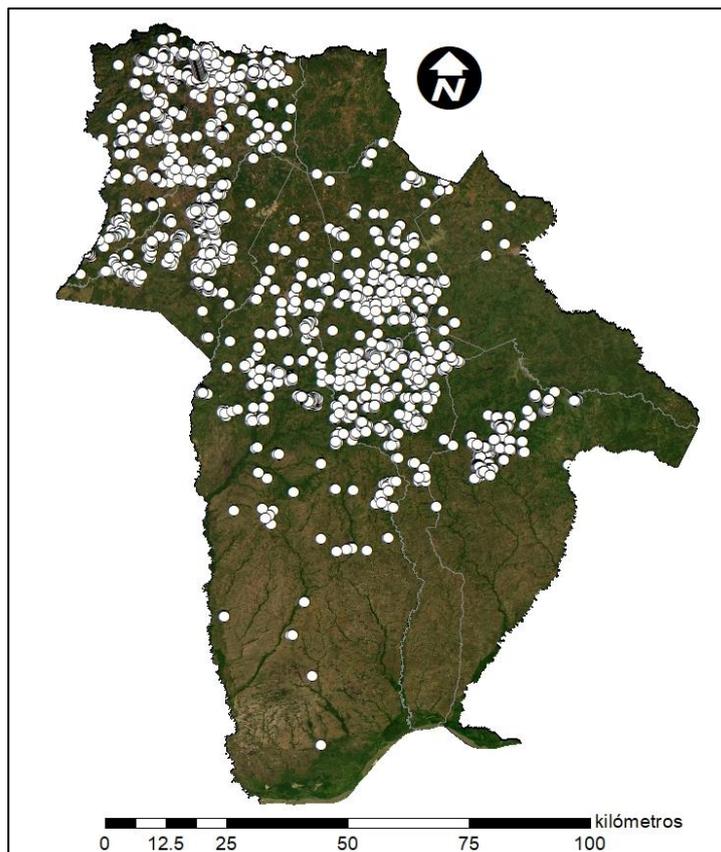
**Figura 1.** Ubicación relativa del Área de estudio en la región oriental del estado Guárico, Venezuela.

La zona corresponde a una zona de vida de bosque seco tropical, con precipitaciones promedio que oscilan entre 900 y 1.450 mm anuales (Fick y Hijmans, 2017), y dominancia de formaciones geológicas del del Oligoceno a Mioceno (Formación Quebradón, Quiamare, Uchirito, Capaya, La Pica, Chaguaramas),

Pleistoceno (Formación Mesa) y del Pleistoceno al Holoceno (zonas aluviales). El paisaje predominante de la zona corresponde a colinas dominadas por relieves de los llanos altos centrales ondulados y disectados, con presencia de suelos de los órdenes entisoles, alfisoles, ultisoles y oxisoles (MARN, 1983).

### Datos de suelo

En este estudio se utilizó una cantidad de 2.279 muestras de suelo compuestas, derivadas del ingreso al laboratorio de servicio del Centro de investigación y Extensión de Suelos y Aguas de la Universidad Rómulo Gallegos (CIESA-UNERG). Las muestras de suelo provienen de zonas agrícolas bajo la siembra principal de maíz y sorgo, a una profundidad de muestreo de 20 cm en promedio, y corresponden a datos acumulados desde el año 2000 hasta la fecha actual (Figura 2). La determinación del carbono orgánico de los suelos se realizó utilizando el método de Walkley-Black (1934) basado en combustión húmeda, con medición por espectrofotometría de luz visible a 650 nm. Los datos de suelo fueron georreferenciados con apoyo de las coordenadas UTM del Censo Agrícola (INE, 2007), complementados con hojas cartográficas a escala 1:100.000 y 1:25.000 del Instituto Geográfico Venezolano Simón Bolívar (IGVSB, 2004), y la plataforma Google Earth Pro®.



**Figura 2.** Distribución de los sitios de muestreo de los suelos de la región Oriental del estado Guárico.

## Variables auxiliares

Como variables ambientales auxiliares se utilizaron diversos mapas en formato ráster de 92 m de resolución espacial, tales como: Modelo digital de elevación (SRTM, *Shuttle Radar Topographic Mission*), geología (USGD, U.S. *Geological Survey*), precipitación media anual (PM Anual) (Wordclim 2.1, 1970-2000) (Fick y Hijmans, 2017); índice de vegetación de diferencia normalizada (NDVI, *normalized difference vegetation index*) (plataforma MODIS, *Moderate Resolution Imaging Spectroradiometer*), cobertura y uso de la tierra (sensor Sentinel 2A). Los atributos geomorfológicos incluyeron: altitud, pendiente, índice topográfico de humedad (ITH), orientación del terreno, área de captación, posición relativa, Factor LS (índice de transporte de sedimentos), nivel base de la red de drenaje (CNBL, *Channel Network Base Level*), distancia vertical a la red del canal o fluvial (VDCN, *Vertical Distance Channel Network*), los cuales fueron calculados con el programa SAGA v. 2.1.4. (Conrad *et al.*, 2015) a partir de la base de datos digitales de elevación SRTM.

## Reserva de carbono orgánico del suelo

La reserva de COS expresada en t ha<sup>-1</sup> se calculó utilizando la fórmula empleada por Penman *et al.* (2003):

$$\text{COS (t ha}^{-1}\text{)} = \text{CO} \times \text{Da} \times p \times (1 - \%eg) \times 10 \quad (1)$$

donde **CO** es el contenido de CO en g kg<sup>-1</sup>, **Da** es la densidad aparente del suelo en t m<sup>-3</sup>, **p** es el espesor o profundidad del suelo en metros (0,20 m) en estos casos, y **eg** es la corrección por presencia de material grueso en los suelos (> 2 mm). Para la densidad aparente del suelo se consideraron los valores medios de las clases de textura dominante (Fa, F, FAa, A, AL) con énfasis en los contenidos de arena basado en funciones de pedotransferencia (Tranter *et al.*, 2007), y un contenido de esqueleto grueso menor al 5% por tratarse de muestras derivadas de suelos agrícolas, donde las áreas pedregosas son descartadas para el muestreo de suelos y el establecimiento de cultivos.

## Métodos de cartografía digital

En este estudio se utilizó el modelo SCORPAN como base de la cartografía digital del suelo (McBratney *et al.*, 2003) que establece las relaciones entre los atributos o clases del suelo y los factores de formación espacialmente referenciados [incluyendo otros datos de suelo (**s**), clima (**c**), organismos (**o**), relieve (**r**), material parental (**p**), tiempo (**a**) y localización espacial (**n**)]. Para la predicción y evaluación de la distribución espacial de COS se emplearon métodos geoestadísticos avanzados como Kriging Bayesiano Empírico (KBE) y regresión kriging Bayesiano Empírico (RKBE), y el algoritmo de aprendizaje automático denominado bosques aleatorios (RF, *random forest*). Como predictores se utilizaron variables auxiliares (índices espectrales y parámetros topográficos) derivadas de modelos de elevación digital e imágenes de satélite, con la finalidad de evaluar la distribución espacial y la variabilidad del COS, y comparar los modelos en función de su rendimiento

predictivo. Para la estimación de COS con RKBE se utilizaron las variables geología, cobertura y uso de la tierra, precipitación media anual, variables auxiliares derivadas de modelos digitales de elevación e índices de imágenes de satélite (NDVI, índice de vegetación de diferencia normalizada), y para la estimación con RF se utilizaron las variables de mayor importancia, tales como geología, cobertura y uso de la tierra, NDVI, precipitación media anual y altitud.

### ***Kriging Bayesiano Empírico (KBE)***

El Kriging bayesiano empírico (KBE) es un método de interpolación de estadísticas espaciales que automatiza los aspectos más complejos del modelo kriging ordinario, y calcula automáticamente los parámetros del semivariograma por medio de un proceso de creación de subconjuntos y simulaciones. El Kriging bayesiano empírico se diferencia además de otros métodos kriging en que tiene en cuenta el error introducido al estimar el semivariograma subyacente (Krivoruchko, 2012). Otros métodos kriging calculan el semivariograma a partir de ubicaciones de datos conocidas y utilizan este único semivariograma para realizar predicciones en ubicaciones desconocidas. En este proceso se asume implícitamente que el semivariograma estimado es el verdadero semivariograma para la región de interpolación. La ventaja del método consiste en tomar en cuenta la incertidumbre de la estimación del semivariograma, para estimar adecuadamente los errores estándar de la predicción (Krivoruchko y Gribov, 2019).

### ***Regresión Kriging Bayesiano Empírico (RKEB)***

El método de predicción RKEB es un método de interpolación geoestadística avanzada que utiliza el kriging empírico bayesiano (KBE) con mapas ráster de variables explicativas que se conoce que afectan el valor de los datos de interpolación. RKBE se diferencia de los métodos de kriging tradicionales porque toma el error causado teniendo en cuenta la estimación del modelo del semivariograma (Krivoruchko, 2012). Esto es llevado a cabo para obtener nuevos modelos de semivariograma cada vez a partir de los datos simulados del semivariograma anterior, lo que convierte a KBE en un interpolador automático fiable. Este enfoque combina kriging con el análisis de regresión para hacer predicciones que son más precisas que las que pueden lograr la regresión o el kriging por sí solos. El método es una extensión de KBE y un modelo de predicción basado en regresión espacial para la predicción precisa de datos a escala local (Giustini *et al.* 2019). La ventaja de RKEB sobre otros métodos lineales o enfoques geoestadísticos es que considera la no estacionariedad y la relación entre variables primarias y secundarias en consideración. Aparte de esto, RKBE aborda el problema de la multicolinealidad con las variables explicativas tomando en cuenta sus componentes principales, como un medio de reducción de la dimensión para mejorar las predicciones. En este artículo, se utilizó el software ArcGIS Pro® para llevar a cabo la predicción del modelo RKBE utilizando las variables explicativas representativas de la zona. En este estudio se utilizaron mapas ráster de geología, precipitación anual, altitud

(msnm), uso y cobertura de la tierra e índice de vegetación de diferencia normalizada (NDVI) de la zona estudiada.

### **Bosques Aleatorios (RF, Random Forest)**

El método *Random Forest* (RF) es un algoritmo de aprendizaje automático que se ha utilizado con éxito para predecir el carbono orgánico del suelo en varias regiones y condiciones de suelo. *Random Forest* es un algoritmo de aprendizaje automático de tipo ensamble, lo cual significa que combina varios algoritmos de aprendizaje individuales para producir una mejor precisión en las predicciones. Este algoritmo crea varios árboles de decisión y combina las predicciones hechas por cada árbol para producir una predicción final.

RF es un método de aprendizaje automático (ML, *machine learning*) ampliamente utilizado, que consiste en un conjunto de clasificación aleatoria y árboles de regresión. El algoritmo de RF hace crecer diferentes árboles seleccionando al azar y repetidamente variables predictoras y casos de entrenamiento para desarrollar una población aleatoria de árboles. El algoritmo crece un conjunto de árboles de regresión basados en partición recursiva binaria, donde el espacio predictor en cada árbol el nodo se divide en función de divisiones binarias en un subconjunto de predictores seleccionados aleatoriamente. La salida de RF es el promedio de las predicciones de árboles individuales. Se ha demostrado que el algoritmo de RF puede ser muy eficiente, especialmente cuando el número de descriptores es muy grande. El modelo RF es capaz de manejar simultáneamente variables categóricas y continuas, así como relaciones complejas de variables de alto orden como la no linealidad y los efectos de interacción (Breiman, 2001).

### **Análisis estadístico**

Los datos de carbono orgánico de suelo se sometieron a un análisis exploratorio (AED) con apoyo del paquete estadístico SPSS® (IBM® *Statistics*, versión 20), para determinar los estadísticos descriptivos, como: media, mediana, varianza, coeficiente de variación, valores máximos y mínimos, y los índices de asimetría y curtosis. Para detectar la presencia de valores atípicos se utilizó la metodología de cercas externas e internas de Tukey (1977), y posteriormente se realizó la prueba de normalidad de Kolmogorov-Smirnov, para evaluar la distribución de los datos.

### **Evaluación de los modelos de carbono orgánico del suelo**

Los modelos obtenidos por los métodos estadísticos avanzados (KBE, RKBE) se generaron con el total de los datos, y la exactitud de los mapas se obtuvo mediante validaciones cruzadas. En la evaluación de la precisión de los modelos se utilizaron cinco (5) índices: el error medio (EM), el error medio estandarizado (EME), la raíz del error cuadrático medio (RECM), la raíz del error cuadrático medio estandarizado (RECME), y el error estándar promedio (EEP). El EM evalúa el error sistemático e indica la presencia de subestimación (-) o sobrestimación del modelo (+) y el EME muestra la desviación del modelo obtenido. RECM evalúa la precisión de la

predicción y mide la cantidad de error que hay entre los conjuntos de datos medidos (observados) e inferidos (estimados); es decir, compara un valor predicho y un valor conocido. El índice RECME es más preciso mientras más se aproxime al valor ideal de la unidad (1), si el error estandarizado de la raíz cuadrada media es mayor que 1, existe subestimación de la variabilidad en sus predicciones, y si es menor que 1, se sobreestima la variabilidad en sus predicciones. Los índices RECM y EM fueron estandarizados con la desviación estándar de los residuos de los datos. El EEP indica la variabilidad de las predicciones, cuyas estimaciones serán más adecuadas si sus valores están más cerca del ECM. Para la evaluación de la confiabilidad de los modelos de predicción de COS se utilizó el coeficiente de determinación y el coeficiente de concordancia (CC) (AC, por *agreement concordance*). Ambos coeficientes son medidas estadísticas utilizadas para evaluar la variación, y la consistencia o el grado de acuerdos entre lo medido y lo inferido.

## RESULTADOS Y DISCUSIÓN

### Análisis estadístico de los datos

La base de datos conformada por 2.279 muestras de COS fue depurada de valores atípicos, faltantes e incongruentes, cuyos resultados se presentan en el Cuadro 1. Los estadísticos descriptivos indicaron que los valores promedios del COS son similares a la mediana y la moda de los datos, indicando que no fue necesaria la transformación de los mismos para su procesamiento.

**Cuadro 1.** Estadísticos descriptivos de los datos de carbono orgánico del suelo.

Estadístico	Valor
Numero de datos (n)	2.216
Mediana	1,28
Media	1,26
Moda	1,22
Desviación estándar (S)	0,44
Varianza (S <sup>2</sup> )	0,19
Coficiente de Variación (%)	34,8
Curtosis (K)	-0,13
Asimetría (As)	-0,05
Primer cuartil (1Q)	0,99
Tercer cuartil (3Q)	1,57
Mínimo	0,10
Máximo	2,96

Fuente: Datos CIESA UNERG, 2000-2022.

Así mismo, los valores de COS oscilan entre 0,10 y 3 %, con desviaciones y varianza en mínimas proporciones y con coeficientes de variación cercano al 35%. Finalmente, en cuanto a la prueba de normalidad, se verificó que los datos de CO

proceden de poblaciones normales, ya que los valores de la prueba estadística son significativos ( $p > 0,05$ ).

### Estimación del carbono orgánico del suelo

Los resultados de la estimación del COS por el método geoestadístico KBE indicaron que el ajuste de los datos se realizó con varios subconjuntos de 100 datos y un total de 100 simulaciones, mediante un modelo matemático tipo potencia (*Power*) (8 sectores), con un alcance de 407 metros de radio y una función de regresión  $Y = -1,08921372335183 * x + 2,36621585739209$ . La figura 3 señala el semivariograma seleccionado que proporcionó el mejor ajuste para las semivarianzas empíricas (cruces azules del gráfico). En esta situación, las semivarianzas empíricas están en la parte central del espectro del semivariograma. Semivarianzas empíricas caen en el medio del espectro, ajustándose entre los cuartiles  $Q_1$  y  $Q_3$  (líneas rojas segmentadas). De igual manera, la figuras 4 indica la expresión de la relación entre los valores observados y los valores estimados, y los errores del modelo de predicción.

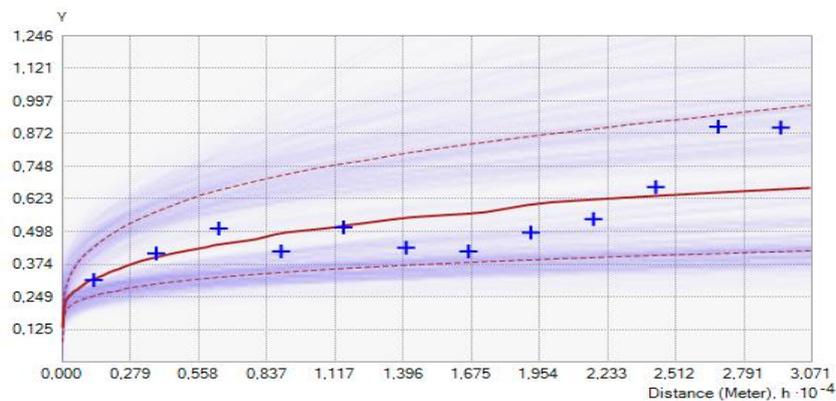


Figura 3. Semivariograma empírico del modelo KBE ajustado.

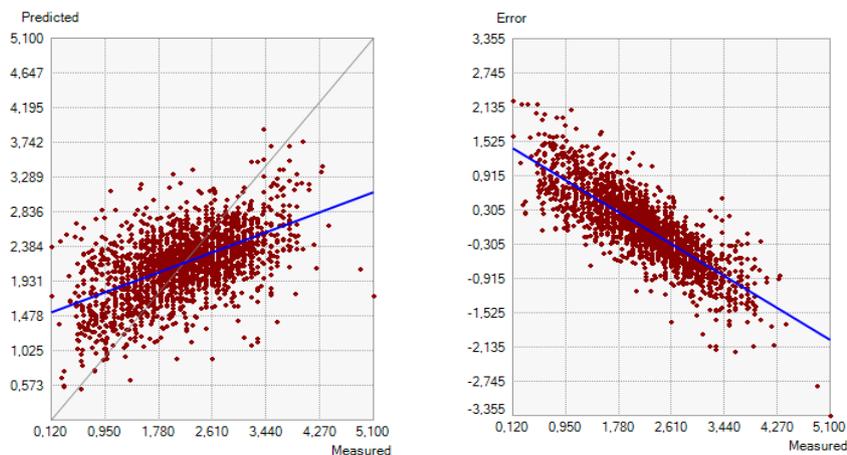
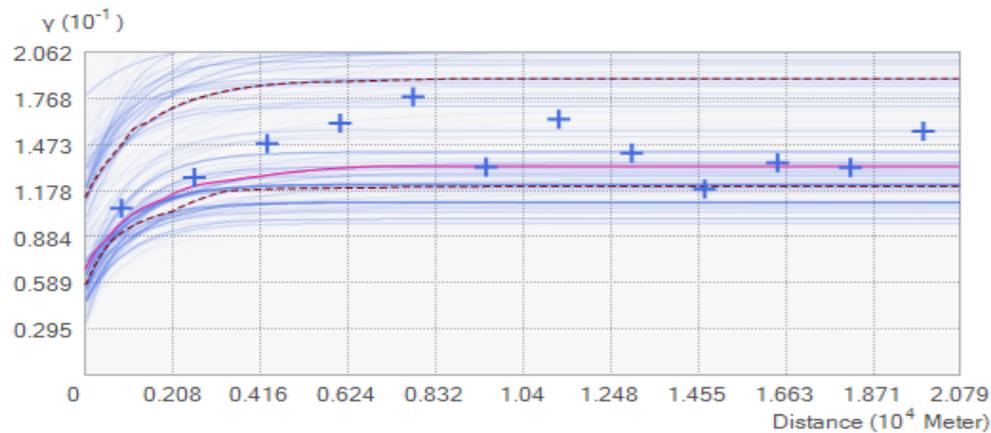
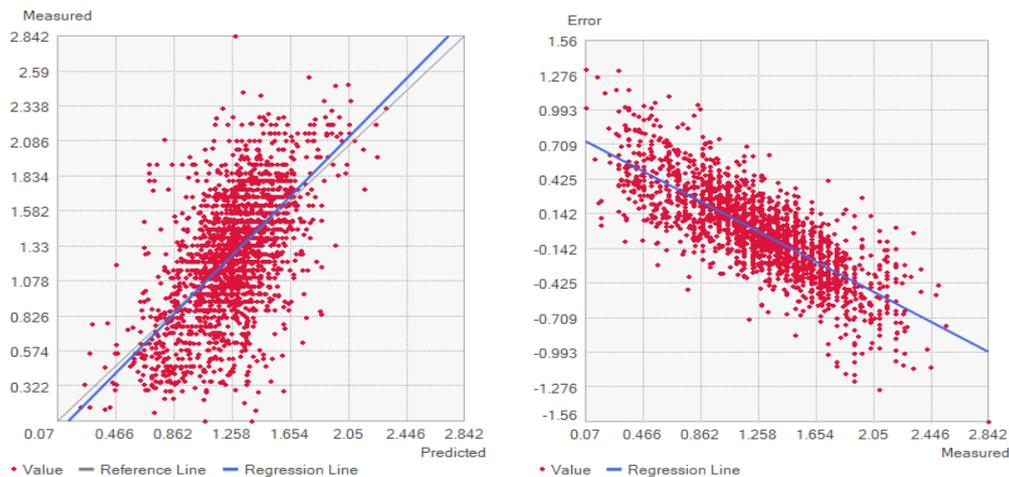


Figura 4. Variación de datos de carbono orgánico del suelo y error de la estimación con el método KBE.

En cuanto a la aplicación del método RKBE, la figura 5 muestra el semivariograma empírico ajustado con un modelo de tipo exponencial a través de la función de regresión:  $Y = 0.305872426668787 * x + 0.872446919115828$ . En la figura 6 se visualiza la relación entre los valores observados y estimados por el modelo, y el grado de incertidumbre de la estimación mediante la técnica de validación cruzada. Este método presenta menor grado de incertidumbre que el KEB debido principalmente a la utilización de variables regresoras como geología, precipitación anual, NDVI, altitud y la cobertura y uso de la tierra de la región oriental del estado Guárico.



**Figura 5.** Semivariograma empírico del método RKBE ajustado con un modelo matemático tipo exponencial.



**Figura 6.** Variación de datos del carbono orgánico del suelo y error de la estimación con el método RBKE.

Con relación a los resultados de la aplicación del método de bosques aleatorios (RF), en el Cuadro 2 se visualiza el aporte de las principales variables auxiliares en la predicción del COS en la zona.

**Cuadro 2.** Importancia de las variables auxiliares en la estimación del COS en la zona oriental del estado Guárico.

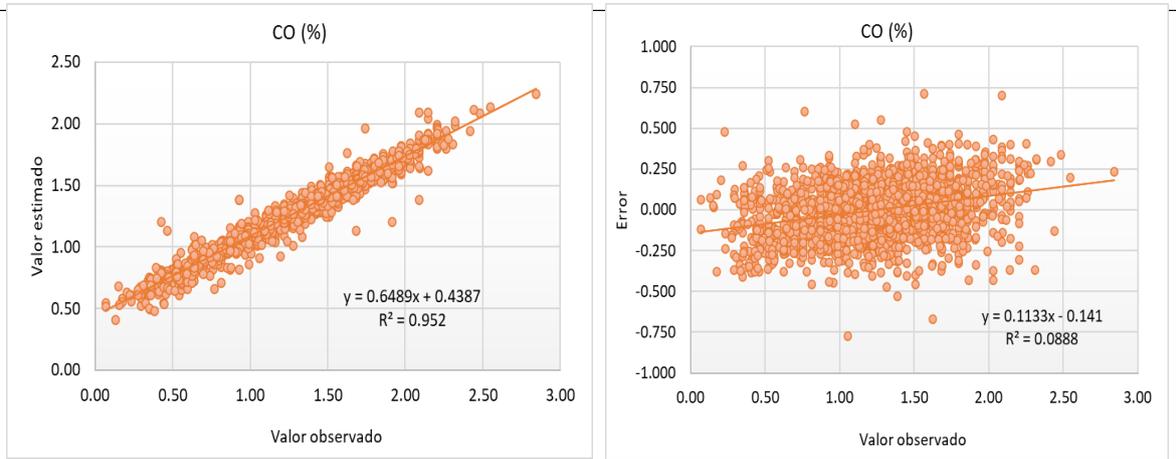
Variable	Importancia	%
CNBL	36,39	10
PM Anual	35,75	10
NDVI	34,98	10
Altitud	34,28	10
ITH	29,90	8
Orientación	29,59	8
Área de Captación	29,27	8
VDCN	28,25	8
Posición Relativa	27,10	8
Pendiente	25,99	7
Factor LS	25,92	7
Geología	12,45	3
Uso y Cobertura	6,59	2

CNBL: Nivel base de la red de drenaje (altura del terreno sobre la red de drenaje), VDCN: Distancia vertical a la red del canal o fluvial, PM: Precipitación anual, NDVI: índice de vegetación de diferencia normalizada, ITH: Índice Topográfico de Humedad, Factor LS: Índice de Transporte de Sedimentos.

El mayor peso en la estimación del COS esta dado por las variables altura del terreno sobre la red de drenaje, la precipitación media anual, la altitud y el índice de vegetación, con una importancia relativa superior al 10%. Muchas de las variables utilizadas aportan menos del 8% en el proceso de interpolación del COS, tales como: ITH, posición relativa, pendiente del terreno, factor LS, geología, cobertura y uso de la tierra. En la figura 7 se visualiza la relación entre los valores observados y estimados por el modelo RF, con el 90% de los datos de COS. Al respecto se evidencia que dicho modelo es capaz de explicar el 93% de la variación del CO en los suelos agrícolas de la zona Oriental del estado Guárico. La validación con un grupo de datos independientes, indico que el modelo presenta una consistencia aceptable.

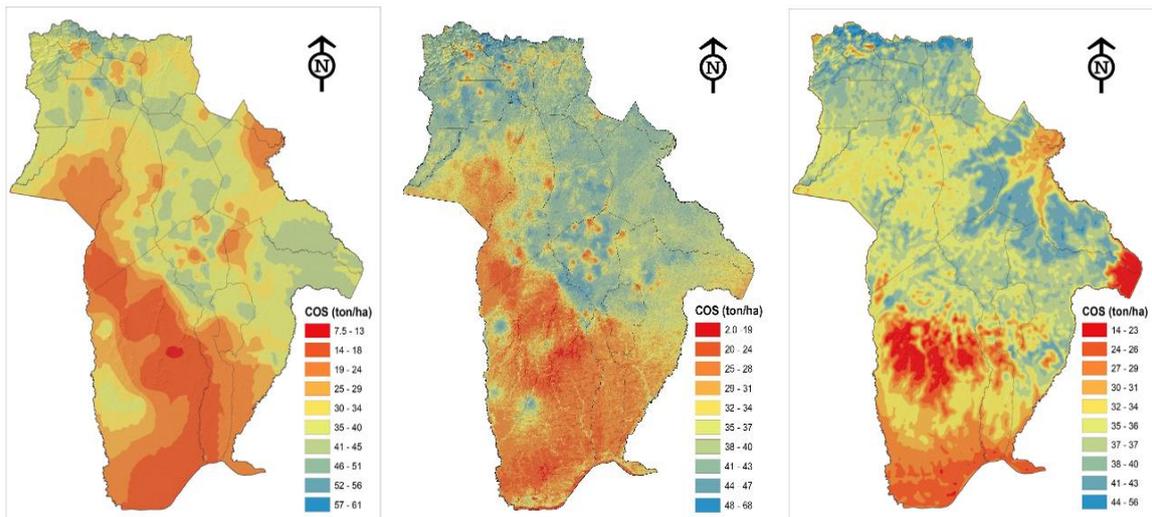
### **Cartografía digital del carbono orgánico del suelo**

Los resultados de los mapas continuos del COS con la aplicación de los tres métodos de predicción se presentan en la Figura 8, con variaciones que oscilan entre 2 y 67 t h<sup>-1</sup> en forma global. En los tres casos, el color rojo indica la menor reserva de COS en la capa superficial y el color azul refleja la mayor cantidad.



**Figura 7.** Relación de valores estimados y observados en la generación del modelo con el método de bosques aleatorios (RF).

En todos los mapas se evidencia un incremento del contenido de carbono orgánico en la zona Norte, la cual corresponde a áreas montañosas del municipio Monagas, con dominancia de vegetación boscosa y la existencia de condiciones hidrológicas bien definidas, y una mayor precipitación media anual, lo que ha permitido la acumulación de materiales orgánicos en la capa superficial del suelo.



**Figura 8.** Variación espacial del COS con los métodos kriging empírico bayesiano (KBE), Regresión KBE y bosques aleatorios (RF).

La región central de la zona de estudio presenta variaciones moderadas de la reserva de COS, cuyo comportamiento responde la extensa actividad agrícola realizada en dicha área de manera convencional y tradicional desde tiempos remotos. Las variaciones corresponden posiblemente a la influencia de acumulación de humedad dado por el área de captación y el índice topográfico de humedad en suelos de relieves colinosos de los municipios Chaguaramas, Infante, Ribas, El Socorro y Zaraza.

Los modelos digitales también reflejan un patrón de distribución que diferencia el contenido de CO en los suelos de la parte Sur del área de estudio, donde predominan suelos formados sobre materiales geológicos con dominancia de la Formación Mesa, y una dominancia de cobertura de pastos naturales. Ello parece proporcionar al suelo una menor capacidad de reserva de COS, al igual que la zona más oriental, cuya reserva está dada por un mayor grado de incertidumbre determinado por la menor cantidad de muestras representativas de la zona. Este factor interfiere en las estimaciones más apropiadas.

Los resultados de CO obtenidos en los suelos agrícolas de la región estudiada se encuentran entre los rangos de la reserva obtenida en suelos de otras regiones. De esta manera, Odebiri et al. (2022) encontraron en suelos de África del Sur que las existencias de COS eran relativamente proporcionales a la cobertura de la tierra, con valores promedios de 46,1 t h<sup>-1</sup> en cuanto a la capacidad de secuestro de carbono. También Park et al. (2021) emplearon técnicas de aprendizaje automático y encontraron en áreas agrícolas de Corea del Sur, valores comprendidos entre 35 y 87 t h<sup>-1</sup> para profundidades de 30 y 100 cm, respectivamente. De igual manera, Farooq et al. (2022) emplearon el método de bosques aleatorios en suelos agrícolas de la región del Himalaya, y encontraron que la concentración de COS varió de 1,12 t h<sup>-1</sup> a 70,60 t h<sup>-1</sup>, cuyo método proporcionó una mejor estimación y variabilidad espacial de dicho atributo.

Con base a los resultados de COS derivados de las grandes unidades de producción agrícola, ha sido de gran interés el uso de las covariables o variables auxiliares como base del mapeo digital del carbono orgánico, lo cual cada día cobra una mayor atención.

Sin embargo, es importante indicar que bajo el sistema de producción actual llevado a cabo en la zona guariqueña, el retorno de la soca de los cultivos tradicionales (maíz, sorgo) y los residuos después de la cosecha, aumentan el aporte de carbono al suelo (Huang et al., 2007), mejora la estructura del suelo y la disponibilidad de nutrientes y aumenta el tamaño de la población microbiana del suelo (Li et al., 2020), razones por la que es apropiado tomar acciones para un manejo sostenible de los suelos de la región.

### **Evaluación de modelos digitales de carbono orgánico del suelo**

Los resultados comparativos de la evaluación de los modelos obtenidos se reflejan en el Cuadro 3. Se observa que los modelos que mejor se adaptan a los datos empleados cumplen con los requerimientos de ECM pequeño, EEP pequeño próximo a RECM, y RECME cerca de 1. En el caso del modelo RF, el error estandarizado de la raíz cuadrada media es ligeramente menor que 1, lo que significa que en el desarrollo del modelo existe cierto grado de sobreajuste en las

predicciones. Sin embargo, en la validación del modelo, el índice RECME es mayor que 1, lo cual indica que existe una ligera subestimación de la variabilidad de las predicciones.

**Cuadro 3.** Estimación del grado de incertidumbre de los modelos digitales de COS

Índice	KEB*	RKEB*	RF <sup>1</sup>	RF <sup>2</sup>
EM	-0.0003	-0.001	0.001	0.026
RECM	0.366	0.349	0.164	0.350
EME	-0.001	-0.000	-0.012	0.040
RECME	0.990	0.949	0.790	1.248
EEP	0.364	0.368	0.213	0.205

\*Obtenido por validación cruzada. <sup>1</sup>Entrenamiento con 90% de datos. <sup>2</sup>Validación con 10% de datos. EM: error medio, EME: Error medio estandarizado, ECM: error cuadrático medio; ECME: error cuadrático medio estandarizado; EEP: error estándar promedio.

**Cuadro 3.** Estimación de la confiabilidad del modelo digital de COS estimado por el método RF

Índice	KEB*	RKEB*	RF <sup>1</sup>	RF <sup>2</sup>
R <sup>2</sup>	0.297	0.352	0.952	0.377
CC	0.670	0.710	0.947	0.832

\*Obtenido por validación cruzada. <sup>1</sup>Entrenamiento con 90% de datos. <sup>2</sup>Validación con 10% de datos. CC: coeficiente de concordancia.

Todos los modelos fueron capaces de estimar la reserva de COS de acuerdo a sus alcances, con muy bajos grados de incertidumbre. Los dos primeros modelos (KBE y RKBE) muestran grandes fortalezas, pero éstas son superiores en el segundo método, cuya precisión está dada por el uso de las variables auxiliares utilizadas como apoyo en la estimación. Sin embargo, el método RF generó un modelo con un alto grado de concordancia entre los valores observados y estimados, y un alto coeficiente de determinación, lo cual demostró una alta potencialidad en zonas de alta complejidad y extensión geográfica. De acuerdo con los resultados, el RKBE no identifica las variables independientes que están fuertemente asociadas con las variables de respuesta ni la importancia de las variables explicativas que influyen en las variables de predicción, por lo que es necesario seleccionar las variables a través de juicio de expertos con algún método estadístico de reducción de variables. Sin embargo, el método de bosque aleatorio (RF) demostró ser una técnica de aprendizaje automático capaz de resolver las limitaciones asociadas con el método RKBE (Molla *et al.*, 2022).

La situación descrita, evidencia la complejidad que presenta la variación espacial del carbono orgánico del suelo en el ámbito geográfico. Sin embargo, la Cartografía

Digital presenta una gran oportunidad de utilizar diferentes métodos para generar la distribución espacial de propiedades del suelo, desde los más tradicionales hasta los más sofisticados. Esto amplía la visión del experto para seleccionar los modelos más apropiados de acuerdo con los intereses u objetivos planteados, especialmente en suelos de amplia trayectoria y tradición agrícola.

## CONCLUSIONES

Los mapas de COS proporcionaron información útil para monitorear las condiciones del suelo, identificar áreas degradadas, y tomar decisiones basadas en evidencias cuantificables, para mitigar y adaptarse a un clima cambiante.

El mapeo digital del suelo permitió la evaluación de la reserva de COS en áreas agrícolas extensas, lo cual evidencia la importancia de conocer su distribución espacial para la exploración de zonas potenciales de secuestro de COS.

La aplicación de las técnicas de cartografía digital generó información base que puede permitir un manejo de suelo apropiadamente, lo cual puede contribuir a reducir los gases de efecto invernadero y permitir la implementación de medidas ante los efectos del cambio climático.

Los métodos empleados pueden ser utilizados para la evaluación de la variabilidad espacial de COS en un espacio geográfico extenso complejo, y por lo tanto los modelos finales podrían ser empleados por los planificadores como una herramienta de apoyo a la toma de decisiones en el manejo agrícola, con miras hacia una agricultura de mayor precisión en el espacio y mayor sostenibilidad en el tiempo.

## AGRADECIMIENTO

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), municipio Juan German Roscio, estado Guárico, Venezuela.

## REFERENCIAS

- Bashir, O.; Ali, T.; Baba, Z.A.; Rather, G.H.; Bangroo, S.A.; Mukhtar, S.D.; Naik, N.; Mohiuddin, R.; Bharati, V.; Bhat, R.A. (2021). Soil Organic Matter and Its Impact on Soil Properties and Nutrient Status. In *Microbiota and Biofertilizers*; Springer: Berlin/Heidelberg, Germany, Volume 2, pp. 129–159.
- Breiman, L. (2001). Random Forests. *Machine Learning* 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Conrad, O., B. Bechtel, M. Bock, H. Dietrich, E. Fischer, L. Gerlitz, J. Wehberg, V. Wichmann, Böhner, J. (2015). *Geosci. Model Dev. Discuss.*, 8, 2271–2312, 2015. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. doi:10.5194/gmdd-8-2271-2015

- Dobos, E. F. Carré; T. Hengl; H.L. Reuter; Tóth, G. (2006). Digital Soil Mapping as a support to production of functional maps. EUR 22123 EN. Office for Official Publications of the European Communities, Luxemburg. 68 p.
- Farooq, I.; Bangroo, S.A.; Bashir, O.; Shah, T.I.; Malik, A.A.; Iqbal, A.M.; Mahdi, S.S.; Wani, O.A.; Nazir, N.; Biswas, A. (2022). Comparison of Random Forest and Kriging Models for Soil Organic Carbon Mapping in the Himalayan Region of Kashmir. *Land* (11): 2180. <https://doi.org/10.3390/land11122180>
- Fick, S.E.; Hijmans, R.J. (2017). WorldClim 2: New 1km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* 37 (12): 4302-4315.
- Grunwald S. (2022). Artificial intelligence and soil carbon modeling demystified: power, potentials, and perils. *Carbon Footprints*, (2)1:6. <https://dx.doi.org/10.20517/cf.2022.03>
- Huang B, Sun W, Zhao Y, Zhu J, Yang R, Zou Z, Ding F and Su J. 2007. Temporal and spatial variability of soil organic matter and total nitrogen in an agricultural ecosystem as affected by farming practices *Geoderma*, 139 336-45
- IGVSB. (2004). Cartografía Nacional. Hojas cartográficas a escala 1:25.000 y 1:100.000 producidas en el Instituto Geográfico Venezolano Simón Bolívar. Caracas, Venezuela.
- INE. (2007). Censo de Población y Vivienda. Instituto Nacional de Estadísticas. Venezuela. [En línea]: [http://www.ine.gob.ve/documentos/Demografia/CensodePoblacionyVivienda/pdf/tendencia\\_pobreza\\_censo2007.pdf](http://www.ine.gob.ve/documentos/Demografia/CensodePoblacionyVivienda/pdf/tendencia_pobreza_censo2007.pdf)
- IPCC. (2016). Intergovernmental Panel on Climate Change (IPCC) (2016). <http://www.ipcc.ch/> (acceso noviembre 2022).
- Krivoruchko, K. (2012). Empirical Bayesian Kriging Implemented in ArcGIS Geostatistical Analyst. Senior Research Associate, Software Development Team, Esri. <https://www.esri.com/news/arcuser/1012/files/ebk.pdf>
- Krivoruchko, K., Gribov, A. (2019). Evaluation of empirical Bayesian kriging. *Spatial Statistics*, 32, 100368. <https://doi.org/10.1016/j.spasta.2019.100368>
- Li Y.; Li Z, Chang S X, Cui S, Jagadamma S, Zhang Q; Cai Y. 2020. Residue retention promotes soil carbon accumulation in minimum tillage systems: implications for conservation agriculture *Sci. Total Environ.* 740 140147
- Liu, S.; An, N.; Yang, J.; Dong, S.; Wang, C.; Yin, Y. (2015). Prediction of soil organic matter variability associated with different land use types in mountainous landscape in southwestern Yunnan province, China. *Catena*, 133: 137-144.
- McBratney, A.B.; Mendonça Santos, M.L.; B. Minasny, B. (2003). On digital soil mapping. *Geoderma* Volume 117 (1-2) 3-52.
- MARN. 1983. Sistemas Ambientales Venezolanos. Proyecto VEN 79/ 001. Región Natural 18 Llanos Altos Centrales. Serie II: Los Recursos Naturales Renovables y las Regiones Naturales, Sección 2, Documento No. 18. Código II-2-18. Caracas. 226 p.
- Molla, A.; Zhang, W.; Zuo, S.; Ren, Y.; Han, J. (2022). A machine learning and geostatistical hybrid method to improve spatial prediction accuracy of soil potentially toxic elements. *Research Square*. DOI: <https://doi.org/10.21203/rs.3.rs-1306764/v1>
- Odebiri, O., Mutanga, O., Odindi, J., Peerbhay, K., Dovey, S., Ismail, R. (2020). Estimating soil organic carbon stocks under commercial forestry using topo-climate variables in KwaZulu-Natal, South Africa. *S. Afr. J. Sci.* 116 (3-4), 1-8.
- Penman, J., Gytarsky, M., Hiraishi, T., Krug, T., Kruger, D., Pipatti, R., Wagner, F. (2003). Good practice guidance for landuse, land-use change and forestry. Japan: IPCC [accessed Mar 26 2023].

- 
- Santhi, C.; J.G. Arnold, J.R. Williams; W.A. Dugas; R. Srinivasan; Hauck, L.M. (2001). Validation of the SWAT model on a large river basin with point and nonpoint sources. *J. American Water Resources Assoc.* 37(5): 1169-1188.
- Park, Seong-Jin; Soon-Ik Kwon, Seong-Heon Kim, Jaehong Shim, Yun-Hae Lee; Taek-Keun Oh. 2021. Estimation of Soil Organic Carbon (SOC) Stock in South Korea Using Digital Soil Mapping Technique. *Korean J. Soil Sci. Fert.* Vol.54, No.2, pp.247-256, pISSN: 0367-6315, eISSN: 2288-2162. <https://doi.org/10.7745/KJSSF.2021.54.2.247>
- Tranter, G., B. Minasny, A.B. Mcbratney, B. Murphy, N.J. Mckenzie, M. Grundy, D. Brough. 2007. Building and testing conceptual and empirical models for predicting soil bulk density. *Soil Use Manage.* 23(4):437-443.
- Walkley, A., Black, I. A. (1934). An examination of the method for determining soil organic matter, and a proposed modification of the chromic acid titration method. *Soil Science* 37 (29-38).
- Willmott C.J., Robeson S.M., Matsuura, K. (2012). A refined index of model performance. *International Journal of Climatology*, 32: 2088-2094.