



Madera y bosques

ISSN: 1405-0471

ISSN: 2448-7597

Instituto de Ecología A.C.

Pérez-Vázquez, Zaira Rosario; Ángeles-Pérez, Gregorio; Chávez-Vergara, Bruno; Valdez-Lazalde, José René; Ramírez-Guzmán, Martha Elva
Enfoque espacial para modelación de carbono en el mantillo de bosques bajo manejo forestal maderable
Madera y bosques, vol. 27, núm. 1, e2712122, 2021
Instituto de Ecología A.C.

DOI: <https://doi.org/10.21829/myb.2021.2712122>

Disponible en: <https://www.redalyc.org/articulo.oa?id=61766374013>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org



Enfoque espacial para modelación de carbono en el mantillo de bosques bajo manejo forestal maderable

Spatial approach for modeling litter carbon in forests under management for timber production

Zaira Rosario Pérez-Vázquez¹, Gregorio Ángeles-Pérez^{1*}, Bruno Chávez-Vergara^{2,3},
José René Valdez-Lazalde¹ y Martha Elva Ramírez-Guzmán⁴

¹ Colegio de Postgraduados. Postgrado en Ciencias Forestales. Montecillo, Estado de México, México.

³ Laboratorio Nacional de Geoquímica y Mineralogía. Ciudad de México, México.

* Autor de correspondencia. gangeles@colpos.mx

² Universidad Nacional Autónoma de México. Instituto de Geología. Departamento de Ciencias Ambientales y del Suelo. Ciudad de México, México.

⁴ Colegio de Postgraduados. Postgrado en Estadística. Montecillo, Estado de México, México.

RESUMEN

El piso forestal o mantillo es el almacén de carbono que regula la mayoría de los procesos funcionales de los ecosistemas forestales, influyendo directamente en la fertilidad del suelo y en la productividad del sitio. El contenido de carbono en el piso forestal es altamente variable en espacio y tiempo; por ello, obtener evaluaciones precisas del carbono contenido en este almacén representa un desafío metodológico importante a cualquier escala. En este estudio, se compararon cuatro métodos de modelación espacial para mapear el contenido de carbono en el piso forestal de un bosque templado. Los métodos fueron *kriging* ordinario, modelo lineal generalizado, modelo aditivo generalizado y *random forest*. Las estimaciones del contenido de carbono fueron realizadas para 2013 y 2018. Las variables predictoras representan la estructura espacial, del dosel y topográfica presente en el área de estudio. Todos los modelos fueron evaluados mediante validación cruzada y se determinó el error medio absoluto, el error cuadrático medio y el coeficiente de determinación. El desempeño de los métodos fue, en orden decreciente: *random forest*, modelo aditivo generalizado, modelo lineal generalizado y *kriging* ordinario. El método *kriging* ordinario reflejó el grado de dependencia espacial del contenido de carbono, pero las estimaciones espaciales fueron poco realistas ($R^2 \leq 0.35$). El modelo aditivo generalizado y el modelo lineal generalizado mostraron buen desempeño ($R^2 \geq 0.70$), pero mayor sobreestimación; *random forest* obtuvo el mejor ajuste ($R^2 \geq 0.86$) para modelar contenido de carbono en ambos años evaluados. Se concluye que *random forest* es un método prometedor, con gran potencial para mejorar las estimaciones de carbono en el mantillo a escala de paisaje.

PALABRAS CLAVE: almacenes de carbono; geoestadística; modelo aditivo generalizado; modelo lineal generalizado; *random forest*; variabilidad espacial.

ABSTRACT

Forest floor or mulch is the carbon stock that regulates most of the functional processes of forest ecosystems, directly influencing soil fertility and site productivity. The forest floor carbon content is highly variable in space and time; therefore, obtaining accurate assessment of the carbon contained in this stock represents a significant methodological challenge at any scale. In this study, four spatial modeling methods were compared to map forest floor carbon content in a temperate forest. The methods were ordinary kriging, generalized linear model, generalized additive model and random forest. Carbon content estimates were made for 2013 and 2018. The predictor variables represent the spatial, canopy and topographic structure present in the study area. All models were evaluated by ten-fold cross validation and the mean absolute error, root mean square error and the coefficient of determination (R^2) were determined. The performance of the methods was, in decreasing order, *random forest*, generalized additive model, generalized linear model and ordinary kriging. The ordinary kriging method reflected the degree of spatial dependence of carbon content, but the spatial estimates were unrealistic ($R^2 \leq 0.35$). Generalized additive model and generalized linear model showed good performance ($R^2 \geq 0.70$), but higher overestimation; random forest obtained the best fit ($R^2 \geq 0.86$) to model carbon content in both years evaluated. It is concluded that random forest is a promising method with high potential to improve estimates of forest floor carbon at landscape scale.

KEYWORDS: carbon stocks; geostatistics; generalized additive model; generalized lineal model; forest floor; *random forest*; spatial variability.

INTRODUCCIÓN

El piso forestal o mantillo es la capa superficial del suelo formada por residuos orgánicos en diferentes grados de descomposición. Esta capa orgánica puede almacenar considerables cantidades de carbono (C) y regular la mayoría de los procesos funcionales de los ecosistemas, por influir directamente en la productividad forestal y la fertilidad de los suelos (Schlesinger y Benhardt, 2013). El Panel Intergubernamental sobre Cambio Climático [IPCC] (2006) reconoce al piso forestal como el tercer almacén de C más importante de los ecosistemas terrestres, representando a escala global $\sim 5\%$ ($43 \text{ Pg} \pm 3 \text{ Pg}$ de C) de todas las reservas de C de los ecosistemas forestales (Pan *et al.*, 2011). Dada su importancia, es crítico estimar y monitorear con precisión los almacenes de carbono en el piso forestal.

En México, las estimaciones de los almacenes de carbono del piso forestal se basan casi exclusivamente en mediciones de campo (Comisión Nacional Forestal [Conafor], 2012; López-Escobar *et al.*, 2017) y, a la fecha, son pocos los estudios que exploran su variabilidad espacial (Martínez-Yrízar, Nuñez, Miranda y Bürquez, 1999; Jaramillo, Kauffman, Rentería-Rodríguez, Cummings y Ellingson, 2003; Anaya, 2007; Pérez-Suárez, Arredondo-Moreno y Huber-Sannwald, 2012). Este tipo de evaluaciones son un desafío metodológico debido a la alta variación espacial de factores biofísicos en el paisaje y a la influencia de perturbaciones humanas y/o naturales que, al interactuar entre sí, forman patrones complejos de C en el espacio y en el tiempo (Grunwald, 2009; Zhang *et al.*, 2014).

La introducción de tecnologías digitales, como los sensores remotos y los sistemas de información geográfica, así como una mejora en la velocidad del procesamiento computacional, el manejo de datos espaciales y el desarrollo de nuevos métodos cuantitativos, han proporcionado nuevas oportunidades para describir patrones, propiedades y procesos del suelo. El mapeo digital del suelo (DSM, por sus siglas en inglés) es uno de los principales enfoques que combinan las tecnologías anteriores para cuantificar y modelar la distribución espacial de los almacenes de carbono en el suelo, desde escalas locales hasta globales

(Grunwald, 2009). El enfoque DSM utiliza diversos métodos numéricos para calibrar y validar modelos estadísticos de una variable regionalizada, a partir de covariables ambientales asociadas (Beguin, Fuglstad, Mansuy y Paré, 2017). En DSM, se suele incluir métodos geoestadísticos (Zhang y McGrath, 2004; Bhunia, Shit y Maiti, 2018; Yao *et al.*, 2019), modelos de regresión (Meersmans, De Ridder, Canters, De Baets y Van Molle, 2008), y más recientemente, métodos de *machine learning* (Cao, Domke, Russell y Walters, 2019). Estos modelos han demostrado ser una alternativa efectiva y poco costosa para estimar carbono en los distintos componentes del ecosistema, en comparación con enfoques de estimación tradicionales. Sin embargo, debido a la complejidad del atributo de interés, así como a la calidad y disponibilidad de información, algunos de estos métodos no son capaces de mapear con precisión las existencias de C, y en algunas ocasiones, seleccionar un método óptimo de predicción no siempre es prometedor. Por lo cual, una adecuada selección del método con menor incertidumbre proporcionará la base para evaluar el potencial de almacenamiento de C en el piso forestal, además contribuirá al desarrollo de estrategias para un manejo forestal sostenible.

OBJETIVOS

El objetivo de este estudio fue comparar el desempeño de cuatro métodos para modelar la distribución espacial del contenido de carbono en el piso forestal, los cuales representan las rutinas estadísticas más utilizadas en el mapeo digital del suelo. Los métodos fueron: a) *kriging* ordinario (interpolación espacial), b) modelo lineal generalizado (modelo de regresión), c) modelo aditivo generalizado (modelo de regresión) y d) *random forest* (*machine learning*).

MATERIALES Y MÉTODOS

Área de estudio

Este estudio se desarrolló en el Sitio de Monitoreo Intensivo del Carbono (SMIC-Atopixco), localizado en el Municipio de Zacualtipán, Hidalgo, México. El SMIC-



Atopixco forma parte de la Red Mexicana de Sitios de Monitoreo Intensivo de Carbono (Red Mex-SMIC) (Ángeles-Pérez *et al.*, 2015) y se encuentra delimitado dentro del paisaje forestal por un polígono de 9 km² (900 ha), dividido en nueve cuadros de 1 km² (Fig. 1). Se ubica entre las coordenadas 20°40'17" y 20°34'51" latitud norte y 98°40'07" y 98°34'22" longitud oeste. La elevación varía entre 1836 m y 2097 m s.n.m. Predominan los suelos Feozem háplico y Regosol calcárico (Santiago-García, de los Santos-Posadas, Ángeles-Pérez, Valdez-Lazalde y Ramírez-Valverde, 2013). El clima es C(fm)w'b(e)g, templado húmedo con una marcada estación de lluvias entre junio y octubre (García, 1981). La temperatura media anual es de 13.5 °C y la precipitación total anual es de 2050 mm.

Esquema de muestreo

Las muestras fueron colectadas en 40 unidades de muestreo permanente (Fig. 1), establecidas entre 2012 y 2013 dentro del SMIC-Atopixco (Ángeles-Pérez *et al.*, 2015). Estas unidades (conglomerados) son similares a las utilizadas en el Inventario Nacional Forestal y de Suelos (INFyS) (Conafor, 2012) y siguen el esquema de muestreo propuesto por Hollinger (2008): a) el cuadro central cuenta con una torre de flujos turbulentos y tiene establecidos 16 conglomerados dispuestos en forma sistemática; b) los ocho cuadrantes periféricos cuentan con un conglomerado en su centro, siguiendo el diseño sistemático; c) se cuenta con 16 conglomerados adicionales distribuidos bajo un enfoque de cronosecuencia (establecidos en rodales con diferente edad que no fueron representados dentro del esquema sistemático).

Colecta y análisis de muestras

Para propósitos de este estudio, el piso forestal fue caracterizado en dos capas que representan las dos etapas de descomposición del material orgánico: 1) la capa Ho, identificada como los residuos vegetales poco alterados por su descomposición y donde aún se pueden identificar claramente los componentes; y 2) la capa de fermentación

(F), en el que la materia orgánica está descompuesta a tal grado que no se reconocen estructuras completas (Conafor, 2012). Con fines de comparación, se incluyó la capa total del piso forestal (Tot), obtenida por la suma de la capa Ho y F.

La colecta de muestras se realizó en los 160 sitios de muestreo y fue llevada a cabo en dos períodos (Fig. 1). El primer periodo fue de noviembre 2013 a enero 2014, el segundo periodo se realizó de noviembre 2018 a enero 2019, en los mismos conglomerados. Las muestras se colectaron con base en el protocolo estandarizado seguido por Ángeles-Pérez *et al.* (2015). En laboratorio, las muestras se secaron a 70 °C, se registró su peso y se molieron para determinar la concentración de carbono por el método de digestión seca a 900 °C, en el determinador automático de C (TOC SSM 5050^a Shimadzu). El contenido de carbono fue obtenido por multiplicar la concentración de C por la biomasa de cada muestra (López-Merlín *et al.*, 2016).

Variables predictoras

Se utilizó la información de doce variables ambientales predictoras obtenidas para los años 2013 y 2018. Estas variables se clasificaron en tres categorías (Tabla 1): topográficas (elevación, pendiente, exposición), de estructura espacial (coordenadas geográficas X y Y) y de estructura de dosel (edad del rodal, altura dominante, área basal, apertura del dosel, riqueza de especies, índice de Shannon-Wiener, cobertura arbórea). La información fue proyectada en la misma referencia geográfica (WGS84 UTM Zona 14), remuestreada a una misma resolución espacial (1 m) y transformada a formato *raster* mediante el software ArcGIS 10.5 (Environmental System Research Institute, ESRI Inc., Redlands, CA).

Las variables topográficas y de estructura espacial se derivaron de un modelo de elevación digital con resolución espacial de 1 m, generado para toda el área de estudio por Soriano-Luna *et al.* (2018), a partir de datos derivados de métricas LiDAR (*Light Detection and Ranging*). La información utilizada para caracterizar la estructura del dosel fue extraída de tres fuentes: métricas LiDAR,

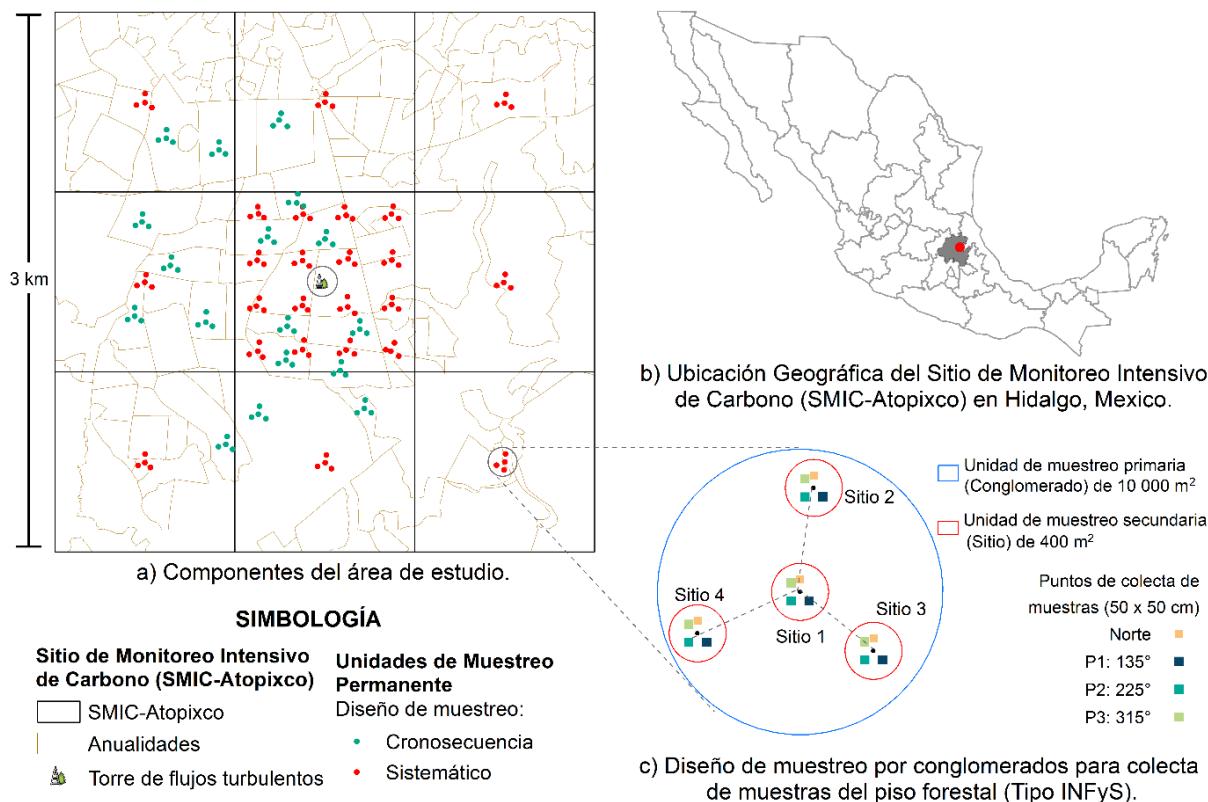


FIGURA 1. Componentes del Sitio de Monitoreo Intensivo de Carbono (SMIC-Atopixco): a) ubicación de las unidades de muestreo y anualidades; b) ubicación geográfica del área de estudio en Hidalgo, México; c) diseño del conglomerado tipo INFyS y esquema de muestreo para la colecta de información.

TABLA 1. Descripción de las variables ambientales utilizadas para predecir el contenido de carbono en el piso forestal.

| Categoría | Variable predictoria | Resolución espacial | Fuente | Referencia | Unidades |
|----------------------|--------------------------------|---------------------|--------------------------|-----------------------------------|-----------------------------|
| Topografía | Elevación | 1 m | LiDAR | Soriano-Luna <i>et al.</i> , 2018 | m s.n.m. |
| | Pendiente | 1 m | LiDAR | - | grados |
| | Exposición | 1 m | LiDAR | - | grados |
| Estructura espacial | Coordenadas X | 1 m | LiDAR | - | m |
| | Coordenadas Y | 1 m | LiDAR | - | m |
| Estructura del dosel | Edad del rodal | 1 m | Archivo vectorial | Soriano-Luna <i>et al.</i> , 2015 | años |
| | Altura dominante | 5 m | Inventario/LiDAR | Soriano-Luna <i>et al.</i> , 2018 | m |
| | Área basal | sitio* | Inventario | - | $m^2 \text{ ha}^{-1}$ |
| | Fracción de apertura del dosel | sitio | Fotografías hemisféricas | - | % |
| | Riqueza de especies | sitio | Inventario | Villegas-Macedo, 2019 | árboles sitio ⁻¹ |
| | Índice de Shannon-Wiener | 1 m | Archivo raster | Villegas-Macedo, 2019 | - |
| | Cobertura arbórea | 5 m | LiDAR | Soriano-Luna <i>et al.</i> , 2018 | % |

*Sitio: unidad de muestreo secundaria con una superficie de 400 m².



inventarios dasométricos y fotografías hemisféricas. Los datos de altura dominante, área basal, apertura del dosel y cobertura arbórea se obtuvieron a partir de métricas LiDAR (Soriano-Luna *et al.*, 2018) y de dos inventarios dasométricos, realizados en 2013 y 2018, correspondientes al mismo periodo de colecta de muestras. Este conjunto de datos fue obtenido en campo mediante la metodología establecida por el INFyS (Conafor, 2012). La edad del rodal se definió como el tiempo transcurrido a partir del año en que se aplicó la corta de repoblación, la información de esta variable fue extraída de polígonos en formato vectorial (Soriano Luna *et al.*, 2015). Los datos de riqueza de especies (S) y la estimación espacialmente explícita del índice de Shannon-Wiener fue obtenida por Villegas-Macedo (2019) y convertida a la misma resolución espacial.

Análisis estadístico

Análisis exploratorio

El análisis estadístico fue realizado en R v.3.5.1. (R Development Core Team, 2018). Se realizó un análisis descriptivo del CC en cada capa del piso forestal mediante la estadística de resumen tradicional (media, mediana, mínimo, máximo, desviación estándar, coeficiente de variación, curtosis y asimetría). Se comprobó la normalidad de los datos mediante la prueba de Kolmogorov-Smirnov (Brown y Forsythe, 1974) y se utilizó el coeficiente de correlación de Spearman para detectar problemas de multicolinealidad entre variables predictoras, seleccionando para el entrenamiento de los modelos, aquellas variables que presentaron una correlación menor a 80% ($r < 0.8$).

Métodos de predicción espacial

Los métodos DSM más utilizados para predicción y modelación espacial pueden clasificarse de acuerdo con las técnicas estadísticas empleadas en: a) interpolación espacial, b) regresión espacial y c) *machine learning* (Minasny, McBratney, Malone y Wheeler, 2013; Chen *et al.*, 2019). Por lo anterior, se comparó el desempeño predictivo de cuatro métodos DSM, pertenecientes a cada una de las clases, para estimar la distribución espacial del CC por capa de

descomposición del piso forestal, en 2013 y 2018. Los métodos utilizados fueron: 1) *kriging* ordinario (interpolación espacial), 2) modelo lineal generalizado (regresión espacial), 3) modelo aditivo generalizado (regresión espacial) y 4) *random forest* (*machine learning*). La elección de estos métodos responde a tres aspectos: a) incluyen algoritmos que han demostrado ser eficientes en el modelado predictivo de carbono en el suelo, b) abarcan la mayoría de las rutinas estadísticas más aplicadas y representativas de las técnicas DSM y c) pueden manejar estructuras de dependencia espacial (Scull, Franklin, Chadwick y McArthur, 2003).

a) Método de interpolación espacial. *Kriging* ordinario (KO) es un método geoestadístico univariado que estima el valor de un atributo continuo a cualquier ubicación no muestreada (Goovaerts, 1999; Webster y Oliver, 2007). Se realizó un análisis estructural de la variable de interés, en el que se obtuvo el semivariograma empírico del CC de cada una de las capas del piso forestal y se ajustó a uno de los modelos teóricos disponibles (Bivand, Pebesma y Gómez-Rubio, 2008). El ajuste del semivariograma está en función de tres parámetros: rango (distancia a la cual la autocorrelación se desvanece), meseta (máxima variabilidad en ausencia de dependencia espacial) y *nugget* (variabilidad intrínseca de los datos) (Berndt y Haberlandt, 2018; Bhunia *et al.*, 2018). La relación *nugget*/meseta caracteriza la dependencia espacial de los valores, considerando los siguientes umbrales: a) < 25% (fuerte) b) entre 25% y 75% (moderada) c) > 75% (débil) (Cambardella *et al.*, 1994). Los paquetes utilizados para la generación de semivariogramas y la interpolación *kriging* fueron *geoR* (Ribeiro y Diggle, 2006), *gstat* (Pebesma, 2004) y *automap* (Hemstra, 2013).

b) Métodos de regresión espacial. El modelo lineal generalizado y el modelo aditivo generalizado (GLM y GAM respectivamente, por siglas en inglés) son métodos que incorporan la regresión semi paramétrica y son capaces de modelar relaciones no lineales entre las variables (Antúnez, Hernández-Díaz, Wehenkel y Clark-Tapia, 2017). La estructura general de los modelos GLM y GAM se muestra en la ecuación 1 y 2, respectivamente (Wood y Augustin, 2002; Wood, 2017; Toriz-Robles, Ramírez-

Guzmán, Fernández-Ordoñez, Soria-Ruiz y Ybarra Moncada, 2019).

$$g(\mu) = X\beta \quad (1)$$

$$g(\mu) = X\beta + f_1(x_1) + f_2(x_2) + \dots + f_k(x_k) \quad (2)$$

Donde:

$X\beta$ = predictor lineal

g = función conocida (función liga)

X = matriz diseño de variables predictoras

β = vector de parámetros

f_k = funciones de suavizamiento continuas

Los modelos de regresión fueron generados para 2013 y 2018, mediante los paquetes de R: “*stats*” (Bolar, 2019) para GLM y “*mgcv*” (Wood y Wood, 2019) para GAM. Se definió al CC como variable dependiente y se agregó (como variable factorial) la capa del piso forestal perteneciente, a fin de obtener las estimaciones para cada una de las capas evaluadas. Las variables predictoras fueron las mismas para ambos modelos. Para minimizar el efecto de autocorrelación espacial y sobreajuste, se incluyó la dependencia espacial dentro de la estructura del algoritmo, como una función de las coordenadas geográficas y se determinó, mediante el índice de Moran, la autocorrelación espacial de los residuales con el paquete “*spdep*” (Bivand, 2019). Por último, para controlar la complejidad de los modelos, se realizó la comprobación del cumplimiento de supuestos en residuales y se determinó el criterio de información de Akaike (AIC) y el Criterio de Información Bayesiano (BIC).

c) Método de Machine Learning. *Random forest* (RF) es un modelo no paramétrico de clasificación y regresión que se caracteriza, en comparación con los métodos anteriores, por su mayor flexibilidad en el cumplimiento de supuestos estadísticos clásicos (Breiman, 2001; Biau 2012; Biau y Scornet, 2016). Se generaron dos modelos RF para el conjunto de datos de 2013 y 2018 mediante los paquetes “*randomForest*” (Liaw y Wiener, 2002) y “*ModelMap*” (Freeman, Frescino y Moisen, 2018). La definición de

variables fue igual que en los modelos GAM y GLM. Para el entrenamiento de cada modelo, se desarrolló un análisis de sensibilidad extensivo sobre dos parámetros (*mtry* y *ntree*) contenidos en el algoritmo de RF, hasta obtener el mejor desempeño predictivo (Breiman, 2001).

Validación y desempeño de los modelos

El desempeño del modelo en la predicción fue evaluado a través del método de validación cruzada ($k = 10$). La validación cruzada proporciona una estructura para crear varias divisiones aleatorias de entrenamiento/prueba en el conjunto de datos, de esta forma se garantiza que cada dato se encuentre en el conjunto de entrenamiento y de prueba al menos una vez (Zeraatpisheh, Ayoubi, Jafari, Tajik y Finke, 2019). Por lo cual, del conjunto total de datos, 90% fueron aleatoriamente seleccionados para entrenamiento y 10% para validación del modelo; este proceso fue repetido 100 veces. Los paquetes de R utilizados para cada modelo fueron: “*gstat*” (Pebesma, 2004) para KO; “*boot*” (Canty y Ripley, 2019) para GLM; “*gamclass*” (Mairdonald, 2018) para GAM y “*ModelMap*” (Freeman *et al.*, 2018) para RF. Los criterios utilizados para evaluar el desempeño de estos modelos fueron: a) el porcentaje de varianza explicada, obtenida con el paquete “*modEvA*” (Barbosa, Brown, Jiménez-Valverde y Real, 2020); b) el coeficiente de determinación (R^2) (Ecuación 3); c) el error medio absoluto (MAE) (Ecuación 4) y d) el error cuadrático medio (RMSE) (Ecuación 5)

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (3)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (4)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (5)$$

Donde;

\hat{y}_i = i ésimo valor obtenido de la predicción



y_i = i ésimo valor observado

\bar{y}_i = valor medio de y_i

n = número de valores predichos u observados con $i = 1, 2, \dots, n$

Este procedimiento mejora la precisión en la estimación de la incertidumbre del modelo y previene que los análisis presenten problemas de sobreajuste (Beguin *et al.*, 2017).

Mapas de distribución espacial

Para cada capa y año evaluado, se generaron mapas de distribución espacial del CC, utilizando los modelos ajustados. Los paquetes utilizados fueron: “*gstat*” (Pebesma, 2004) para KO, “*raster*” (Hijmans, 2019) para GLM y GAM, “*ModelMap*” (Freeman *et al.*, 2018) para RF.

RESULTADOS

El carbono en el piso forestal fue mayor en la capa F, lo que sugiere un efecto del grado de descomposición del material

orgánico. Se observó una disminución en el CC en las dos las capas del piso forestal de 2013 a 2018 (Tabla 2). El valor medio total del CC fue de $12.19 \text{ Mg ha}^{-1} \pm 5.48 \text{ Mg ha}^{-1}$ para 2013 y de $5.76 \text{ Mg ha}^{-1} \pm 2.34 \text{ Mg ha}^{-1}$ para 2018. Los valores mínimos iguales a cero en ambas capas corresponden principalmente a zonas con suelo desnudo, que ocurren generalmente en sitios cercanos a caminos. La estimación del CC para ambos años reveló una variabilidad menor en la capa Ho ($CV \sim 36\%$) con respecto a la capa F ($CV \sim 52\%$), sin embargo, la variabilidad de las estimaciones en cada capa se mantuvo. Las variables que presentaron una relación lineal significativa con el CC fueron el área basal, la edad del rodal y la altura dominante (Tabla 3). Las únicas variables con multicolinealidad fueron la riqueza de especies (S) y el índice de Shannon (H). Dado que H fue estimado a partir de S (Villegas-Macedo, 2019), solo se seleccionó a la variable S para los análisis posteriores.

TABLA 2. Estadística descriptiva del contenido de carbono (Mg ha^{-1}), estratificado por año y capa del piso forestal.

| Año | Capa | Media | Mediana | Min | Max | DS | CV | Kur | Skew | KS _p |
|------|------|-------|---------|------|-------|------|-------|------|-------|-----------------|
| 2013 | Ho | 2.03 | 1.95 | 0.00 | 4.72 | 0.74 | 36.36 | 4.75 | 0.58 | 0.34* |
| | F | 10.16 | 9.37 | 0.00 | 26.33 | 5.34 | 52.53 | 3.27 | 0.54 | 0.61* |
| | Tot | 12.19 | 11.47 | 0.77 | 27.75 | 5.48 | 44.98 | 3.18 | 0.48 | 0.54* |
| 2018 | Ho | 1.85 | 1.78 | 0.28 | 5.23 | 0.68 | 36.85 | 6.63 | 1.07 | 0.33* |
| | F | 3.91 | 3.83 | 0.00 | 8.81 | 1.97 | 50.45 | 2.55 | 0.08 | 0.99* |
| | Tot | 5.76 | 5.65 | 0.44 | 11.80 | 2.34 | 40.54 | 2.65 | -0.06 | 0.99* |

Ho: hojarasca, F: fermentación, Tot: suma de Ho y F; Min: mínimo, Max: máximo, DS: desviación estándar; CV: coeficiente de variación (%); Kur: curtosis; Skew: asimetría y KS_p: Prueba de Kolmogórov-Smirnov (*nivel de significancia $\alpha > 0.05$).

TABLA 3. Coeficientes de correlación de Spearman (r^2) entre el contenido de carbono por capa del piso forestal y cada una de las variables predictoras, para 2013 y 2018.

| Año | Capa | X | Y | ELE | SLO | ASP | SA | BA | DH | S | H | TCO | CO |
|------|------|------|-------------------|-------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|--------------------|
| 2013 | Ho | 0.01 | 0.11 | 0.13 | -0.06 | 0.01 | ^b 0.21 | ^b 0.24 | ^a 0.19 | ^a 0.19 | ^a 0.20 | 0.13 | ^a -0.17 |
| | F | 0.07 | 0.12 | 0.12 | ^a 0.01 | ^a 0.04 | 0.33 | 0.37 | 0.35 | 0.16 | 0.11 | 0.20 | -0.01 |
| | Tot | 0.06 | 0.14 | 0.15 | 0.00 | 0.05 | ^b 0.36 | ^b 0.39 | ^b 0.36 | ^a 0.20 | 0.15 | ^b 0.23 | -0.03 |
| 2018 | Ho | 0.11 | ^a 0.16 | 0.05 | -0.09 | 0.06 | ^b 0.42 | ^b 0.36 | ^b 0.38 | ^b 0.23 | ^b 0.29 | 0.10 | 0.10 |
| | F | 0.12 | 0.00 | -0.02 | -0.09 | 0.05 | ^b 0.24 | ^b 0.32 | ^b 0.25 | ^a 0.20 | ^a 0.17 | ^a 0.19 | ^a -0.19 |
| | Tot | 0.14 | 0.04 | -0.02 | -0.09 | 0.04 | ^b 0.29 | ^b 0.35 | ^b 0.29 | ^b 0.23 | ^b 0.21 | 0.11 | -0.11 |

X, Y: Coordenadas UTM X, UTM Y; ELE: elevación; SLO: pendiente; ASP: exposición; SA: edad del rodal; BA: área basal; DH: altura dominante; S: riqueza de especies; H: índice de Shannon-Wiener; TCO: cobertura arbórea; CO: Apertura del dosel. Nivel de significancia: (a < 0.05), (b < 0.01).

Entrenamiento de los modelos

Kriging ordinario

Los modelos teóricos *Matern*, *Stein* y *Periódico* presentaron el mejor ajuste al semivariograma empírico omnidireccional (Tabla 4). Los datos mostraron una variación considerable a pequeña escala, principalmente para 2018, con excepción de las capas F y Tot en 2013. Este comportamiento espacial sugiere que el diseño y la intensidad del muestreo no fue suficientemente intensivo para revelar patrones espaciales del CC bajo este método. Por el contrario, la relación *nugget*/meseta mostró diferencias en el grado de dependencia espacial del CC en las tres capas del piso forestal.

Modelos GAM, GLM y RF

Se detectó que evaluar el CC por capas del piso forestal contribuyó significativamente en el ajuste de ambos modelos por año (Tabla 5). Para GLM, las variables con mayor contribución en las estimaciones fueron la elevación y el área basal ($\alpha < 0.001$), la altura dominante y la cobertura arbórea ($\alpha < 0.01$) y en menor medida, la estructura espacial ($\alpha < 0.05$). Para GAM, las variables con mayor efecto en la predicción de CC fueron la elevación, la edad del rodal, la cobertura arbórea y la estructura espacial ($\alpha < 0.001$), con grados de libertad estimados (EDF, por sus siglas en inglés) ligeramente superiores a 8.0 (Tabla 6).

Los residuales para GAM y GLM idealmente deben asumir una distribución idéntica e independiente, sin

embargo, se obtuvieron valores más cercanos a cero para GLM en ambos años. Para GAM, los residuales presentaron una ligera autocorrelación espacial, lo que puede indicar un ligero sobreajuste en los modelos. Respecto al entrenamiento de RF, en ambos años el error fue estabilizado mediante el crecimiento de 2000 árboles (*ntree*), mientras que el análisis de sensibilidad del modelo permitió definir *mtry* en nueve y siete variables por cada nodo, para 2013 y 2018, respectivamente (Tabla 6).

Desempeño de los modelos

Aunque hay marcadas diferencias en el desempeño de los modelos predictivos, la superioridad de RF, GAM y GLM sobre KO en la predicción del CC en el piso forestal fue muy notable, con valores de R^2 superiores a 0.70 (Tabla 7). En vista de que KO es un método univariado que fue generado de forma individual para cada una de las capas del piso forestal, su principal limitante es que los valores de MAE y RMSE no pueden ser comparados con el resto de los modelos. Sin embargo, los valores de R^2 dan una idea del bajo desempeño que KO presentó para modelar CC en el piso forestal, con valores de 0.3 para 2013 y entre 0.7 a 0.3 para 2018. Los valores observados vs predichos revelaron una sobreestimación del CC en todos los valores antes de la intersección, mientras que los valores localizados después de la intersección mostraron una subestimación del CC (Fig. 2).

TABLA 4. Parámetros de los modelos de semivariograma con mejor ajuste, estratificado por año y capa del piso forestal.

| Año | Capa | Ajuste de modelo teórico | Nugget | Meseta | Rango (m) | Kappa | Nugget/meseta |
|------|------|--------------------------|--------|--------|-----------|-------|---------------|
| 2013 | Ho | Matern | 0.39 | 0.56 | 137 | 1.9 | 0.69 |
| | F | Stein | 0.00 | 29.0 | 50 | 1.1 | 0.00 |
| | Tot | Stein | 0.00 | 31.0 | 53 | 0.5 | 0.00 |
| 2018 | Ho | Periódico | 0.37 | 0.41 | 58 | 0.0 | 0.90 |
| | F | Stein | 1.5 | 3.8 | 116 | 0.2 | 0.39 |
| | Tot | Stein | 1.5 | 5.4 | 91 | 0.2 | 0.27 |

Ho: hojarasca, F: fermentación, Tot: suma de Ho y F.



TABLA 5. Parámetros y significancia de los predictores en los modelos lineal generalizado (GLM) y aditivo generalizado (GAM), para predecir contenido de carbono en las capas del piso forestal, en 2013 y 2018.

| Ajuste del modelo | GLM | | | | | GAM | | | | |
|-------------------|-----------|----------|---------------------|----------|---------------------|----------|---------------------|----------|---------------------|--|
| | 2013 | | 2018 | | 2013 | | 2018 | | | |
| (X_i) | β_i | α | β_i | α | EDF | α | EDF | α | | |
| Capa | Ho | 2645.0 | 0.0323 ^a | -504.2 | 0.3765 | 2.048 | 0.0001 ^c | 3.915 | 0.0001 ^c | |
| | F | 2653.0 | 0.0317 ^a | -502.1 | 0.3748 | 10.149 | 0.0001 ^c | 1.851 | 0.0001 ^c | |
| | Tot | 2655.0 | 0.0316 ^a | -500.3 | 0.3785 | 12.177 | 0.0001 ^c | 5.765 | 0.0001 ^c | |
| | X | 0.0012 | 0.0266 ^a | -0.0004 | 0.1086 | 1.000 | 0.5407 | 9.000 | 0.0001 ^c | |
| | Y | -0.0015 | 0.0121 ^a | 0.0003 | 0.2479 | 2.186 | 0.2298 | 8.470 | 0.0005 ^c | |
| | ELE | 0.0454 | 0.0008 ^c | -0.0123 | 0.0374 ^a | 8.426 | 0.0001 ^c | 8.308 | 0.0001 ^c | |
| | SLO | 0.0495 | 0.1397 | -0.0320 | 0.0177 ^a | 1.821 | 0.0699 | 6.320 | 0.1293 | |
| | ASP | 0.0001 | 0.9560 | -0.0006 | 0.4951 | 1.000 | 0.1640 | 7.775 | 0.0903 | |
| | SA | 0.0160 | 0.2730 | -0.0039 | 0.5189 | 8.548 | 0.0001 ^c | 7.789 | 0.0001 ^c | |
| | BA | 0.1422 | 0.0001 ^c | 0.0394 | 0.0251 ^a | 3.162 | 0.0015 ^b | 8.483 | 0.0001 ^c | |
| | DH | 0.1324 | 0.0035 ^b | -0.0312 | 0.0234 ^a | 7.455 | 0.0575 | 1.000 | 0.9246 | |
| | S | -0.0478 | 0.6370 | -0.0021 | 0.9569 | 8.152 | 0.0439 ^a | 8.463 | 0.0009 ^c | |
| | TCO | -0.0387 | 0.0082 ^b | 0.0252 | 0.0001 ^c | 8.636 | 0.0001 ^c | 5.733 | 0.1048 | |
| | CO | 0.0175 | 0.6868 | 0.0382 | 0.0826 | 5.910 | 0.0093 ^b | 4.690 | 0.0002 ^c | |

β_i : Parámetro desconocido de cada predictor; α : nivel de significancia de los predictores ^a($\alpha \leq 0.05$), ^b($\alpha \leq 0.01$) y ^c($\alpha \leq 0.001$); EDF: grados de libertad estimados para GAM; X, Y: Coordenadas UTM X, UTM Y; ELE: elevación; SLO: pendiente; ASP: exposición; SA: edad del rodal; BA: área basal; DH: altura dominante; S: riqueza de especies; D: índice de Shannon-Wiener; TCO: cobertura arbórea, CO: apertura del dosel.

TABLA 6. Parámetros de ajuste de los modelos lineal generalizado (GLM), aditivo generalizado (GAM) y *random forest* (RF) para predecir contenido de carbono en el piso forestal.

| Año | GLM | | | | GAM | | | | RF | |
|------|--------|--------|----------|----------|--------|--------|----------|----------|-------|------|
| | AIC | BIC | I. Moran | α | AIC | BIC | I. Moran | α | ntree | mtry |
| 2013 | 2590.5 | 2652.2 | -0.05 | 0.6604 | 2498.2 | 2746.3 | -0.31 | 1.00 | 2000 | 9 |
| 2018 | 1773.7 | 1835.4 | 0.14 | 0.0028 | 1631.4 | 1960.8 | -0.24 | 1.00 | 2000 | 7 |

AIC: Criterio de Información de Akaike; BIC: Criterio de Información Bayesiano; I. Moran: Índice de autocorrelación espacial de Moran ($\alpha > 0.05$); ntree: Número de árboles; mtry: Número de variables por nodo.

TABLA 7. Estadísticas de validación cruzada de las tres técnicas empleados para modelar el contenido de carbono en el piso forestal.

| Año | Técnica | Método de estimación | Bondad de ajuste | | Error (Mg ha ⁻¹) | |
|------|------------------------|----------------------|------------------|------------------------|------------------------------|------|
| | | | R ² | Varianza explicada (%) | MAE | RMSE |
| 2013 | Interpolación espacial | KO (Capa Ho) | 0.34 | ND | 0.51 | 0.69 |
| | | KO (Capa F) | 0.34 | ND | 3.97 | 5.01 |
| | | KO (Capa Tot) | 0.31 | ND | 4.15 | 5.21 |
| | Regresión espacial | GLM | 0.75 | 84.1 | 2.96 | 4.21 |
| | | GAM | 0.77 | 89.4 | 3.20 | 4.02 |
| | Machine Learning | RF | 0.90 | 86.2 | 1.81 | 2.62 |
| 2018 | Interpolación espacial | KO (Capa Ho) | 0.07 | ND | 0.53 | 0.72 |
| | | KO (Capa F) | 0.29 | ND | 1.56 | 1.88 |
| | | KO (Capa Tot) | 0.31 | ND | 1.80 | 2.21 |
| | Regresión espacial | GLM | 0.72 | 86.6 | 1.25 | 1.73 |
| | | GAM | 0.76 | 92.7 | 1.25 | 1.58 |
| | Machine Learning | RF | 0.86 | 69.1 | 0.85 | 1.51 |

R²: coeficiente de determinación, MAE: error medio absoluto, RMSE: error cuadrático medio. ND: No definido para interpolación con *Kriging Ordinario* (KO); GLM: Modelo lineal generalizado; GAM: Modelo aditivo generalizado; RF: Modelo *Random Forest*.

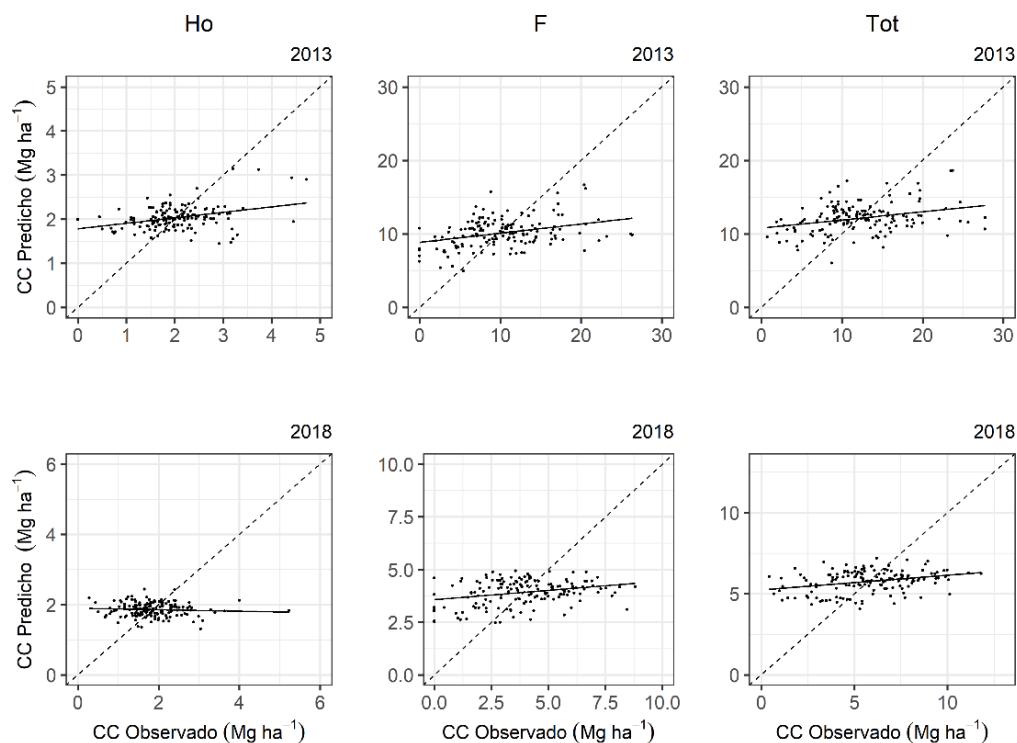


FIGURA 2. Valores observados *vs* predichos del conjunto de datos de validación cruzada ($k=10$) del modelo *Kriging ordinario*, para predecir el contenido de carbono en las capas del piso forestal: Ho (hojarasca), F (fermentación) y Tot (Totales), en 2013 y 2018. La línea negra es de regresión y la línea punteada es la relación 1:1



Al comparar las técnicas de regresión y ML, los resultados indicaron que RF fue el método con mejor desempeño para modelar la distribución del CC en el piso forestal, con valores de R^2 de 0.90 para 2013 y 0.86 para 2018 (Tabla 7). GAM y GLM mostraron una mayor varianza explicada que RF, sin embargo, considerando la ligera dependencia espacial en los residuales, puede existir un ligero sobreajuste (Fig. 3).

Mapas de CC a partir de las técnicas de predicción

Los mapas de distribución espacial del CC en cada una de las capas del piso forestal se muestran en la figura 4 para el

año 2013, y en la figura 5 para el año 2018. Las regiones en blanco fueron excluidas en la modelación por ser asentamientos humanos. Debido a que el CC varío a distancias muy cortas, la interpolación espacial fue generada a partir de malla fina de pixeles (1 m de resolución) que cubre toda el área de estudio (3 km \times 3 km). Para facilitar la comparación en las predicciones entre años, capas y modelos, las predicciones espaciales fueron reclasificadas en intervalos geométricos. El CC exhibió una notable variación espacial dentro del SMIC-Atopixco. Para ambos años y entre los cuatro métodos evaluados, se observó una diferencia considerable entre los patrones de distribución espacial del CC en las capas Ho y F.

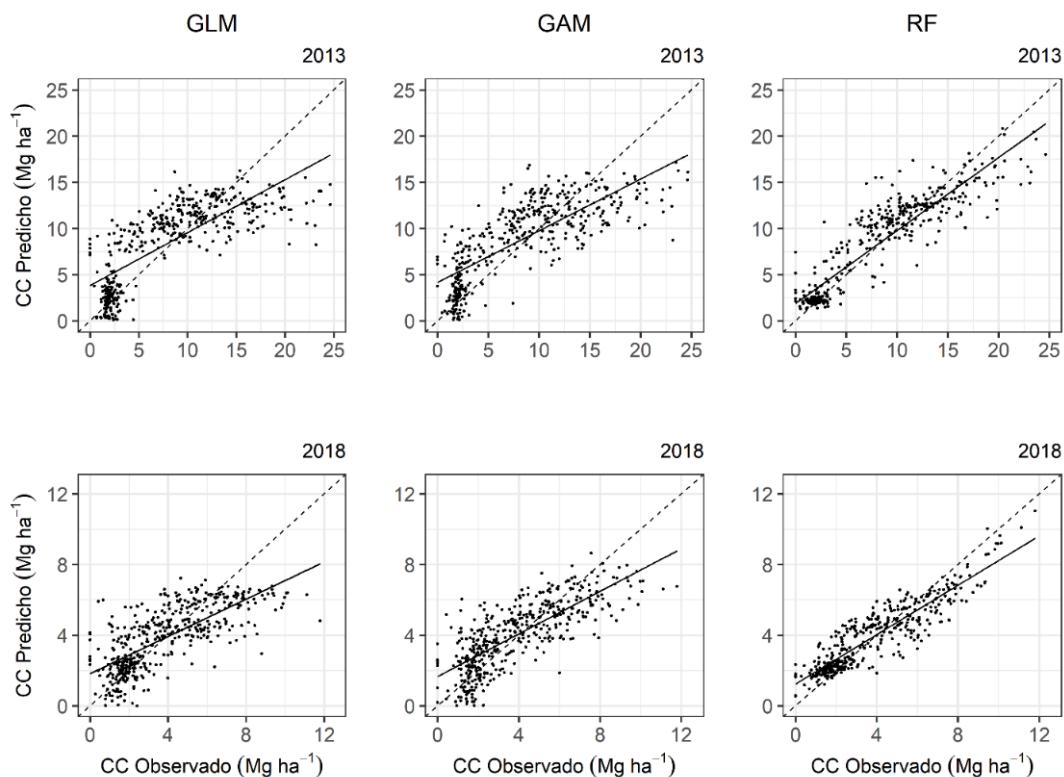


FIGURA 3. Valores observados vs predichos del conjunto de datos de validación cruzada ($k=10$) de los modelos lineal generalizado (GLM), aditivo generalizado (GAM) y *random forest* (RF), para predecir el contenido de carbono en el piso forestal, en 2013 y 2018. La línea negra es de regresión y la línea punteada es la relación 1:1.

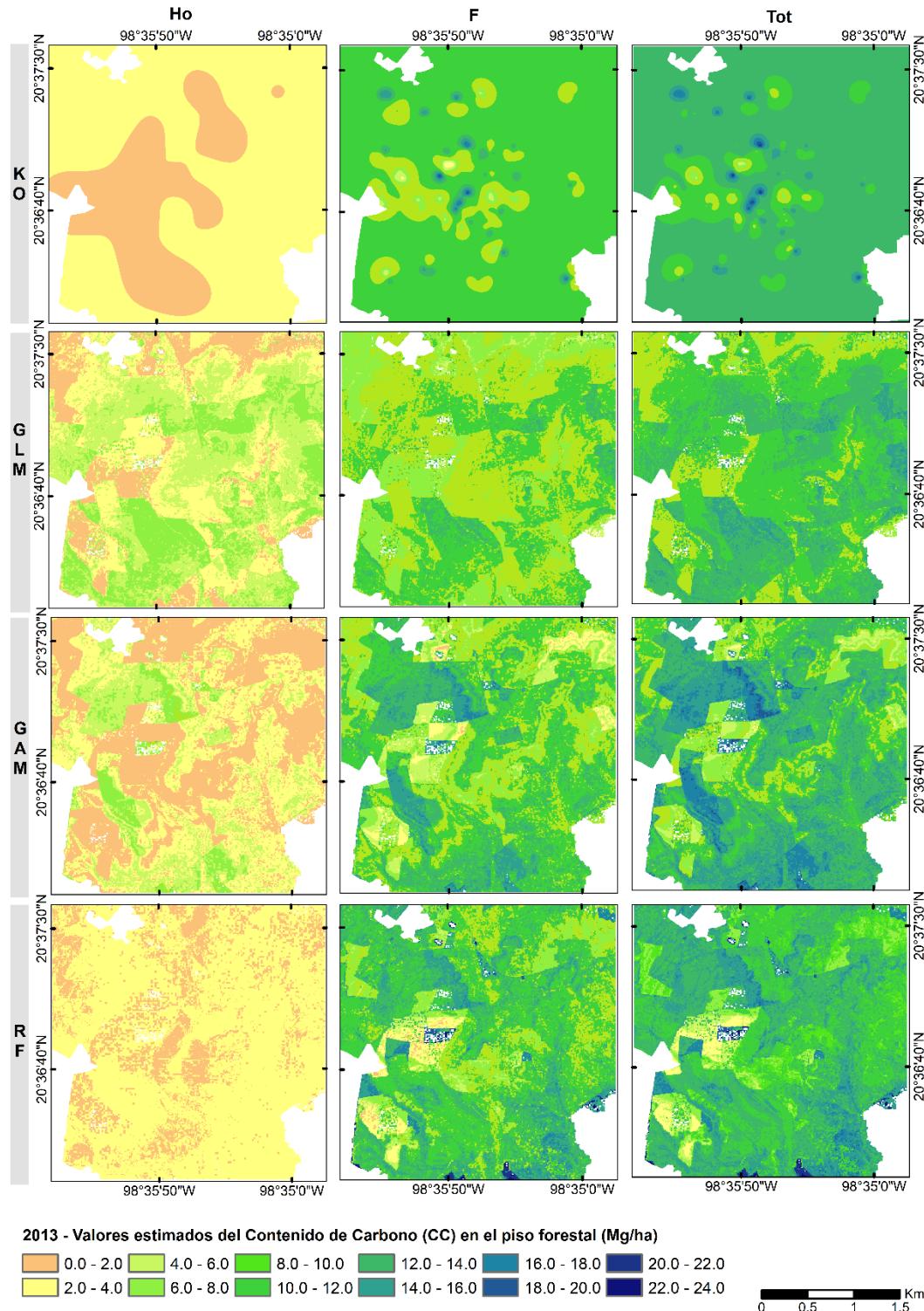


FIGURA 4. Predicciones espaciales para cada modelo espacial del CC en el piso forestal. Año 2013. Capas del piso forestal: Ho: hojarasca, F: fermentación, Tot: Suma de Ho y F. Modelos espaciales: *Kriging Ordinario* (KO), *Modelo Lineal Generalizado* (GLM), *Modelo Aditivo Generalizado* (GAM) y RF (*Random Forest*). Color blanco: Asentamientos humanos (excluidos de análisis).

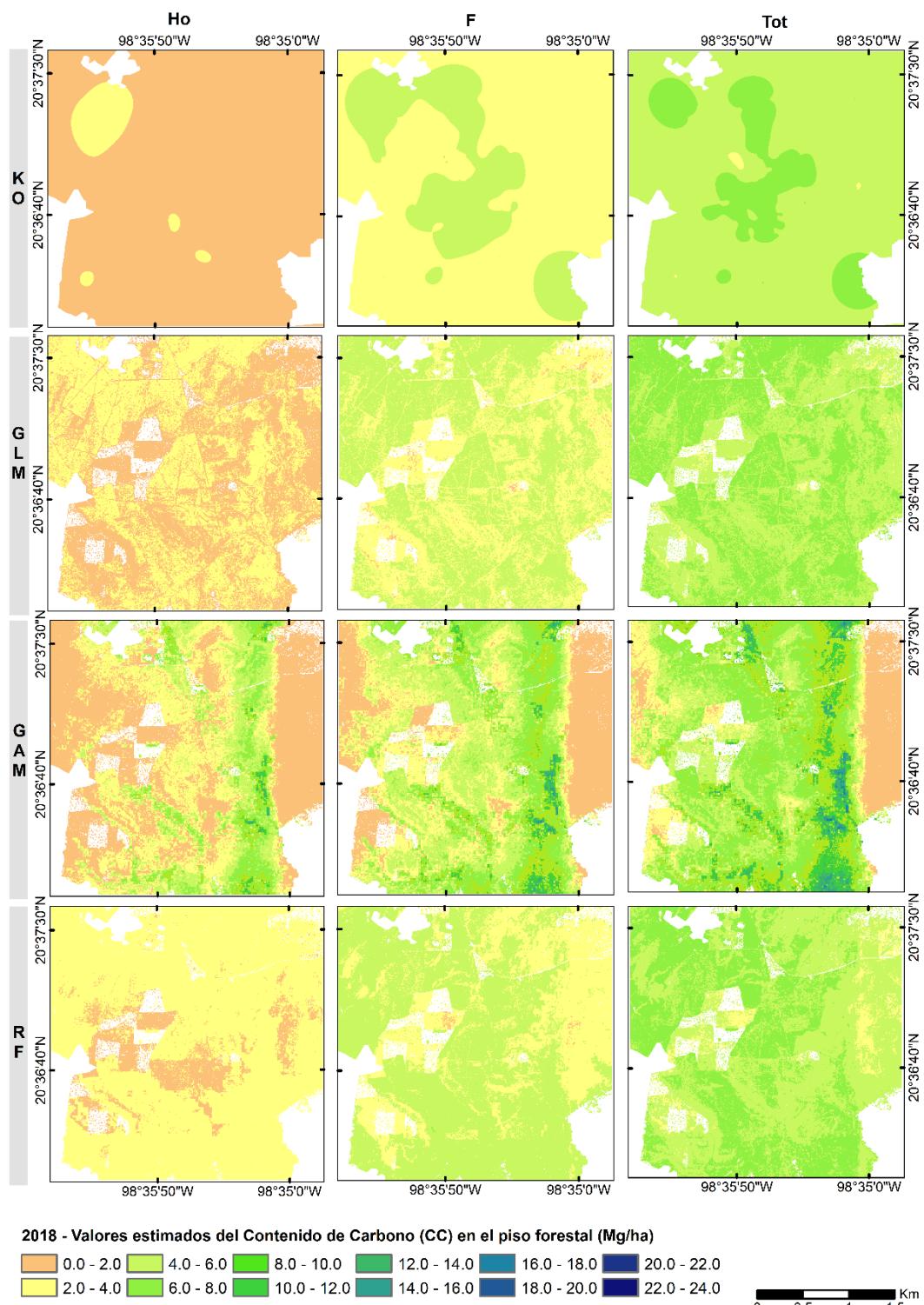


FIGURA 5. Predicciones espaciales para cada modelo espacial del CC en el piso forestal. Año 2018. Capas del piso forestal: Ho: hojarasca, F: fermentación, Tot: Suma de Ho y F. Modelos espaciales: Kriging Ordinario (KO), Modelo Lineal Generalizado (GLM), Modelo Aditivo Generalizado (GAM) y RF (Random Forest). Color blanco: Asentamientos humanos (excluidos de análisis).

La interpolación con KO exhibió patrones circulares alrededor de los puntos de muestreo y presentó clústeres en áreas con baja densidad de muestreo, por lo que se consideró una representación poco realista de la distribución espacial del CC en las diferentes capas del piso forestal. Las predicciones espaciales de GAM y GLM fueron notablemente diferentes entre ellos y para ambos años, los modelos sobreestimaron el CC con valores superiores al máximo valor observado en el análisis exploratorio, sobre todo en la región noreste de la zona de estudio, sin embargo, RF logró reducir este tipo de clústeres espaciales. GAM presentó un mejor desempeño para modelar el CC en el año 2013 y GLM en el año 2018. Las predicciones espaciales de RF, para ambos años fueron muy cercanas a los valores mínimos y máximos obtenidos en el análisis exploratorio, por lo cual, se consideró como la representación espacial más aproximada a la realidad. Para 2013, el CC total (Tot) varió entre 0 Mg ha⁻¹ y 24 Mg ha⁻¹, con un RMSE estimado de 2.62 Mg ha⁻¹; para 2018 el CC total varió entre 0 Mg ha⁻¹ y 12 Mg ha⁻¹, con un RMSE estimado de 1.51 Mg ha⁻¹. Para ambos años, la capa F presentó mayor heterogeneidad espacial, por lo que fue mejor representada por RF, en comparación con la capa Ho.

DISCUSIÓN

Selección del mejor método

La selección de un modelo apropiado es un aspecto central en la estimación de almacenes de carbono en los ecosistemas forestales (Zhang *et al.*, 2014). El empleo de diferentes técnicas como la interpolación espacial, modelos de regresión y ML permitió identificar y comparar los requerimientos estadísticos, el desempeño y las principales limitantes de cada modelo para estimar la distribución espacial del CC en el piso forestal. Aunado a lo anterior, el emplear dos conjuntos de datos de diferentes años reflejó una aproximación del potencial de aplicación de cada modelo para estimaciones futuras. Los resultados mostraron que la interpolación espacial con KO subestima en gran medida el CC en cada una de las capas del piso forestal y consecuentemente, no representa correctamente

la distribución espacial del CC. Por su parte, los modelos de regresión presentaron un desempeño aceptable, aunque ligeramente sesgado por el incumplimiento del supuesto estadístico de dependencia, que incurrieron en un sobreajuste y/o sobreestimaciones substanciales en el CC. Con respecto a la técnica de ML, el alto ajuste obtenido por RF para ambos años ($R^2 > 0.85$) sugiere que esta técnica tiene un gran potencial para mejorar las estimaciones de carbono en el piso forestal, tanto a escalas locales como regionales. Esto se atribuye a tres ventajas sobresalientes de RF con respecto al resto de los modelos evaluados: a) es un método robusto con mayor flexibilidad en los supuestos estadísticos clásicos, b) mejor resistencia al sobreajuste, c) no requiere la preselección de variables, por efectos de multicolinealidad (Breiman, 2001; Beguin *et al.*, 2017).

La utilización de KO permitió explorar el comportamiento espacial del CC dentro del área de estudio mediante el análisis estructural y, a pesar de haber obtenido valores de R^2 relativamente bajos ($R^2 < 0.35$), estos fueron similares a los observados en otros estudios sobre estimaciones de reservas de carbono en las capas orgánicas del suelo (Dai *et al.*, 2018; Yao *et al.*, 2019). El bajo desempeño de KO se atribuye a tres aspectos: a) las características intrínsecas del modelo, b) la alta variabilidad del CC a pequeñas distancias y c) a la densidad y distribución espacial de los puntos de muestreo. Por otra parte, GLM y GAM no presentan discrepancias substanciales entre ellos (Antúnez *et al.*, 2017), sin embargo, la mayor explicación de varianza obtenida para GAM se relaciona con la inclusión de relaciones no lineales mediante funciones suavizadas (sugeridas en el análisis exploratorio y en el entrenamiento de los modelos).

El ajuste del RF fue similar al obtenido por Wiesmeier *et al.* (2014) y Deng *et al.* (2018), quienes observaron un alto desempeño de RF en predecir carbono en las capas orgánicas del suelo, en comparación con otras técnicas DSM. Lo que permite sugerir que RF es un algoritmo adecuado para estimaciones aceptables de C en el piso forestal, a diferentes escalas espaciales.



Fuentes de incertidumbre

El desempeño de los modelos aplicados puede ser altamente variable y como lo sugieren Zeraatpisheh *et al.* (2017, 2019), la precisión en las predicciones estará en función de cuatro aspectos clave: a) la complejidad inherente del atributo de interés; b) las características de los datos de entrada; c) el esquema de muestreo y d) la selección adecuada de los análisis estadísticos y algoritmos requeridos para la predicción espacial.

Una de las principales limitantes para la aplicación de los métodos DSM es la disponibilidad de datos ($n = 40$, con 160 sitios de muestreo), la cual puede conducir a estimaciones imprecisas con cierto grado de sub o sobreestimación, a pesar del substancial esfuerzo para asegurar datos de alta calidad en las mediciones de campo. Metodológicamente, los modelos comparados son flexibles para modelar un atributo sin requerir forzosamente un esquema de muestreo específico, en cambio su desempeño se verá influenciado por la capacidad de estos para interpolar bajo diversas distribuciones de puntos de muestreo. De acuerdo con Carvalho Gomes *et al.* (2019), una disponibilidad limitada de mediciones en campo y baja representatividad de puntos de muestreo en el SMIC-Atopixco representaron una fuente importante de error en los métodos empleados, principalmente para KO. No obstante, lo anterior se compensó mediante la combinación de datos auxiliares derivados de distintas fuentes. La utilización de información LiDAR con alta resolución redujo el esfuerzo de muestreo, mejoró la capacidad predictiva de los modelos y, consecuentemente, permitió estimaciones más precisas del CC. Estos resultados son consistentes con algunas evaluaciones similares que han comparado el desempeño de las mismas técnicas DSM para predecir la variabilidad espacial CC en las capas orgánicas del suelo (Zhang *et al.*, 2014; Beguin *et al.*, 2017), quienes obtuvieron un mejor desempeño de RF por incorporar información armonizada de variables auxiliares.

El incumplimiento de supuestos en los modelos de regresión también representa una importante fuente de error. El CC es una variable espacialmente dependiente y puede tener algún grado de redundancia con la información

contenida en las covariables ambientales, por lo que frecuentemente violan el supuesto de independencia en residuales. Cuando la independencia es asumida incorrectamente, las predicciones e inferencia estadística puede ser inexacta o mal interpretada (Wood, 2017). En este caso, la inclusión de la estructura espacial dentro del algoritmo de GAM y GLM permitió reducir significativamente el efecto de la dependencia espacial.

Implicaciones para la estimación del CC

Estos resultados tienen una serie de implicaciones importantes para futuras modelaciones con estas técnicas DSM. Bajo las condiciones de este estudio, la utilización de semivariogramas permitió identificar el grado de dependencia espacial, la distancia a la cual el CC varía dentro del área de estudio y las diferencias entre capas del piso forestal. Por consiguiente, el conocimiento del comportamiento espacial del CC sugirió incluir la estructura espacial en los métodos de regresión y RF, lo que permitió estimaciones más robustas para estas técnicas.

Asimismo, el ajuste en los modelos de regresión y *machine learning* fue alto mediante la evaluación del CC por capas de descomposición del piso forestal. Algunas evaluaciones espaciales que han clasificado el piso forestal en capas, de acuerdo con el grado de descomposición, han enfatizado en este aspecto, al obtener mejores evaluaciones, en términos de precisión, de las reservas de carbono en este componente (Penne, Ahrends, Duerer y Böttcher, 2010; Jeyanny, Balasundram, Ahmad-Husni y Wan-Rasidah, 2016).

Una segunda implicación parte de la selección de covariables utilizadas en la predicción del CC. El desempeño de los modelos depende del tipo de variables predictoras utilizadas para representar las relaciones espaciales dentro del paisaje. En este caso, las covariables fueron seleccionadas a priori, en función de evaluaciones previas dentro del área de estudio (Ángeles-Pérez *et al.*, 2015; Soriano-Luna *et al.*, 2018; Villegas-Macedo, 2019). Finalmente, en este estudio se demostró que la combinación de datos provenientes de inventarios

dasométricos e información derivada de sensores remotos, así como la resolución a la cual se desplegó toda la información, permiten una mejora en el entrenamiento y validación de los modelos GAM, GLM y RF. Consecuentemente, mejoran su capacidad predictiva, lo cual es consistente con los resultados obtenidos por otros autores (Domke *et al.*, 2016; Cao *et al.*, 2019).

CONCLUSIONES

La comparación entre los cuatro métodos de DSM mostró que el método de interpolación espacial tuvo la menor precisión, atribuido a la estrategia de muestreo, la complejidad intrínseca del contenido de carbono (CC) y las características inherentes del modelo. Debido a las relaciones complejas entre el CC y las variables predictoras, los modelos GAM y RF presentaron los mejores desempeños, sin embargo, GAM y GLM mostraron un ligero sobreajuste por incumplimiento de supuestos estadísticos. El alto desempeño de RF se atribuye su mayor flexibilidad en el cumplimiento de supuestos estadísticos clásicos. La distribución espacial del CC, obtenida con *kriging* ordinario fue poco realista; los mapas derivados de GAM y GLM mostraron sobreestimaciones superiores al máximo valor observado en campo. RF presentó los mapas más aproximados a la realidad. En comparación con estimaciones tradicionales de carbono en el piso forestal, la aplicación de estos métodos permitió combinar información derivada de múltiples fuentes, lo que mejoró el desempeño predictivo de los modelos evaluados, redujo el esfuerzo de muestreo y puede ser potencialmente reproducible a diferentes escalas.

RECONOCIMIENTOS

Este estudio fue financiado por el Programa de Paisajes Sostenibles de la Agencia para el Desarrollo Internacional de los Estados Unidos de América, a través de la Oficina de Programas Internacionales del Servicio Forestal de la USDA, proyecto “Reducing Greenhouse Gas Emissions and Improving Forest Management in Mexico” (Acuerdo No. 12-IJ-11242306-033). También fue financiado por el Consejo Nacional de Ciencia y Tecnología de México, a

través del proyecto “El papel de los bosques bajo gestión forestal comunitaria en la mitigación del cambio climático” (PN 2017-6231).

REFERENCIAS

Anaya M., C. A. (2007). *Dinámica del C y N en el mantillo de un bosque tropical caducifolio en Jalisco, México*. Tesis de doctorado, Universidad Nacional Autónoma de México, México.

Ángeles-Pérez, G., Méndez-López, B., Valdez-Lazalde, R., Plascencia-Escalante, F. O., de los Santos-Posadas, H. M., Chávez-Aguilar, G., Ortiz-Reyes, A. D., Soriano L., M. A., Zaragoza-Castañeda, Z., Ventura P., E., & Martínez-López, A. (2015). *Estudio de Caso del Sitio de Monitoreo Intensivo del Carbono en Hidalgo*. Colegio de Postgraduados, Montecillo, México.

Antúnez, P., Hernández-Díaz, J. C., Wehenkel, C., & Clark-Tapia, R. (2017). Generalized models: an application to identify environmental variables that significantly affect the abundance of three tree species. *Forests*, 8(3), 1-14. doi: 10.3390/f8030059

Barbosa, A. M., Brown, J. A., Jiménez-Valverde A., & Real, M. (2020). R package “modEva”: Model Evaluation and Analysis. Recuperado de: <https://cran.r-project.org/web/packages/modEva/index.html>

Beguin, J., Fuglstad, G. A., Mansuy, N., & Paré, D. (2017). Predicting soil properties in the Canadian boreal forest with limited data: Comparison of spatial and non-spatial statistical approaches. *Geoderma*, 306, 195-200. doi: 10.1016/j.geoderma.2017.06.016

Berndt, C., & Haberlandt, U. (2018). Spatial interpolation of climate variables in Northern Germany-Influence of temporal resolution and network density. *Journal of Hydrology: Regional Studies*, 15, 184-202. doi: 10.1016/j.ejrh.2018.02.002

Bhunia, G. S., Shit, P. K., & Maiti, R. (2018). Comparison of GIS-based interpolation methods for spatial distribution of soil organic carbon (SOC). *Journal of the Saudi Society of Agricultural Sciences*, 17(2), 114-126. doi: 10.1016/j.jssas.2016.02.001

Biau, G. (2012). Analysis of a Random Forest. *Journal of Machine Learning Research*, 13, 1063-1095.

Biau, G., & Scornet, E. (2016). A random forest guided tour. *TEST*, 25, 197-227. doi: 10.1007/s11749-016-0481-7

Bivand, R. S. (2019). R Package “spdep”: Spatial dependence. Recuperado de: <https://cran.r-project.org/web/packages/spdep/index.html>

Bivand, R. S., Pebesma, E. J., & Gómez-Rubio, V. (2008). *Applied spatial data analysis in R*. New York: Springer.

Bolar, K. (2019). Package STAT - Interactive document for working with basic statistical analysis. Recuperado de: <https://cran.r-project.org/web/packages/STAT/STAT.pdf>



Breiman, L., 2001. Random Forest. *Machine Learning*, 45, 5-32.

Brown, M. B., & Forsythe, A. B. (1974). Robust test for the equality of variances. *Journal of American Statistical Association*, 69, 364-367.

Cambardella, C. A., Moorman, T. B., Novak, J. M., Parkin, T. B., Karlen, D. L., Turco, R. F., & Konopka, A. E. (1994). Field-Scale Variability of Soil Properties in Central Iowa Soils. *Soil Sciences Society of America Journal*, 58(5), 1501-1511. doi: 10.2136/sssaj1994.03615995005800050033x

Canty, A., & Ripley, B. (2019). R Package "boot": Bootstrap Functions. Recuperado de <https://cran.r-project.org/web/packages/boot/index.html>

Cao, B., Domke, G. M., Russell, M. B., Walters, B. F. (2019). Spatial modeling of litter and soil carbon stocks on forest land in the conterminous United States. *Science of the Total Environment*, 654, 94-106. doi: 10.1016/j.scitotenv.2018.10.359

Carvalho Gomes, L., Moniz F., R., de Souza, E., Vieira V., G., Schaefer, C. E. G. R., Fernandes F., E. I. (2019). Modelling and mapping soil organic carbon stocks in Brazil. *Geoderma*, 340, 337-350. doi: 10.1016/j.geoderma.2019.01.007

Chen, L., Chunying, R., Li, L., Wang, Y., Zhang, B., Wang, Z., & Li, L. (2019). A comparative assessment of Geostatistical, Machine Learning, and Hybrid Approaches for Mapping Topsoil Organic Carbon Content. *ISPRS International Journal of Geoinformation*, 8(4), 1-18. doi: 10.3390/ijgi8040174

Comisión Nacional Forestal [Conafor]. (2012). *Manual y procedimientos para el remuestreo de campo. Re-muestreo 2012*. México: Secretaría de Medio Ambiente y Recursos Naturales.

Dai, W., Fu, W., Jiang, P., Zhao, K., Li, Y., & Tao, J. (2018). Spatial pattern of carbon stocks in forest ecosystem of a typical subtropical region of southeastern China. *Forest Ecology and Management*, 409, 288-297. doi: 10.1016/j.foreco.2017.11.036

Deng, X., Chen, X., Ma, W., Ren, Z., Zhang, M., Grieneisen, M.L., Long, W., Ni, Z., Zhan, Y., & Lv, X. (2018). Baseline map of organic carbon stock in farmland topsoil in East China. *Agriculture, Ecosystem and Environment*, 254, 213-223. doi: 10.1016/j.agee.2017.11.022

Domke, G. M., Perry, C. H., Walters, B. F., Woodall, C. W., Russell, M. B., & Smith, J. E. (2016). Estimating litter carbon stocks on forest land in the United States. *Science of the Total Environmental*, 557-558, 469-478. doi: 10.1016/j.scitotenv.2016.03.090

García, E. (1981). *Modificaciones al sistema de clasificación climática de Köppen*. Serie Libros. México: Universidad Nacional Autónoma de México - Instituto de Geografía.

Goovaerts, P. (1999). Geostatistics in soil science: state-of-the-art and perspectives. *Geoderma*, 89, 1-45. doi: 10.1016/S0016-7061(98)00078-0

Grunwald, S. (2009). Multi-criteria characterization of recent digital soil mapping and modelling approaches. *Geoderma*, 152, 195-207. doi: 10.1016/j.geoderma.2009.06.003

Freeman, E. A., Frescino, T. S., & Moisen, G. G. (2018). *ModelMap: an R Package for Model Creation and Map Production*. Recuperado de <https://cran.r-project.org/web/packages/ModelMap/vignettes/VMModelMap.pdf>

Hiemstra, P. (2013). R Package "automap": Automatic interpolation package. Recuperado de <https://cran.r-project.org/web/packages/automap/index.html>

Hijmans, R. J. (2019). R Package "raster": Geographic Data Analysis and Modelling. Recuperado de <https://cran.r-project.org/web/packages/raster/index.html>

Hollinger, D. (2008). Defining a Landscape-Scale Monitoring Tier for the North American Carbon Program. En: C. Hoover (Ed.), *Field Measurements for Forest Carbon Monitoring* (pp. 3-16). USA: Springer-NY.

Intergovernmental Panel on Climate Change [IPCC]. (2006). *IPCC Guidelines for National Greenhouse Gas Inventories. Institute for Global Environmental Strategies*. Japan. Recuperado de https://www.ipcc-nggip.iges.or.jp/support/Primer_2006GLs.pdf

Jaramillo, V. J., Kauffman, J. B., Rentería-Rodríguez, L., Cummings, D. L., & Ellingson, L. J. (2003). Biomass, Carbon, and Nitrogen Pools in Mexican Tropical Dry Forest Landscapes. *Ecosystems*, 6, 609-629. doi: 10.1007/s10021-002-0195-4

Jeyanny, V., Balasundram, S. K., Ahmad-Husni, M. H., & Wan-Rasidah, K. (2016). Spatial variability of forest floor thickness for estimation of refined carbon stocks in a tropical montane forest. *Journal of Tropical Forest Science*, 28(3), 285-297.

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, 2(3), 18-22.

López-Escobar, N. F., Gómez-Guerrero, A., Velázquez-Martínez, A., Fierros-González, A. M., Castruita-Esparza, L. U., & Vera-Castillo, J. A. G. (2017). Reservoirs and nutrient dynamics in two stands of *Pinus montezumae* Lamb. in Tlaxcala, Mexico. *Revista Chapingo Serie Ciencias Forestales y del Ambiente*, 24(1), 115-129. doi: 10.5154/r.rchscfa.2017.09.055

López-Merlín, D., Maldonado, V., Wayson, C., Carrillo, O., Dupuy Rada, J. M., Ángeles-Pérez, G., Caamal Sosa, J. P., Méndez-López, B., Sánchez-Santos, G., Chávez-Angular, G., Johnson, K., Tamayo, M., & Puc, S. (2016). Capítulo III. Reservorios de carbono en

parcelas permanentes. En: *Protocolo para la estimación de la dinámica del carbono forestal en sitios de medición intensiva: un enfoque multi-escala*. Fortalecimiento REDD+ y Cooperación Sur-Sur, México, Noruega. 16-47 pp.

Maindonald, J. (2018). *Package “gamclass” – Functions and Data for a Course on Modern Regression and Classification*. Recuperado de: <https://cran.r-project.org/web/packages/gamclass/gamclass.pdf>

Martínez-Yrizar, A., Nuñez, S., Miranda, H., & Búrquez, A. (1999). Temporal and Spatial Variation of Litter Production in Sonoran Desert Communities. *Plant Ecology*, 145(1), 37-48. doi: 10.1023/A:1009896201047

Meersmans, J., De Ridder, F., Canters, F., De Baets, S., & Van Molle, M. (2008). A multiple regression approach to assess the spatial distribution of Soil Organic Carbon (SOC) at the regional scale (Flanders, Belgium). *Geoderma*, 143, 1-13. doi: 10.1016/j.geoderma.2007.08.025

Minasny, B., McBratney, A. B., Malone, B. P., & Wheeler, I. (2013). Digital Mapping of Soil Carbon. *Advances in Agronomy*, 118, 1-47. doi: 10.1016/B978-0-12-405942-9.00001-3

Pan, Y., Birdsey, R., Fang, J., Houghton, R., Kauppi, P. E., Kurz, W. A., Phillips, O. L., Shvidenko, A., Lewis, S. L., Canadell, J. G., Ciais, P., Jackson, R. B., Pacala, S. W., McGuire, A. D., Piao, S., Rautiainen, A., Sitch, S., & Hayes, D. (2011). A Large and Persistent Carbon Sink in the World's Forests. *Science*, 333, 988-993. doi: 10.1126/science.1201609

Pebesma, E. J. (2004). Multivariable geostatistics in S: the gstat package. *Computer & Geosciences*, 30(7), 683-691. doi: 10.1016/j.cageo.2004.03.012

Penne, C., Ahrends, B., Deurer, M., & Böttcher, J. (2010). The impact of the canopy structure on the spatial variability in forest floor carbon stocks. *Geoderma*, 158, 282-297. doi: 10.1016/j.geoderma.2010.05.007

Pérez-Suárez, M., Arredondo-Moreno, J. T., & Huber-Sannwald, E. (2012). Early stage of single and mixed leaf litter decomposition in semiarid forest pine-oak: the role of rainfall and microsite. *Biogeochemistry*, 108, 245-258. doi: 10.1007/s10533-011-9594-y

R Development Core Team (2018). *R: a Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Recuperado de: <http://www.r-project.org/>

Ribeiro, P. J., & Diggle, P. J. (2006). geoR: A package for Geostatistical Data Analysis. *R News*, 1(2), 15-18.

Santiago-García, W., de los Santos-Posadas, H. M., Ángeles-Pérez, G., Valdez-Lazalde, J. R., & Ramírez-Valverde, G. (2013). Sistema compatible de crecimiento y rendimiento para rodales coetáneos de *Pinus patula*. *Revista Fitotecnia Mexicana*, 36(2), 163-172.

Schlesinger, W. H., & Bernhardt, E. S. (2013). *Biogeochemistry: An Analysis of Global Change* (3ra ed.). Oxford, United Kingdom: Academic Press.

Scull, P., Flanklin, J., Chadwick, O. A., & McArthur, D. (2003). Predicting soil mapping: a review. *Progress in Physical Geography*, 27(2), 171-197. doi: 10.1191/030913303pp366ra

Soriano-Luna, M. A., Ángeles-Pérez, G., Guevara, M., Birdsey, R., Pan, Y., Vaquera-Huerta, H., Valdez-Lazalde, J. R., Johnson, K. D., & Vargas, R. (2018). Determinants of Above-Ground Biomass and Its Spatial Variability in a Temperate Forest Managed for Timber Production. *Forests*, 9(490), 1-20. doi: 10.3390/f9080490

Soriano-Luna, M. A., Ángeles-Pérez, G., Martínez-Trinidad, T., Plascencia-Escalante, F. O., & Razo-Zarate, R. (2015). Estimación de biomasa aérea por componente estructural en Zacualtipán, Hidalgo, México. *Agrociencia*, 49, 423-438.

Toríz-Robles, N., Ramírez-Guzmán, M. E., Fernández-Ordoñez, Y. M., Soria-Ruiz, J., & Ybarra Moncada, M. C. (2019). Comparación de modelos lineales y no lineales para estimar el riesgo de contaminación de suelos. *Agrociencia* 53(2), 269-283.

Villegas-Macedo, A. Y. (2019). *Caracterización y mapeo de hábitat en bosque templado bajo manejo maderable*. Tesis de Maestría en Ciencias. Colegio de Postgraduados. Montecillo, México.

Webster, R., & Oliver, M. A. (2007). *Geostatistics for Environmental Scientist* (2da ed.). England: John Wiley & Sons, Ltd.

Wiesmeier, M., Barthold, F., Spörlein, P., Geuß, U., Hangen, E., Reischl, A., Schilling, B., Angst, G., von Lützow, M., & Kögel-Knabner, I. (2014). Estimation of total organic carbon storage and its driving factors in soils of Bavaria (southeast Germany). *Geoderma Regional*, 1, 67-78. doi: 10.1016/j.geodrs.2014.09.001

Wood, S. (2017). *Generalized Additive Models. An Introduction with R* (2da ed.). United Kingdom: CRC Press. A Chapman & Hall Book.

Wood, S., & Augustin, N. H. (2002). GAMs with integrated model selection using penalized regression splines and application to environmental modelling. *Ecological Modelling*, 157(2-3), 157-177. doi: 10.1016/S0304-3800(02)00193-X

Wood, S., & Wood, M. S. (2019). *Package “mgcv” – Mixed GAM Computational Vehicle with Automatic Smoothness Estimation*. Recuperado de: <https://cran.r-project.org/web/packages/mgcv/mgcv.pdf>

Yao, X., Yu, K., Deng, Y., Zeng, Q., Lai, Z., & Liu, J. (2019). Spatial distribution of soil organic carbon stocks in Masson pine (*Pinus*



massoniana) forests in subtropical China. *Catena*, 178, 189-198. doi: 10.1016/j.catena.2019.03.004

Zeraatpisheh, M., Ayoubi, S., Jafari, A., Tajik, S., & Finke, P. (2017). Comparing the efficiency of digital and conventional soil mapping to predict soil types in a semi-arid region in Iran. *Geomorphology*, 285, 186-204. doi: 10.1016/j.geomorph.2017.02.015

Zeraatpisheh, M., Ayoubi, S., Jafari, A., Tajik, S., & Finke, P. (2019). Digital mapping of soil properties using multiple machine learning in a semiarid region, central Iran. *Geoderma*, 338, 445-452. doi: 10.1016/j.geoderma.2018.09.006

Zhang, J., Huang, S., Hogg, E. H., Lieffers, V., Qin, Y., & He, F. (2014). Estimating spatial variation in Alberta forest biomass from a combination of forest inventory and remote sensing data. *Biogeosciences*, 11(10), 2793-2808. doi: 10.5194/bg-11-2793-2014

Zhang, C., & McGrath, D. (2004). Geostatistical and GIS analyses on soil organic carbon concentrations in grassland of southeastern

Ireland from two different periods. *Geoderma*, 119, 261-275. doi: 10.1016/j.geoderma.2003.08.004

Manuscrito recibido el 24 de febrero de 2020

Aceptado el 11 de mayo de 2020

Publicado el 1 de julio de 2021

Este documento se debe citar como:

Pérez-Vázquez, Z. R., Ángeles-Pérez, G., Chávez-Vergara, B., Valdez-Lazalde, J. R., & Ramírez-Guzmán, M. E. (2021). Enfoque espacial para modelación de carbono en el mantillo de bosques bajo manejo forestal maderable. *Madera y Bosques*, 27(1), e2712122. doi: 10.21829/myb.2021.2712122



Madera y Bosques por Instituto de Ecología, A.C. se distribuye bajo una Licencia Creative Commons Atribución-NoComercialCompartirIgual 4.0 Internacional.