



Método para estimar la distribución espacial del contenido de carbono orgánico en el suelo de páramo con base en datos de sensores remotos.

Pablo Cesar Serrano Agudelo

Universidad Nacional de Colombia

Facultad de Ciencias Agrarias

Bogotá, Colombia

2023

Method to estimate the spatial distribution of organic carbon content in paramo soil based on remote sensing data

Pablo Cesar Serrano Agudelo

Trabajo de investigación presentado como requisito parcial para optar al título de:
Magister en Geomática

Director:

Luis Joel Martínez Martínez

Línea de Investigación:

Geoinformación para el uso sostenible de recursos naturales

Grupo de Investigación:

Gipso

Universidad Nacional de Colombia

Facultad de Ciencias Agrarias

Bogotá, Colombia

2023

Para mi papá, mamá y hermana por su constante apoyo y por siempre ser la luz en la oscuridad.

“Siempre lo intentaste. Siempre fallaste. No importa. Inténtalo otra vez. Falla de nuevo. Falla mejor”

Samuel Beckett

Declaración de obra original

Yo declaro lo siguiente:

He leído el Acuerdo 035 de 2003 del Consejo Académico de la Universidad Nacional. «Reglamento sobre propiedad intelectual» y la Normatividad Nacional relacionada al respeto de los derechos de autor. Esta disertación representa mi trabajo original, excepto donde he reconocido las ideas, las palabras, o materiales de otros autores.

Cuando se han presentado ideas o palabras de otros autores en esta disertación, he realizado su respectivo reconocimiento aplicando correctamente los esquemas de citas y referencias bibliográficas en el estilo requerido.

He obtenido el permiso del autor o editor para incluir cualquier material con derechos de autor (por ejemplo, tablas, figuras, instrumentos de encuesta o grandes porciones de texto).

Por último, he sometido esta disertación a la herramienta de integridad académica, definida por la universidad.

Pablo C. Jarama

Nombre Pablo Cesar Serrano Agudelo

Fecha 31/07/2023

Agradecimientos

A la Universidad Nacional de Colombia, Sede Bogotá por la formación académica y personal.

A mis profesores, en especial al profesor Luis Joel Martínez Martínez por su paciencia, tiempo, sus valiosas palabras y por guiarme en este maravilloso proceso.

A todos mi amigos y amigas que fueron soporte, me ayudaron en diferentes maneras con ideas y aportes o simplemente sacándome durante estos años una sonrisa con su apoyo y compañía.

De nuevo agradecimientos a mi familia por todo su apoyo incondicional durante toda mi formación académica y formación personal.

Resumen

En esta investigación se desarrolló un método para estimar la distribución espacial del carbono orgánico en el suelo de páramo basado en datos derivados de sensores remotos y utilizando algoritmo de aprendizaje automatizado. Para este fin se efectuaron análisis de correlaciones entre el contenido del carbono orgánico de 169 muestras a dos profundidades (0-15cm, 15-30cm) con las covariables derivadas de los sensores Sentinel 1, Sentinel 2, modelos digitales de elevación de Alos Palsar y datos de clima obtenidos de la plataforma WorldClim. Las covariables que mayor correlación tuvieron con el contenido de carbono orgánico del suelo (COS), fueron la temperatura, la altura, los índices derivados del modelo de elevación digital, los índices espectrales en especial el NDVI, y el índice VH de Sentinel 1. El mejor desempeño fue para los modelos desarrollados con random forest. Por último, se validó y documentó el método, permitiendo hacer una estimación de la distribución espacial del COS en los suelos de páramo, de gran utilidad para apoyar la toma de decisiones sobre uso y manejo de conservación de estos ecosistemas.

Palabras clave: (Carbono Orgánico del Suelo, Sensores Remotos, Aprendizaje Automatizado, Páramo).

Abstract

In this research, a method was developed to estimate the spatial distribution of organic carbon in páramo soil based on data derived from remote sensors and using machine learning algorithm. For this purpose, correlation analyzes were carried out between the organic carbon content of 169 samples at two depths (0-15cm, 15-30cm) with the covariates derived from the Sentinel 1, Sentinel 2 sensors, digital elevation models of Alos Palsar and climate data obtained from the WorldClim platform. The covariates that had the highest correlation with the soil organic carbon (SOC) content were temperature, height, indices derived from the digital elevation model, spectral indices, especially NDVI, and the Sentinel 1 VH index. The best performance was for the models developed with random forests. Finally, the method was validated and documented, allowing an estimate of the spatial distribution of SOC in páramo soils, very useful to support decision-making on the use and conservation management of these ecosystems.

Keywords: (Soil Organic Carbon, Remote Sensing, Machine Learning, Páramo).

Contenido

	Pág.
1. Introducción	2
2. Objetivos	5
2.1 Objetivo general	5
2.2 Objetivos específicos	5
3. Marco de referencia	7
3.1 Carbono orgánico del suelo	7
3.2 Sensores remotos para el COS	8
3.3 Métodos predictivos basados en aprendizaje automatizado	8
3.4 Avances en la investigación sobre la estimación de la distribución espacial del COS	9
4. Materiales y métodos	13
4.1 Zona de estudio.....	13
4.2 Metodología.....	14
4.2.1 Datos de carbono orgánico del suelo para entrenamiento y validación de los modelos	16
4.2.2 Variables calculadas a partir de la imagen Sentinel 2.....	16
4.2.3 Variables geomorfológicas calculadas a partir de un Modelo digital de elevación.....	19
4.2.4 Variables de clima utilizadas	21
4.2.5 Variables calculadas a partir de la imagen Sentinel 1	22
4.3 Selección de variables	23
4.4 Desarrollo de los modelos predictivos	23
4.5 Evaluación y validación modelos.....	25
4.6 Estimación de la distribución espacial del COS	26
5. Resultados y discusión.	26
5.1.1 Variables para la estimación del contenido del COS	26
5.1.2 Evaluación de los algoritmos bosques aleatorios y redes neuronales para estimación del COS.....	40
5.1.3 Validación del método	52
5.1.4 Estimación del método desarrollado	57
6. Conclusiones	63
Bibliografía	65

Lista de figuras

Pág.

Figura 1 Ubicación zona de estudio	13
Figura 2 <i>Diagrama de flujo metodología.</i>	15
Figura 3 <i>Diagrama de flujo de evaluación y validación de los modelos.</i>	26
Figura 4 <i>Distribución de los datos de COS de 0-15 cm.</i>	27
Figura 5 <i>Distribución espacial de los datos COS de 0-15 cm.</i>	28
Figura 6 <i>Distribución de los datos de COS de 15-30 cm.</i>	29
Figura 7 <i>Distribución espacial de los datos COS de 15-30 cm</i>	30
Figura 8 <i>Correlación entre las covariables predictoras</i>	33
Figura 9 <i>VIF de las covariables</i>	34
Figura 10 <i>Variación espacial de altura zona de estudio.</i>	35
Figura 11 <i>Distribución espacial del NDVI zona de estudio.</i>	37
Figura 12 <i>Mapa de distribución espacial de la PPT en la zona de estudio</i>	38
Figura 13 <i>Mapa de la distribución espacial VH zona de estudio.</i>	39
Figura 14 <i>Importancia relativa de las covariables para el modelo 4 con bosques aleatorios 0-15 cm.</i>	44
Figura 15 <i>Importancia covariables en el modelo 11 de bosques aleatorios.</i>	44
Figura 16 <i>Importancia covariables modelo 12 bosques aleatorios.</i>	45
Figura 17 <i>Importancia relativa de las covariables para el modelo 5 con bosques aleatorios 15-30 cm.</i>	50
Figura 18 <i>Importancia relativa de las covariables para el modelo 8 con bosques aleatorios 15-30 cm.</i>	50
Figura 19 <i>Importancia relativa de las covariables para el caso 5 con bosques aleatorios 15-30 cm.</i>	51
Figura 20 <i>Gráfico de dispersión de los valores predichos/observados para el caso 4 bosques aleatorios 0-15 cm.</i>	53
Figura 21 <i>Gráfico de dispersión de los valores predichos/observados para el caso 11 bosques aleatorios 0-15 cm.</i>	54
Figura 22 <i>Gráfico de dispersión de los valores predichos/observados para el caso 12 bosques aleatorios 0-15 cm.</i>	54
Figura 23 <i>Gráfico de dispersión de los valores predichos/observados para el caso 5 bosques aleatorios 15-30 cm.</i>	55

Figura 24 <i>Gráfico de dispersión de los valores predichos/observados para el caso 8 bosques aleatorios 15-30 cm.</i>	56
Figura 25 <i>Gráfico de dispersión de los valores predichos/observados para el caso 9 bosques aleatorios 15-30 cm.</i>	56
Figura 26 <i>Distribución espacial del COS predicho caso 4 con bosques aleatorios 0-15 cm.</i>	57
Figura 27 <i>Distribución espacial del COS predicho modelo 11 con bosques aleatorios 0-15 cm.</i>	58
Figura 28 <i>Distribución espacial del COS predicho modelo 12 con bosques aleatorios 0-15 cm.</i>	58
Figura 29 <i>Distribución espacial del COS predicho modelo 5 con bosques aleatorios 15-30 cm.</i>	59
Figura 30 <i>Distribución espacial del COS predicho modelo 8 con bosques aleatorios 15-30 cm.</i>	60
Figura 31 <i>Distribución espacial del COS predicho modelo 9 con bosques aleatorios 15-30 cm.</i>	60

Lista de tablas

Pág.

Tabla 1 <i>Bandas de imagen Sentinel 2</i>	17
Tabla 2 Índices espectrales calculados con base en imagen Sentinel 2.....	18
Tabla 3 <i>Especificaciones de imagen Sentinel 1</i>	22
Tabla 4 Modelos evaluados con diversas combinaciones de covariables.....	24
Tabla 5 <i>Estadísticas básico descriptivas del COS (t ha – 1)</i>	26
Tabla 6 <i>Estadísticas básica de las covariables en los sitios muestreados</i>	31
Tabla 7 <i>Correlación y significancia entre el COS y las covariables.(n=169)</i>	32
Tabla 8 <i>Desempeño de cada caso evaluado por bosques aleatorios y redes neuronales para la profundidad 0-15 cm</i>	43
Tabla 9 <i>Desempeño de cada modelo evaluado por bosques aleatorios y redes neuronales para la profundidad 15-30 cm</i>	48
Tabla 10 <i>Estadísticas COS predicho 0-15 cm</i>	53
Tabla 11 <i>Estadísticas COS predicho 15-30 cm</i>	55

1. Introducción

Los páramos son ecosistemas únicos, debido a la vegetación que los caracteriza, a las condiciones climáticas, a sus suelos y a la altitud sobre el nivel del mar (Van der Hammen, 2007); además, son estratégicos ya que brindan servicios ambientales como la capacidad de captura de carbono por parte de la vegetación y su almacenamiento en el suelo a través de la retención de materia orgánica (Cely et al., 2016).

Los suelos son capaces de retener alrededor de tres veces más carbono que la biomasa vegetal y animal sobre la tierra y el doble del carbono contenido en la atmósfera (Pla, 1994; Wang et al., 2009). En el suelo de páramo hay más acumulación de carbono debido a las bajas temperaturas y a la descomposición lenta de los residuos vegetales, (Ayala, 2014), lo cual es de gran relevancia en escenarios de mitigación y adaptación frente a los efectos negativos asociados al cambio climático, ya que ayuda a secuestrar más carbono en el suelo con respecto a otros ecosistemas (Zimmermann et al., 2010).

En el Protocolo de Kyoto se reconoce que las emisiones netas de gases de efecto invernadero pueden ser reducidas, ya sea mediante la disminución de la tasa a la cual se emiten a la atmósfera esos gases, o con el incremento de la tasa por la cual son retirados de ésta gracias a los sumideros (Organización de las Naciones Unidas para la Agricultura y la Alimentación [FAO], 2002), con el fin de mitigar los gases del efecto invernadero se creó un acuerdo entre países industrializados con países en vía de desarrollo para la compra de bonos de carbono. Cada bono de carbono corresponde a una tonelada de dióxido de carbono equivalente (ton CO₂eq.) que ha dejado de ser emitida a la atmósfera.

A partir del primer trimestre de 2016 se implementan en Colombia las negociaciones bajo la administración de la Bolsa Mercantil, con el objetivo de reducir 500.000 toneladas de CO₂. (Cuervo Barahona et al., 2016). Con base en lo anterior, el conocimiento del contenido carbono orgánico de los suelos de páramo y de su dinámica espacial y temporal, podría contribuir a destacar la función de este ecosistema como mitigador ante el efecto de cambio climático y ayudar a fundamentar la adopción de incentivos económicos para apoyar un mejor manejo de los recursos naturales en estos ecosistemas (Ward et al., 2015).

El interés en estimar la distribución espacial del carbono orgánico en el suelo (COS) se fundamenta en que es un indicador de la calidad de suelo y de su fertilidad, y por su incidencia en procesos ambientales como en los flujos de gases efecto invernadero y la erosión del suelo (Stolbovoy et al., 2007). Es por esto que surge la necesidad de optimizar los métodos de mapeo del COS, con conjuntos de datos consistentes capaces de proporcionar información confiable y oportuna para la estimación del contenido de dicho elemento (Croft et al., 2012). Con el avance tecnológico de los sensores remotos montados en plataformas satelitales o en aviones, que captan imágenes a intervalos de tiempo cada vez menores, con mejor resolución espacial y espectral (Valdez-Lazalde et al., 2017), las tendencias actuales están orientadas hacia la evaluación de técnicas de sensores remotos que sean rápidas, rentables y no destructivas, para la estimación del COS (Croft et al., 2012).

En las estimaciones globales del contenido de COS, hay dificultades debido a la alta variabilidad espacial, a las áreas ocupadas por distintos tipos de suelos, insuficientes datos de densidad aparente (necesaria para cuantificar la composición volumétrica) y los complejos efectos de la vegetación y de los cambios en el uso del suelo (Sandoval et al., 2003; Martínez et al., 2008).

De acuerdo con la FAO (2017), se requieren métodos que sean rentables y que puedan cubrir una gran variedad de suelos para estimar el contenido del COS, con el fin de efectuar el monitoreo de forma regular, eficiente y rápida. Por otra parte, es fundamental considerar la variación espacial y temporal del COS de manera que se puedan ver los cambios en el espacio y en el tiempo. Los métodos tradicionales de interpolación requieren un número alto de muestras que deben ser analizadas en el laboratorio, lo que aumenta los costos y dificulta la obtención de información oportuna. Gutiérrez et al., (2020) reportan un estudio como una primera aproximación del COS para los páramos en Colombia y consideran que se debe avanzar en estudios más detallados que sean relevantes para los diferentes páramos en Colombia

Los datos de sensores remotos y los algoritmos de aprendizaje automatizado surgen como estrategias con un gran potencial para efectuar modelos predictivos de la distribución espacial del COS (Zhou et al., 2020). Considerando lo mencionado anteriormente, se propone desarrollar un método basado en análisis espacial para estimar la distribución

espacial del COS en un páramo, integrando datos provenientes de sensores remotos y de los factores que influyen en el contenido de COS con algoritmos de aprendizaje automatizado.

Esta investigación hace un aporte al conocimiento para formular planes de ordenamiento territorial, definir medidas de mitigación y adaptación al cambio climático, planear usos sostenibles y la conservación de los suelos, mantenimiento las funciones de servicios ambientales.

2. Objetivos

2.1 Objetivo general

Desarrollar un método, basado en análisis espacial, para estimar la distribución del contenido de carbono orgánico del suelo en un páramo con datos climáticos y de sensores remotos.

2.2 Objetivos específicos

- Determinar las variables climáticas, de uso y cobertura del suelo, índices de vegetación y parámetros morfométricos, que correlacionan y explican el contenido de carbono orgánico del suelo en un páramo.
- Evaluar el desempeño de los algoritmos bosques aleatorios y redes neuronales artificiales para estimar el contenido de carbono orgánico del suelo en un páramo, con base en las covariables previamente seleccionadas.
- Estimar la distribución espacial el contenido de carbono orgánico del suelo de páramo con el método desarrollado.

3. Marco de referencia

3.1 Carbono orgánico del suelo

De acuerdo con la FAO (2017), el COS es una pequeña parte del ciclo global del carbono. El equilibrio del carbono en la tierra está en función de cuatro reservorios: los océanos, con una cantidad estimada de carbono de 38.000Pg (petagramo es igual a 10^{15} gramos), la atmósfera, con 750Pg; y el sistema terrestre, con 550Pg como biomasa-vegetación y 1.550Pg como COS. Los tres reservorios se encuentran en un equilibrio dinámico, con interacciones entre unos y otros. Surge un cuarto reservorio, el geológico, que tiene $65,5 \times 10^6$ Pg C, con sólo una pequeña porción de éste que se podría aproximar a 4000Pg correspondiente a combustible fósil. En el ecosistema terrestre, el mayor componente es el COS con 1.550Pg C, seguido por el CIS con 750-950Pg C (Burbano, 2018).

El COS es un componente importante del ciclo global del C, el cual ocupa el 69,8 % del C orgánico de la biosfera (FAO, 2001). De manera que, el suelo juega un papel destacado en el C ya que dependiendo de su uso y manejo puede actuar como fuente o reservorio de C (Lal, 1997). Que los suelos actúen como un reservorio o como un emisor de C depende de varios factores. Los cambios más fuertes en la reserva de COS en los primeros centímetros de profundidad obedecen a las modificaciones en la cobertura terrestre. Los cambios de uso del suelo por ejemplo de vegetación nativa a suelos para hacer agricultura, derivan en pérdidas de COS pero además generan emisiones de gases de efecto invernadero (GEI) que contribuyen al cambio climático (Burbano, 2018). Aunque no es la única covariable que influye, el cambio de temperatura también es un factor que puede en varias partes llevar a emisiones de GEI. La topografía y el tipo de suelo influyen directamente en el stock de C. Otros factores, como el contenido de humedad determinado por las condiciones climáticas, incide junto con la temperatura en la actividad microbiana y, por lo tanto, en la descomposición de los materiales orgánicos. La actividad microbiológica y los tipos de materia orgánica son factores que inciden en la mayor o menor

descomposición y mineralización del COS. Son varios los factores que caracterizan la concentración y la cantidad de COS.

3.2 Sensores remotos para el COS

Los sensores remotos, que se basan en el uso de radiación electromagnética para obtener información sobre un objeto o fenómeno sin contacto físico (Elachi y Van Zyl, 2006), es una fuente de datos con gran potencial para el desarrollo de métodos predictivos. La teledetección se ha identificado como un proceso eficiente para evaluar y mapear algunos atributos del suelo (Gholizadeh et al., 2018). Las imágenes de sensores remotos son una fuente de datos para el estudio cuantitativo y cualitativo de las propiedades del suelo. Se han aplicado para cartografiar propiedades del suelo como la humedad, el pH y fósforo total del suelo, entre otras propiedades (Zhou et al., 2020).

Los índices de vegetación se pueden derivar de imágenes de sensores remotos y permiten discriminar la vegetación y establecer algunas de sus características de forma cualitativa y cuantitativa (Bannari et al., 1995). Otra fuente de datos son los modelos digitales de elevación que permiten obtener parámetros morfo métricos cuantitativos para ser integrados con otro tipo de datos en los métodos predictivos de las propiedades del suelo, como el COS.

3.3 Métodos predictivos basados en aprendizaje automatizado

En la actualidad, los métodos predictivos basados en el aprendizaje automatizado son fundamentales para la estimación de diversos procesos de la naturaleza, entre ellos el COS.

Uno de los métodos más utilizados es el Random Forest (RF) o bosques aleatorios, el cual es una técnica de aprendizaje automatizado basado en árboles y aplicada a la clasificación y a la regresión (Breiman, 2001). El algoritmo de aprendizaje automatizado que emplea RF en regresiones tiene en su estructura árboles de clasificación y regresión (CART) (Breiman et al., 2017), que escoge las mejores variables y los puntos de división para reducir el criterio de error cuadrático y absoluto.

En RF se utilizan las divisiones o particiones de las variables predictivas (x_1, x_2, \dots, x_n) que generan regiones para predecir la variable dependiente (Y) con mayor precisión. Hay dos parámetros en el modelado de RF que deben ser definidos por el usuario: i) el número de árboles de regresión generados en el bosque y ii) el número de predictores seleccionados al azar en cada nodo. (Breiman, 2001).

Otro método utilizado en el aprendizaje automatizado es la Red Neuronal Artificial (RNA), que es un modelo de clasificación y regresión que imita el sistema nervioso. Consiste en elementos de procesamiento interconectados, neuronas que cooperan para producir la función de salida. Cada neurona genera su valor de salida utilizando un nodo de agregación y una función de activación para los valores de entrada, la ventaja más importante de red neuronal artificial está en solucionar problemas que son complejos para las técnicas convencionales (Campbell, 2002).

Para la RNA se tienen los valores de entrada, parámetro de sesgo y peso para estos parámetros, que son el parámetro de aprendizaje del modelo. El marco adaptativo cambia su estructura durante el proceso de aprendizaje por retropropagación de errores - cambiando la ponderación de los nodos de entrada- para minimizar el error experimentado en la salida. La RNA puede ser un modelo multicapa (MLP). El MLP puede estar compuesto por varias capas ocultas intermedias que conectan las capas de entrada y salida, y puede utilizarse para modelar relaciones complejas entre las variables de entrada y salida (Lee y Evangelista, 2006; Conforti et al., 2014).

Las ventajas de estos métodos son la rapidez de predicción, la capacidad de manejar datos y la capacidad de identificar fácilmente relaciones complejas. Sin embargo, los datos de entrada siempre deben convertirse en numéricos. Cuando se utiliza este método, existe el riesgo de sobreajuste y debe tenerse en cuenta que el tiempo de procesamiento puede ser muy largo; el tiempo de procesamiento puede aumentar en el caso de redes neuronales de gran tamaño.

3.4 Avances en la investigación sobre la estimación de la distribución espacial del COS

Son varios los estudios que han demostrado la importancia de estimar la distribución espacial del contenido de COS y sus reservas, debido a la capacidad de almacenar o secuestrar el C que tiene el suelo. Esto ha cobrado mayor énfasis en los últimos años debido a los efectos del cambio climático (Zomer et al., 2017; Chenu et al., 2019). Chen et al. (2022) concluyeron que el mayor porcentaje de estudios sobre mapeo de parámetros edáficos corresponde al COS y a la materia orgánica, con un 77.5%.

Los primeros estudios realizados con datos de sensores remotos para estimar el COS han arrojado una fuerte correlación con los datos en campo, como lo muestran Croft et al. (2012) y (Paul, 2013), mientras que en otros casos, las correlaciones fueron débiles entre las mediciones en el laboratorio y los valores estimados en los modelos, como se observa en Gómez et al. (2008), lo cual se atribuyó a factores como los requisitos de procesamiento computacional, disponibilidad de datos y la complejidad técnicas de las estimaciones del COS (Paul, 2013). El crecimiento continuo de las capacidades computacionales, la gran explosión de 'Big Data', el desarrollo de algoritmos de minería de datos, (Minasny y McBratney, 2016; Wang et al., 2018), la disponibilidad de nuevos datos como los DEM y la mejora de la resolución de las imágenes satelitales han permitido que dichos problemas se hayan superado. De acuerdo con Chen et al. (2022), los parámetros del relieve calculados a partir de los DEM se han utilizado en más de 200 artículos y son una covariable importante en la explicación del COS (Chen et al., 2022). Resultados similares reportados por Davy y Koen (2013) en Australia, quienes señalaron que las variables topográficas fueron de gran importancia en contenido del COS.

Schillaci et al. (2017) encontraron que la integración de datos de sensores remotos aumentaba la capacidad predictiva del COS en comparación con los modelos construidos sin covariables provenientes de los sensores remotos. Wang et al. (2018) concluyen que el Índice de Vegetación de Diferencia Normalizada (NDVI) es el más utilizado en estudios de COS, aunque no es el único, como lo demuestran Yang et al. (2008) quienes encontraron que el Índice de Vegetación Mejorada (EVI), explicaba entre el 50-66% de la variación en el stock de COS en sus modelos de predicción. Armas et al. (2017) encontraron que las predicciones del COS dependían de la profundidad del suelo. En los horizontes superficiales de 0 a 30 centímetros, las mejores correlaciones se obtuvieron con la precipitación, el EVI y el índice de planitud de fondo de valle de resolución múltiple, que es un indicador de las dimensiones espaciales de los valles y de las zonas de acumulación,

mientras que, a profundidades superiores a los 30 centímetros, la precipitación y las variables topográficas fueron las que mayor importancia tuvieron con el COS.

En referencia a los métodos de estimación de la distribución espacial del COS, el número de artículos relacionados era bajo antes de 2013 y dominaban los estudios con base en la geoestadística y modelos lineales. A partir del 2014, hubo un gran aumento de los modelos no lineales, que alcanzó su punto máximo en 2019 y 2020, lo cual se atribuye a los avances computacionales que permitieron el desarrollo de los algoritmos de aprendizaje automatizado. Diversos investigadores (Guo et al., 2015; Mouazen et al., 2010) han demostrado que los modelos basados en los algoritmos de aprendizaje automatizado son más precisos que los métodos estadísticos como la regresión lineal paso a paso, la regresión de componentes principales y la regresión de mínimos cuadrados parciales.

Bernardeu et al. (2014), compararon los Modelos Lineales Generalizados (GLM), Modelos Aditivos Generalizados (GAM), Splines de Regresión Adaptativa Multivariante (MARS), Random Forest (RF), Máquinas de Vectores Soporte (SVM), K-vecinos más próximos (KNN) y encontraron que los modelos RF y SVM obtuvieron los mejores resultados para la predicción del COS, mientras que los modelos GLM y GAM son los de desempeño más bajos. Were et al. (2015) compararon métodos predictivos de aprendizaje automatizado como regresión de vectores de soporte (SVR), las redes neuronales artificiales (ANN) y el bosque aleatorio (RF), encontraron que BRT y RF fueron los mejores.

Estudios más recientes (Zhou et al., 2020) utilizaron datos de imágenes multiespectrales y datos de radar con algoritmos de aprendizaje automatizado para la predicción. Evaluaron y compararon los algoritmos Random Forest (RF), Máquinas de Vectores Soporte (SVM), Árboles de regresión potenciados (BRT), Árboles de regresión y clasificación (CART). Encontraron que la elección del algoritmo de aprendizaje automatizado y las variables seleccionadas tuvieron alto impacto en el rendimiento de la predicción del COS, concluyeron que el modelo BRT superó en rendimiento a los otros métodos.

Específicamente en el ecosistema de páramo, Ayala (2019) utilizó Random Forest como modelo predictivo y elaboró un mapa digital del COS con datos de imágenes satelitales, modelo digital de elevación, variables climáticas de una base de datos, geología y taxonomía del suelo. Concluyó que las variables importantes fueron la formación

geológica, taxonomía del suelo, precipitación, altura, orientación, factor LS, índice BI, temperatura y temperatura superficial, con un 82% de poder predictivo.

En Boyacá, Cely et al. (2016) determinaron la línea base del potencial de captura de carbono en los suelos de siete complejos de páramos, y validaron una metodología para su estimación bajo diferentes usos. Este estudio se convierte en un referente para continuar con investigaciones que permitan generar información más detallada con respecto al COS. En el páramo de la Cortadera, Cuervo et al. (2016) determinaron las fracciones del COS en laboratorio. Sin embargo, estos dos estudios no evaluaron la distribución espacial del COS. Gutiérrez et al. (2020) reportan un estudio para la estimación del COS en los suelos de páramo de Colombia utilizando Random Forest y diferentes variables predictoras.

Como conclusiones, los modelos de predicción basados en aprendizaje automatizado que han tenido mejores rendimientos son los bosques aleatorios (Random Forest) y los árboles de regresión potenciados. No existe un método estandarizado para estimar la distribución espacial del COS. Los estudios revisados han utilizado diferentes variables predictoras y diversos métodos de análisis. La importancia de las variables predictoras depende de las condiciones de cada zona referentes a uso, cobertura, suelos, clima y relieve. En consecuencia, se requiere adelantar investigaciones más detalladas para condiciones específicas que permitan hacer estimaciones confiables de la distribución espacial del COS.

4. Materiales y métodos

4.1 Zona de estudio

La zona de estudio incluye el páramo de la Cortadera y la zona de clima muy frío aledaña (Figura 1). Es de gran importancia hídrica, por ser fuente de recarga del embalse de La Copa, y surtir los ríos Pesca, Jordán, Tuta, Salitre, Muche, y por formar parte de las cuencas de los ríos Chicamocha y Upía (Corporación Autónoma Regional de Boyacá, 2015). Incluye partes de los municipios de Pesca, Toca, Tuta, Siachoque, Firavitoba y Rondón del departamento de Boyacá y es parte del complejo de páramos de Tota – Bijagual – Mamapacha.



Figura 1 Ubicación zona de estudio

4.2 Metodología

La metodología empleada en este estudio (Figura 2) partió de una revisión de literatura, para determinar los posibles factores que influyen en el COS, referidos al clima, el relieve, el suelo y la vegetación, con el fin de seleccionar las variables predictoras del COS y con base en esta revisión se buscaron los datos de sensores remotos para estimar las variables predictoras. Posteriormente, se diseñaron diversas simulaciones con los métodos de bosques aleatorios (Random Forest) y redes neuronales, combinando diversas variables teniendo en cuenta la colinealidad de las variables y su relación con los factores que inciden en el COS. Con base en indicadores estadísticos, se evaluó el desempeño de los modelos para predecir el COS teniendo como base el COS determinado en el laboratorio en 169 muestras de suelo tomadas en el campo. Con los modelos de mejor desempeño, se efectuó la estimación de la distribución espacial del COS. A continuación, se presenta la descripción detallada de cada una de las etapas de la investigación que se muestran en la figura 2.

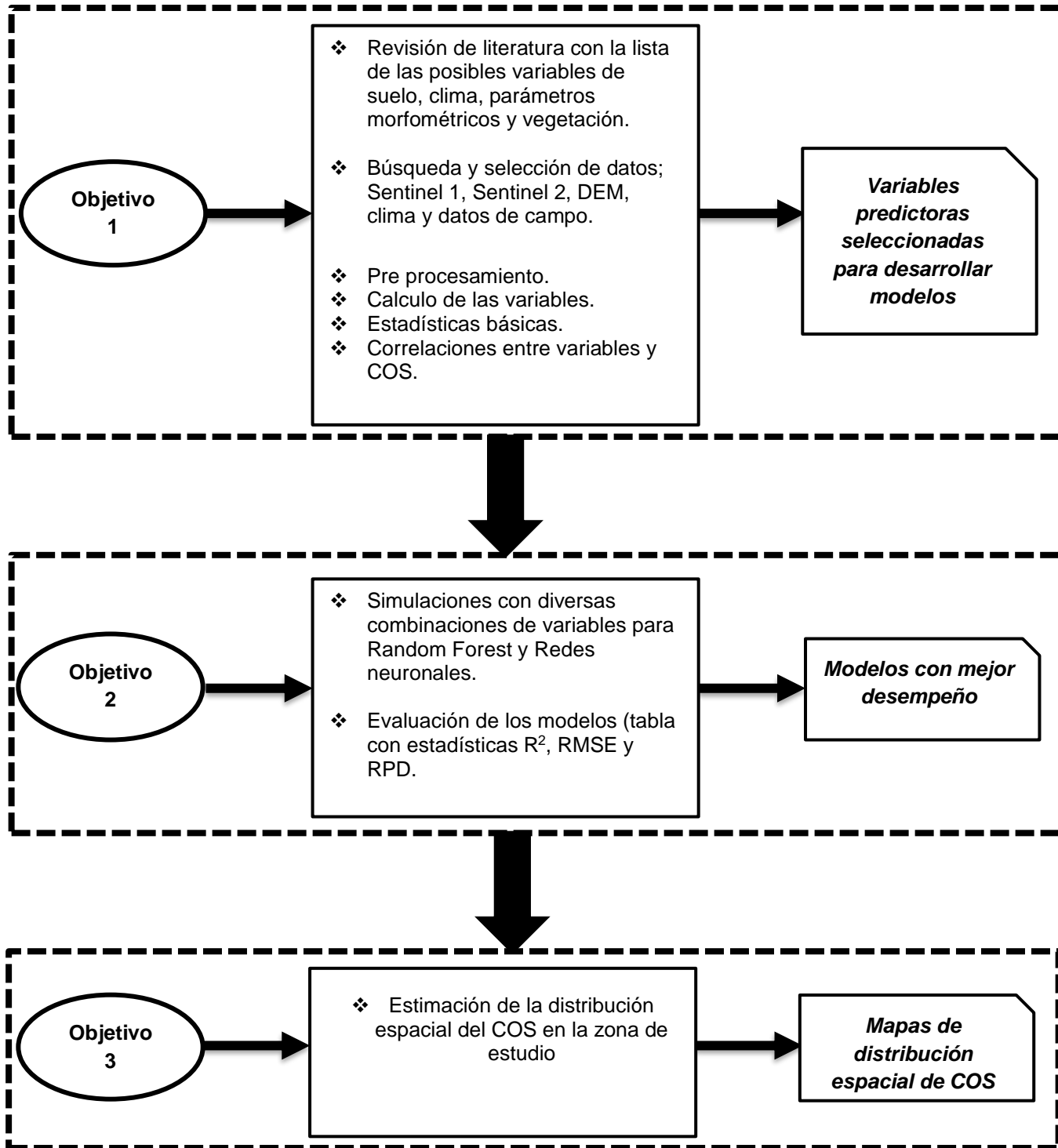


Figura 2 Diagrama de flujo metodología.

4.2.1 Datos de carbono orgánico del suelo para entrenamiento y validación de los modelos

Para los datos de COS, se tomaron un total de 169 muestras georreferenciadas, recopilados en dos muestreos en el 2017 y en el 2020 por el grupo de investigación GIPSO (Universidad Pedagógica y Tecnológica de Colombia). Para cada uno de los puntos, se tomaron muestras de suelo a dos profundidades (0-15 cm y 15-30 cm) y se procesaron en el Laboratorio de Suelos y Aguas de la Universidad Pedagógica y Tecnológica de Colombia, determinando el porcentaje de carbono (CO) por el método de Walkley Black (oxidación húmeda). Se estimó el contenido de carbono orgánico del suelo en toneladas (Ec.1) (Gardi et al., 2014), a partir del contenido de la concentración de carbono del suelo, la profundidad del suelo, la densidad aparente (ρ_b) y la fracción de carbono orgánico (Rügnitz et al., 2009).

$$\text{COS: } C \times \rho_b \times \text{Profundidad} \times (1 - \text{frag}) \times 10 \quad (1)$$

En donde:

COS: Contenido de carbono orgánico del suelo, representativo del tipo de uso del suelo ($\text{t ha}^{-1}\text{C}$).

C: Concentración de carbono orgánico del suelo obtenido en laboratorio

ρ_b : Densidad aparente.

Profundidad: profundidad del horizonte o espesor de la capa de suelo, en metros (m).

Frag: volumen porcentual de fragmentos gruesos/100, sin dimensiones

4.2.2 Variables calculadas a partir de la imagen Sentinel 2

Se utilizó una imagen Sentinel-2A de la Agencia Espacial Europea del día 17 de enero 2019, cuyas especificaciones se muestran en la tabla 1 (ESA, 2015). Las imágenes Sentinel 2 pueden descargarse con diferentes niveles de procesamiento; para este estudio, se utilizó una imagen de nivel 2A que se encuentran orto rectificadas con niveles de reflectancia por debajo de la atmósfera (BOA), corregidas con el algoritmo Sen2Cor, que es una combinación de técnicas para realizar correcciones atmosféricas (ESA, 2015). A

partir de la imagen Sentinel 2, se calcularon los índices espectrales que se muestran en la tabla 3.

Los índices espectrales calculados se utilizaron como indicadores de la vegetación ya que permiten discriminar el comportamiento reflectivo de la vegetación, evaluando su estado de manera cualitativa y cuantitativa (Bannari et al., 1995). El ecosistema es una cubierta de tipo vegetal en donde diferentes actividades antrópicas podrían alterar su estructura y características. En condiciones naturales, los índices caracterizan el estado en el que se encuentre y se convierten en variables que se utilizan para predecir el COS, por su relación con la cobertura vegetal (Ayala, 2019).

Tabla 1 *Bandas de imagen Sentinel 2.*

Banda	Resolución espacial	Longitud de onda central	Descripción
B1	60 m	443 m	Ultra azul (Costa y Aerosol)
B2	10 m	490 m	Azul
B3	10 m	560 m	Verde
B4	10 m	665 m	Rojo
B5	20 m	705 m	Visible e Infrarrojo Cercano (VNIR)
B6	20 m	740 m	Visible e Infrarrojo Cercano (VNIR)
B7	20 m	783 m	Visible e Infrarrojo Cercano (VNIR)
B8	10 m	842 m	Visible e Infrarrojo Cercano (VNIR)
B8a	20 m	865 m	Visible e Infrarrojo Cercano (VNIR)
B9	60 m	940 m	Onda Corta Infrarroja (SWIR)
B10	60 m	1375 m	Onda Corta Infrarroja (SWIR)
B11	20 m	1610 m	Onda Corta Infrarroja (SWIR)
B12	20 m	2190 m	Onda Corta Infrarroja (SWIR)

Fuente: Esa, (2015)

Las imágenes Sentinel 2 pueden ser descargadas con diferentes niveles de procesamiento. Para este estudio, se utilizó una imagen de nivel 2A, que se encuentran orto rectificadas con niveles de reflectancia por debajo de la atmósfera (BOA) y corregida con el algoritmo Sen2Cor, una combinación de técnicas para realizar correcciones atmosféricas (ESA, 2015). A partir de la imagen Sentinel 2, se calcularon los índices espectrales que se muestran en la tabla 2. Se utilizaron las imágenes Sentinel 1 y 2 y,

como demostró Zhou (2020) en su estudio, la relación entre los conjuntos de variables provenientes de Sentinel 2, las variables topográficas y las covariables VV/VH proporcionan buenos resultados en la explicación del COS. Esto mejora los resultados en comparación con estudios que solo utilizaron índices espectrales y muestra un rendimiento superior a los datos de Sentinel sobre otros datos como Landsat y Modis.

Tabla 2 Índices espectrales calculados con base en imagen Sentinel 2

Índice	Fórmula	Fórmula Sentinel 2	Referencia
Índice de vegetación de la diferencia normalizada (NDVI)	$NDVI = \frac{(NIR - Red)}{(NIR + Red)}$	$NDVI \text{ Sentinel 2} = \frac{(B8 - B4)}{(B8 + B4)}$	(Rouse et al., 1973).
Índice de vegetación mejorado (EVI)	$EVI = G * \frac{(NIR - R)}{(NIR + C1 * R - C2 * B + L)}$	$EVI \text{ Sentinel 2} = \frac{2.5 * ((B8 - B4) / (B8 + 6 * B4 - 7.5 * B2 + 1))}{1}$	(Gao et al., 2003).
Índice de suelo desnudo (BSI)	$BSI = \frac{((Red + SWIR) - (NIR + Blue))}{((Red + SWIR) + (NIR + Blue))}$	$BSI \text{ Sentinel 2} = \frac{(B11 + B4) - (B8 + B2)}{(B11 + B4) + (B8 + B2)}$	(Rikimaru et al., 2002).
Índice de vegetación avanzada (AVI)	$AVI = [NIR * (1 - Red) * (NIR - Red)]^{1/3}$	$AVI \text{ (Sentinel 2)} = [B8 * (1 - B4) * (B8 - B4)]^{1/3}$	(Rikimaru et al., 2002).
Índice de vegetación de la diferencia normalizada verde (GNDVI)	$GNDVI = \frac{(NIR - GREEN)}{(NIR + GREEN)}$	$GNDVI \text{ (Sentinel 2)} = \frac{(B8 - B3)}{(B8 + B3)}$	(Gitelson et al., 1996).
Índice de estrés hídrico (MSI)	$MSI = MidIR / NIR$	$MSI \text{ Sentinel 2} = B11 / B8$	(Hunt y Rock, 1989).
Índice diferencial de agua normalizado (NDWI)	$NDWI = \frac{(NIR - SWIR)}{(NIR + SWIR)}$	$NDWI \text{ (Sentinel 2)} = \frac{(B3 - B8)}{(B3 + B8)}$	(Gao, 1995).

Índice normalizado de área quemada (NBR)	$NBR = (NIR - SWIR) / (NIR + SWIR)$	NBRI (Sentinel 2) = $(B8 - B12) / (B8 + B12)$	(Chuvieco et al., 2002).
--	-------------------------------------	---	--------------------------

4.2.3 Variables geomorfométricas calculadas a partir de un Modelo digital de elevación

El modelo digital de elevación (DEM por sus siglas en inglés) utilizado, es un producto corregido radiométricamente por Alaska Satellite Facility, con una resolución espacial de 12,5m (ASF DAAC, 2022).

A partir de este DEM Alos-Palsar, se calcularon las siguientes variables con el software SAGA GIS (Conrad, 2006):

- **Elevación (m)**

Es el parámetro primario que caracteriza a los MDE y se refiere a los valores de alturas de los pixeles, siendo este la base para el cálculo de los demás parámetros.

- **Pendiente (°)**

La pendiente (S) es la primera derivada de una superficie y tiene tanto magnitud como dirección (Chang y Tsai, 1991), la fórmula en donde z es elevación, x-y las coordenadas planas (Ec. 2).

$$S = \sqrt{\left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2} \quad (2)$$

- **Dirección de flujo**

La dirección de flujo está determinada por la dirección del descenso más empinado, o caída máxima, desde cada celda. La cantidad de flujo que recibe cada pixel con pendiente descendente se estima como una función del gradiente de pendiente máxima (Qin et al., 2007) (Ec. 3)

$$d_i = \frac{(\tan \beta_i)^{f(e)} L_i}{\sum_{j=1}^n (\tan \beta_j)^{f(e)} L_j} \quad (3)$$

- **Acumulación de flujo**

La acumulación de flujo representa un flujo acumulado para cada celda, determinado por la acumulación del peso de todas las celdas que fluyen hacia cada celda de pendiente descendente, su cálculo depende de la dirección de flujo y está basado en la cantidad total de celdas, o en una fracción de las mismas, que fluyen hacia cada celda en el ráster de salida.

- **Aspecto**

Se refiere a la dirección de la pendiente y se expresa en grados en el sentido de la manecilla del reloj (Zhou y Liu, 2004) (Ec. 4)

$$A = 270^\circ + \arctan\left(\frac{fy}{fx}\right) - 90^\circ \frac{fx}{|fx|} \quad (4)$$

- **Curvatura del perfil**

La curvatura del perfil es la segunda derivada de la superficie y se calcula en el plano vertical en una línea de flujo (Wilson y Gallant, 2000) (Ec. 5), en donde θ_x es la pendiente en la dirección x y θ_y es la pendiente en la dirección y, θ_{xx} es la segunda derivada de la pendiente en la dirección x, θ_{yy} es la segunda derivada de la pendiente en la dirección y, y θ_{xy} es la segunda derivada del producto de las pendientes en la dirección x & y.

$$C_v = \frac{\theta_{xx}\theta_x^2 + 2\theta_{xy}\theta_x\theta_y + \theta_{yy}\theta_y^2}{\theta_x^2 + \theta_y^2 \sqrt{(\theta_x^2 + \theta_y^2 + 1)^2}} \quad (5)$$

- **Índice de planitud del fondo de valle (MRVBF)**

Es un índice topográfico diseñado para identificar áreas de depósito de material. Maneja los valores cercanos a cero, que indican terreno erosivo, y valores 1 o

superiores, que indican progresivamente mayor depositación (Gallant y Dowling, 2003).

- **Índice topográfico de humedad (TWI)**

Es un parámetro que describe la tendencia de una celda para acumular agua. Está relacionado con la humedad del suelo y refleja la tendencia del suelo a la generación de escorrentías: en cuanto más elevado es el valor del índice, mayor humedad debe presentar la celda en función de su configuración topográfica (Roa y Kamp, 2012) (Ec. 6).

$$TWI = \ln\left(\frac{A_s}{\tan \theta}\right) \quad (6)$$

- **Factor LS**

LS es el factor de longitud de la pendiente que explica los efectos de esta covariable en la erosión (Wischmeier y Smith, 1978) (Ec. 7); L se refiere a la longitud de la pendiente y S al efecto del grado de inclinación de la pendiente (Panagos et al., 2015).

$$LS = (n + 1) \left(\frac{As}{22.13}\right) n \left(\frac{\sin \beta}{0.0896}\right) m \quad (7)$$

4.2.4 Variables de clima utilizadas

Los datos de precipitación y temperatura se obtuvieron de la plataforma Worldclim, versión 2.1 publicada en el año 2020, que corresponden al periodo 1970-2000, con resolución de 30 arco-segundos (la máxima disponible), en formato GeoTIFF. Los datos de esta plataforma están calibrados con datos de estaciones meteorológicas. Se utilizó la temperatura promedio multianual en grados centígrados (°C) y la precipitación (PPT) promedio multianual en milímetros (mm).

4.2.5 Variables calculadas a partir de la imagen Sentinel 1

La imagen utilizada hace parte de la misión Sentinel-1, son datos en la banda C (ESA,2020), tomada el día 28 junio de 2017 (ESA,2020). El modo operativo principal de la misión presenta una franja de 250 km, con una resolución geométrica de 10 m en el nivel 1 de procesamiento con datos detectados en terreno (GRD), sus especificaciones se muestran en la tabla 3 (ESA, 2020).

Tabla 3 Especificaciones de imagen Sentinel 1

Especificaciones	Imagen Sentinel - 1^a
Orbita	Descendiente IW
Modo de imagen	Banda C (5.4 GHz)
Frecuencia Polarización	VV – VH
Tipo de producto	Level -1 GRD
Resolución espacial	10 m

Fuente: ESA, (2015)

El procesamiento de la imagen Sentinel 1 se realizó con el software SNAP, proporcionado por la ESA (Agencia Espacial Europea). La imagen se calibró con el ángulo de incidencia Sigma para obtener los píxeles cuadrados. Se aplicó el filtro multilooking a dos rangos (Mancera, 2019), y el ruido se corrigió con el filtro Lee Sigma en una ventana de 7x7. Aunque la ESA recomienda corregir los datos con el MDE SRTM, en esta investigación la corrección se utilizó Alos-Palsar debido a que tiene mayor resolución y da mejores resultados (Tovar et al., 2019). Los valores de niveles digitales se convirtieron a valores de retrodispersión en la escala de decibelios (ESA, 2015), y se obtuvieron los datos polarimétricos VV-VH, que fueron las covariables utilizadas provenientes de la imagen Sentinel 1.

4.3 Selección de variables

Para seleccionar las variables, se efectuó la revisión de literatura para identificar variables que incidan en el COS, y se tuvo en cuenta la disponibilidad de datos de la zona de estudio, creando una lista de variables potenciales para estimar el COS. Posteriormente, se calculó el coeficiente de correlación de Pearson entre los datos de campo del COS y las covariables que se muestran en la figura 3, con el fin de establecer cuáles son las variables que más influyen en el COS. Igualmente, se calculó el factor de inflación de la varianza (VIF) (Zuur et al., 2007), para establecer la redundancia estadística y la colinealidad. Este método calcula para cada covariable un valor que resume el vector correspondiente de la matriz de correlaciones (Zuur et al., 2007). De acuerdo con los valores resultantes y según Armas et al (2017), se escogen las variables que menor valor de inflación de la varianza tengan y mayor coeficiente de correlación.

4.4 Desarrollo de los modelos predictivos

Después de haber calculado cuales son las covariables con mayor correlación, se utilizaron éstas como datos de entrada para desarrollar los modelos con los algoritmos bosques aleatorios y redes neuronales. Se efectuó pruebas utilizando combinaciones de las covariables para evaluar la influencia de cada una, empezando con la evaluación de un modelo con todas las covariables, aunque no se correlaciono con el COS. Luego, se evaluaron distintos casos omitiendo o agregando covariables, siempre que no fueran colineales, según se muestra en la tabla 4. Para generar los modelos, se utilizó el 70% de las muestras de suelo tomadas en el campo, y para la validación, se empleo el 30%, mediante una selección al azar

Tabla 4 Modelos evaluados con diversas combinaciones de covariables.

0-15 cm	15-30 cm
Modelo 1 (Todas las covariables)	Modelo 1 (Todas las covariables)
Modelo 2 (RF, RN, Altura, Factor LS, MRVBF, NDVI, Temperatura, TWI, VH)	Modelo 2 (Altura, MRVBF, NDVI, Temperatura, PPT, VH)
Modelo 3 (Altura, MRVBF, NDVI, Temperatura, TWI, VH)	Modelo 3 (Altura, NDVI, Temperatura, PPT, VH)
Modelo 4 (Altura, Factor LS, NDVI, Temperatura, TWI, VH)	Modelo 4 (Altura, NDVI, Temperatura, VH)
Modelo 5 (Altura, NDVI, Temperatura, TWI, VH)	Modelo 5 (Altura, MRVBF, NDVI, Temperatura, VH)
Modelo 6 (Altura, Factor LS, MRVBF, NDVI, Temperatura, TWI, VH)	Modelo 6 (Altura, MRVBF, GNDVI, Temperatura, PPT, VH)
Modelo 7 (Altura, Factor LS, MRVBF, MSI, Temperatura, TWI, VH)	Modelo 7 (Altura, MRVBF, NDVI, Temperatura, PPT, VH)
Modelo 8 (Altura, Factor LS, MRVBF, NBRI, Temperatura, TWI, VH)	Modelo 8 (Altura, MRVBF, NDVI, GNDVI, NDVI, Temperatura, PPT, VH)
Modelo 9 (Altura, Factor LS, MRVBF, NDVI, BSI, EVI, GNDVI, MSI, NBRI, NDMI, NDVI, Temperatura, TWI, VH)	Modelo 9 (Altura, MRVBF, NDVI, VH)
Modelo 10 (Altura, NDVI, TWI, VH,)	Modelo 10 (MRVBF, NDVI, Temperatura, VH)
Modelo 11 (NDVI, Temperatura, TWI, VH)	Modelo 11 (MRVBF, NDVI, Temperatura, PPT, VH)
Modelo 12 (Factor LS, MRVBF, NDVI, Temperatura, TWI, VH)	Modelo 12 (Altura, MRVBF, NDVI, PPT, VH)
Modelo 13 (Altura, Factor LS, MRVBF, NDVI, TWI, VH)	

Las librerías utilizadas fueron Pandas para Python, empleadas para el análisis y manipulación de datos, sklearn.model_selection, que permite dividir conjuntos en subconjuntos, sklearn.preprocessing, utilizada para normalizar los datos en un rango entre 0 y 1, sklearn.metrics, que contiene métricas de evaluación de modelos, matplotlib utilizada para las visualizaciones gráficas. Además, para la regresión de bosques aleatorios se utilizó la librería sklearn.ensemble.randomforestregressor, con diferentes interacciones de parámetros para el número de árboles, y la profundidad de cada árbol, y la librería sklearn.model.selectionGridsearchcv para buscar los mejores hiperparámetros para el modelo de bosque de regresión en cada caso evaluado. Para las redes neuronales, se utilizó la librería de tensorflow y sklearn.model.selectionGridsearchcv para escoger los mejores hiperparámetros entre las interacciones en el número de capas ocultas en cada red, el número de neuronas en cada capa oculta, la tasa de aprendizaje utilizada en el

optimizador, el número de épocas de entrenamiento y el tamaño del lote utilizado durante el entrenamiento.

4.5 Evaluación y validación modelos

Para la evaluación y validación de los modelos desarrollados (figura 3), se analizó el error medio cuadrático (RMSE) y el coeficiente de determinación (R^2). Se seleccionaron los modelos de mejor desempeño para efectuar la validación. Se realizó validación cruzada y validación por subconjuntos de datos, tomando 70% para entrenamiento y el restante 30% para validación.

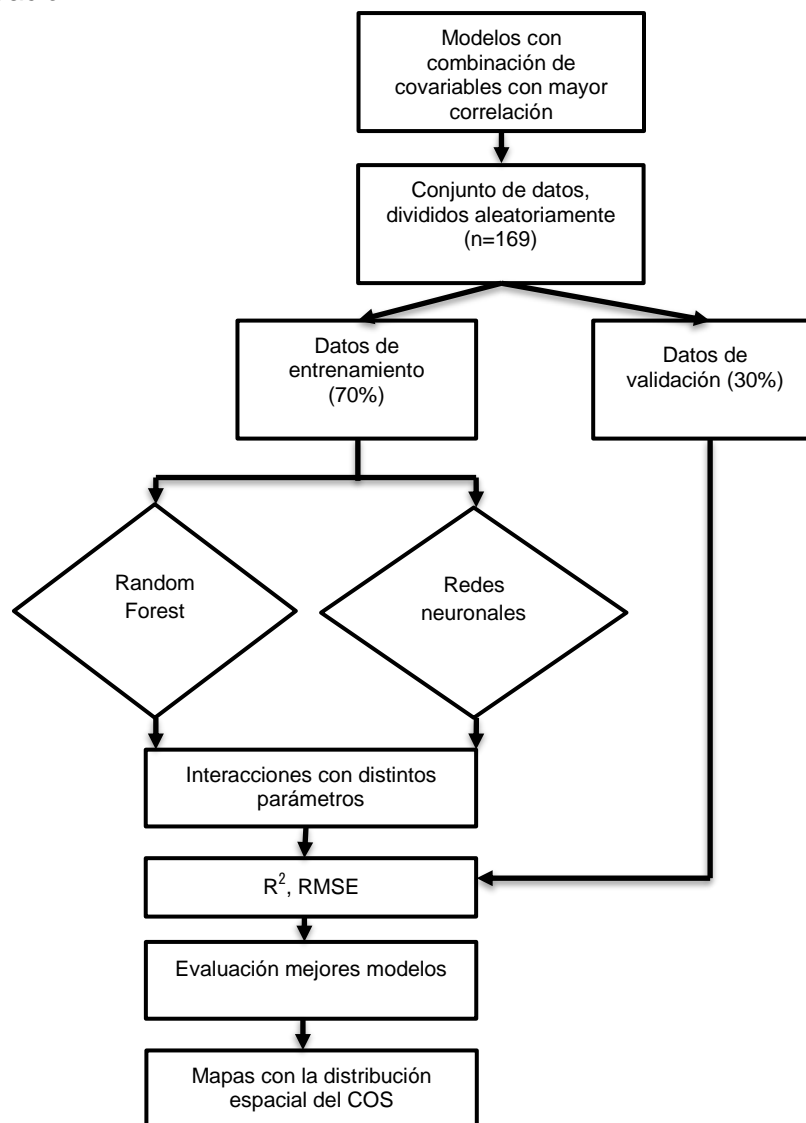


Figura 3 Diagrama de flujo de evaluación y validación de los modelos.

4.6 Estimación de la distribución espacial del COS

Para los modelos de mejor desempeño, se realizó el mapa de la distribución espacial del COS mediante Python, con el fin de normalizar los datos y predecir los valores para así agregar las predicciones al conjunto de datos del mapa. Lo anterior se efectuó con base en la biblioteca Folium, que permite crear un mapa con los valores estimados para cada punto y luego para toda el área, con el modelo entrenado para la estimación del COS.

5. Resultados y discusión.

5.1.1 Variables para la estimación del contenido del COS

En la tabla 5, se muestran los valores del COS para las dos profundidades evaluadas. El contenido oscilo entre 11.4 t ha^{-1} y 187 t ha^{-1} para la profundidad de 0-15 cm, y entre 6.6 t ha^{-1} y 372.6 t ha^{-1} para 15-30 cm, presentando en ambos casos un coeficiente de variación alto.

Tabla 5 Estadísticas básico descriptivas del COS (t ha^{-1}).

	COS 0-15cm	COS 15-30cm
Mínimo	11.4	6.6
1st Qui.	23.25	40.27
Mediana	37.2	61.65
Media	47.44	77.46
3rd Qui.	57.3	95.7
Máximo	187.2	372.6
Desviación estándar	36.38	59.8
Coficiente de variación	77%	85%

Como se puede observar en la figura 4, la distribución de COS para la profundidad de 0-15 cm está sesgada a la izquierda. Esto se debe a la concentración de valores por debajo de 50 t ha^{-1} , en menor proporción aquellos entre 50 y 150 t ha^{-1} , y muy pocos por encima de los 150 t ha^{-1} .

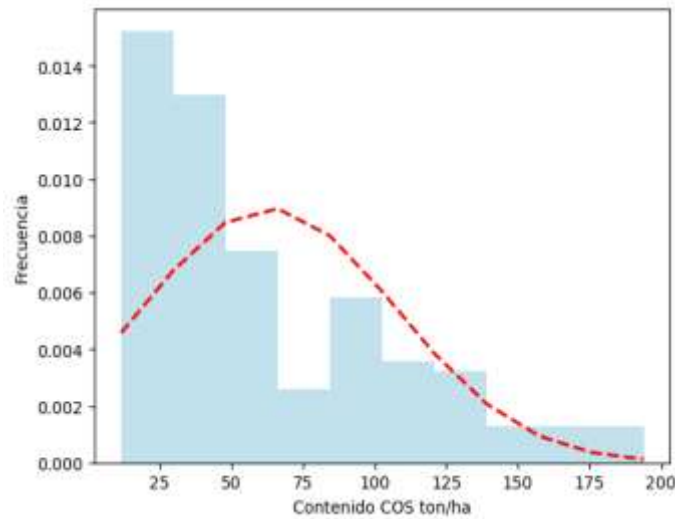


Figura 4 Distribución de los datos de COS de 0-15 cm

En la figura 5, se puede observar la distribución espacial de los puntos muestreados y clasificados por el contenido COS para la profundidad de 0-15 cm. Los menores valores se ubican en la parte norte de la zona de estudio, que corresponde a las áreas de mayor intervención por la actividad humana, mientras que los contenidos más altos se localizan en la parte centro y sur de la zona de estudio, que coincide con la zona más conservada del páramo. Igualmente, se observa un mayor número de muestras en la zona norte y menor en la zona centro y sur, lo cual puede tener incidencia en la estimación del COS. El COS en la zona no intervenida del páramo fue mayor que en aquellas áreas con actividades agropecuarias.

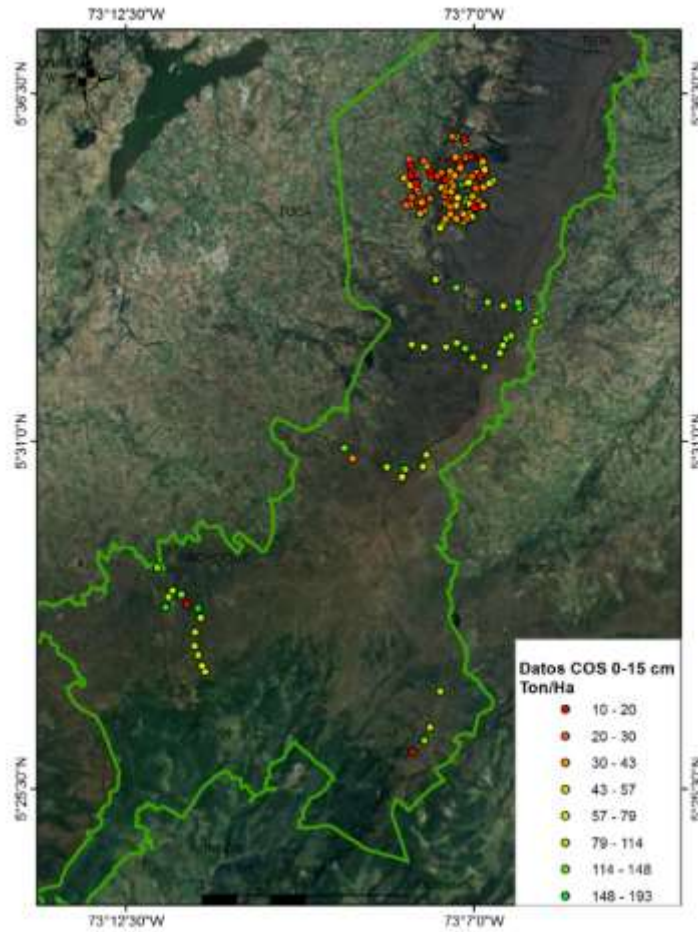


Figura 5 Distribución espacial de los datos COS de 0-15 cm

En la figura 6, se puede observar la distribución del COS para la profundidad de 15–30 cm. Al igual que para 0-15 cm (figura 5), la distribución está sesgada a la izquierda, pero con valores más altos en el extremo derecho, lo cual explicaría el mayor coeficiente de variación de los datos para esta profundidad.

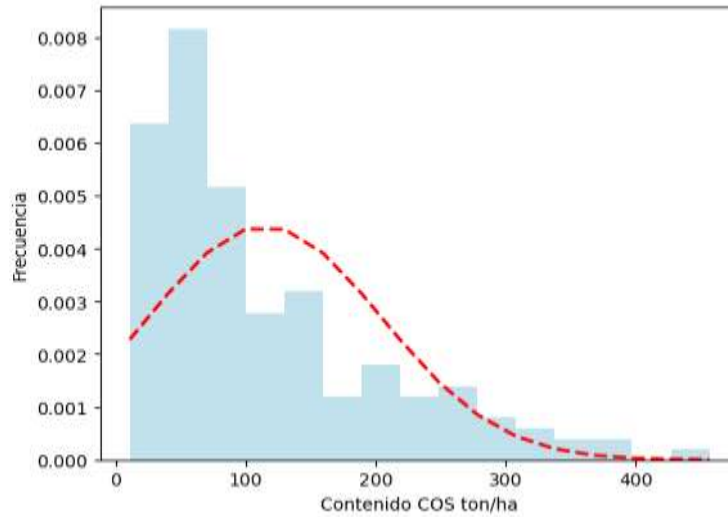


Figura 6 *Distribución de los datos de COS de 15-30 cm*

La distribución espacial de los datos muestreados del COS para la profundidad de 15-30 cm se muestra en la figura 7 y se comporta de manera similar a la profundidad de 0-15 cm (figura 5). Los menores valores están agrupados en la parte norte en comparación con los valores de la zona centro y sur, donde se encuentran los valores de contenido entre medios y altos. Para ambas profundidades, se pueden observar similitudes y diferencias en el conjunto de datos que se utilizó como entrenamiento para el modelo. Los valores de mayor contenido están distribuidos espacialmente en la parte sur y centro para ambas profundidades.

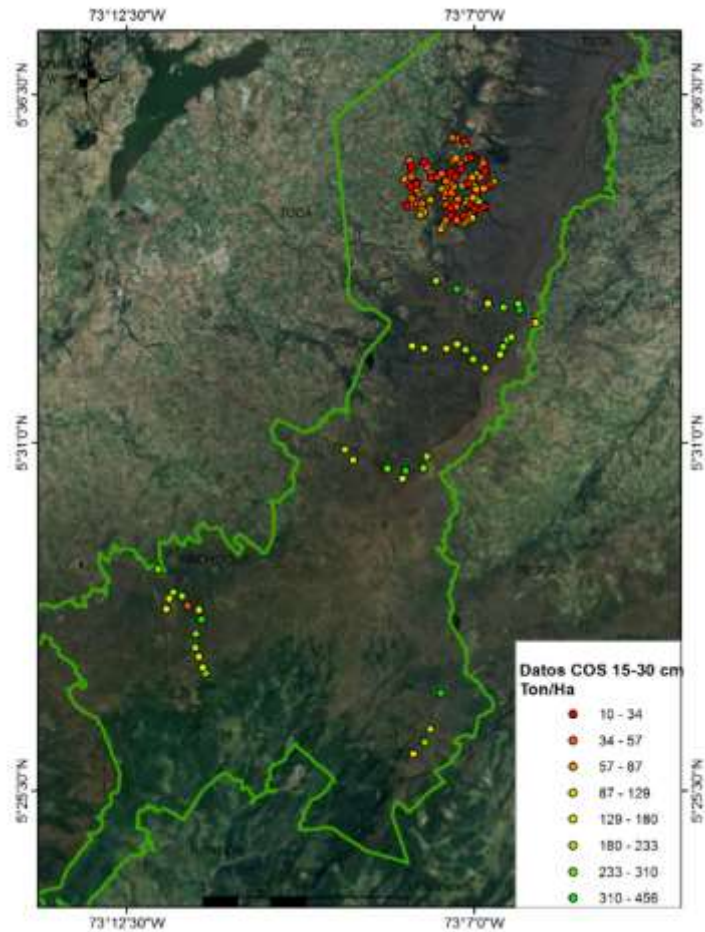


Figura 7 Distribución espacial de los datos COS de 15-30 cm

Los contenidos de COS en los datos de muestra de esta investigación coinciden con Gutiérrez et al. (2020), quienes reportaron para los suelos de páramo de Boyacá un contenido de COS para una profundidad de 0-30 cm, entre 22 t ha^{-1} y 337 t ha^{-1} . En la presente investigación, los menores valores corresponden a los suelos dedicados a la producción agrícola, lo cual indica mayor mineralización del COS posiblemente relacionado con el cambio de la cobertura natural del suelo por cultivos que lo exponen a una mayor incidencia de los factores climáticos.

Para los puntos muestreados, se procedió a extraer los valores de cada una de las covariables a partir de los datos en ráster. La tabla 6 presenta los resultados de las estadísticas básicas de las covariables. Solo el índice espectral MSI y las covariables provenientes de los datos de Sentinel 1 presentaron una distribución normal. Como los métodos utilizados son no paramétricos, no fue necesario realizar transformaciones para la normalidad.

Tabla 6 Estadísticas básica de las covariables en los sitios muestreados.

Datos	Normalidad	Mínimo	Media	Máximo
Altura (m)	0.02	2953	3123	3237
Aspecto	0.22	0	211.45	357.27
AVI	4.40E-06	0.16	0.219	0.32
BSI	0.07704	-0.07	0.197	0.37
Curvatura	3.60E-11	-4.48	-0.02	2.56
EVI	1.10E-08	0.1	0.17	0.3
Acumulación de flujo	0.22	0	110.8	9496
Dirección de flujo	0.22	1	45.88	193
GNDVI	3.80E-06	0.33	0.41	0.57
Factor LS	2.20E-16	0.72	10.56	59.64
MRRTF	2.20E-16	0	0.052	1.86
MRVBF	2.20E-16	0	0.18	1.85
MSI	0.32	0.82	1.6	2.41
NBRI	0.04	-0.3	-0.027	0.31
NDMI	0.01	-0.41	-0.22	0.09
NDVI	3.34E-06	0.2	0.35	0.57
NDWI	5.25E-07	-0.58	-0.44	-0.36
PPT (mm)	2.20E-16	848	985.8	1114
SAVI	7.76E-08	0.04	0.06	0.12
Pendiente (°)	0.07211	2.42	11.7	27.49
Temperatura(°C)	2.20E-16	9.93	10.76	11.53
TWI	2.50E-10	4.39	6.95	13.04
VH	0.92	-22.28	-17.57	-10.83
VV	0.05	-14.81	-12.01	-8.94

En la tabla 7 se presentan las correlaciones para las dos profundidades, entre COS y las covariables. Las variables altura, BSI, EVI, GNDVI, factor LS, MRVBF, MSI, NBRI, NDMI, NDVI, NDWI, SAVI, temperatura, TWI, VH presentaron correlaciones estadísticamente significativas con el COS a las profundidades de 0-15 y 15-30 cm. Mientras que el factor LS solo presentó correlación significativa para la profundidad de 0-15cm y la PPT para la profundidad de 15-30 cm. Las demás variables no presentaron correlaciones significativas. Los mayores coeficientes de correlación se presentaron entre COS y la mayoría de los índices espectrales, siendo el NDVI el de mayor valor con 0.52 para los 0-15 cm y 0.47 para los 15-30 cm. De las covariables calculadas con base en el DEM, la altura obtuvo una

de las mejores correlaciones con 0.55 para los 0-15 cm y 0.64 para los 15-30 cm, el factor LS y el TWI fueron significativas solo para la profundidad de los 0-15 cm. Otras covariables como la PPT fueron significativas solo para la profundidad de los 15-30 cm. La temperatura presentó correlaciones relativamente altas de -0.60 y -0.67 para las dos profundidades.

Tabla 7 *Correlación y significancia entre el COS y las covariables.(n=169)*

	COS.15	Significancia	COS.30	Significancia	VIF
Altura (m)	0.55	***	0.64	***	4.5
Aspecto	-0.03		-0.1		5.2
AVI	0.14		0.16		158
BSI	-0.41	***	-0.37	***	167
Curvatura	0.02		0.08		1.5
EVI	0.29	***	0.3	***	199
Acumulación de flujo	0.02		-0.05		1.4
Dirección de flujo	0.02		0.04		2.9
GNDVI	0.50	***	0.51	***	316
Factor LS	0.22	**	0.05		4.1
MRRTF	0.02		0.08		2.1
MRVBF	0.20	**	0.29	***	2.4
MSI	-0.43	***	-0.39	***	88
NBRI	0.45	***	0.41	***	118
NDMI	0.43	***	0.39	***	134
NDVI	0.52	***	0.48	***	20
NDWI	-0.50	***	-0.51	***	316
PPT (mm)	0.04		0.21	**	161
SAVI	0.32	***	0.33	***	324
Pendiente (°)	-0.08		-0.09		4.3
Temperatura(°C)	-0.60	***	-0.67	***	4.7
TWI	0.26	***	0.12		5.1
VH	0.28	***	-0.34	***	669
VV	-0.07		-0.07		4.4

***p>0.99, **p>0.95, *p > 0.90

En la tabla 7, además de la correlación entre las covariables, se evaluó el factor de inflación de la varianza (VIF). Cuando el VIF es mayor a 5, indica multicolinealidad, lo cual afecta la predicción del COS. Una de las primeras soluciones es eliminar las variables que se correlacionen entre sí, dejando la de mayor correlación con el COS (González, 1998).

Como se observa en la figura 8, los índices espectrales fueron las covariables que más se correlacionaban entre sí, mostrando altos valores del VIF, esto debido a que los cálculos realizados para obtener cada índice tienen como base bandas en común, explicando la alta correlación.

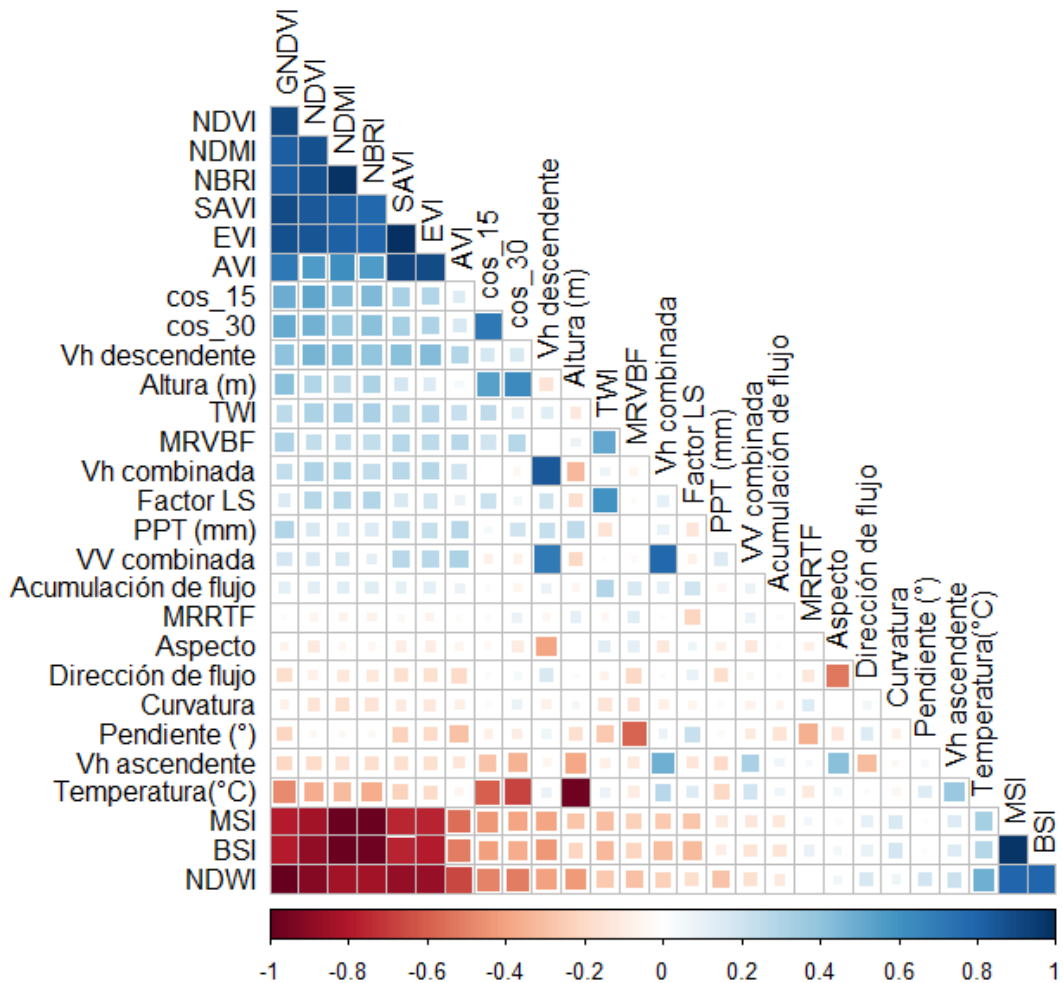


Figura 8 Correlación entre las covariables predictoras

Con base en los análisis anteriormente realizados, se procedió a seleccionar las covariables para uno de los casos a evaluar y a comparar con las covariables de NDVI, la altura, el TWI, MRVBF y el factor LS, la temperatura y la banda VH, como predictores para el modelo de COS. Para confirmar que el problema de la colinealidad entre las variables seleccionadas se haya resuelto, se realizó de nuevo el cálculo de VIF, cuyos

resultados se muestra en la figura 9, indicando que no hay colinealidad entre las variables seleccionadas.

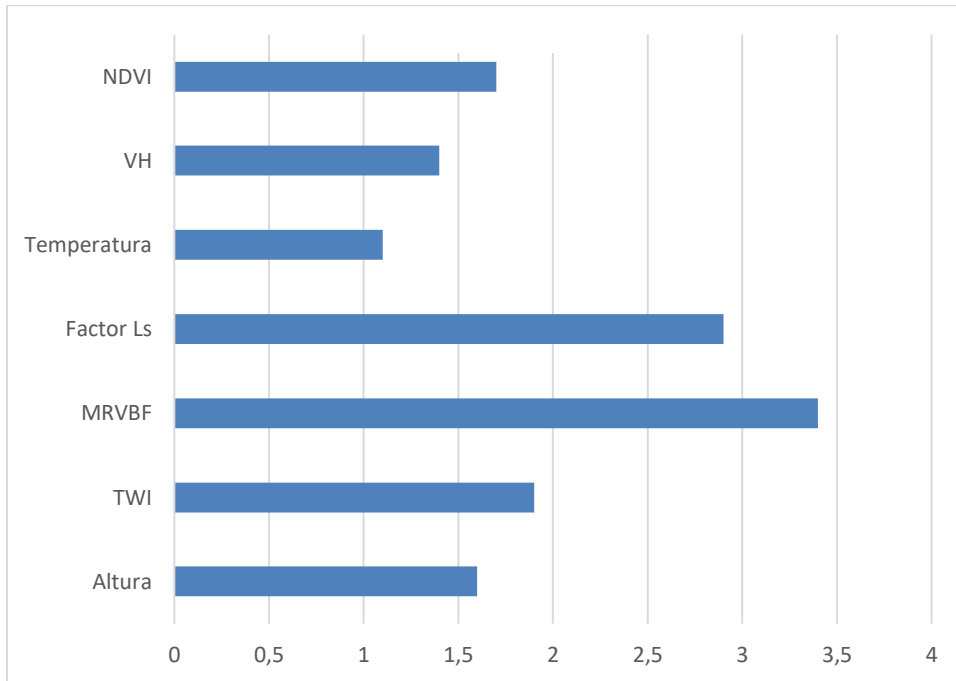


Figura 9 VIF de las covariables

En la figura 10 se presenta la distribución espacial de la altura. Las zonas de mayor altura están ubicadas en la parte central con un máximo de 3856 m, mientras que hacia sur y hacia el norte la altura va disminuyendo hasta un mínimo de 2900 m. Dado que existe una correlación entre la altura y la temperatura, y esta última es uno de los factores que inciden en la descomposición y mineralización de la materia orgánica. Las zonas de mayor altura presentan mayor COS, lo que coincide con los estudios de Quesada et al. (2020), Huamán et al. (2021) y Hunt et al. (2020), que indican que a medida que incrementa la altitud, hay menor temperatura y, por lo tanto, mayor capacidad de almacenamiento de COS.

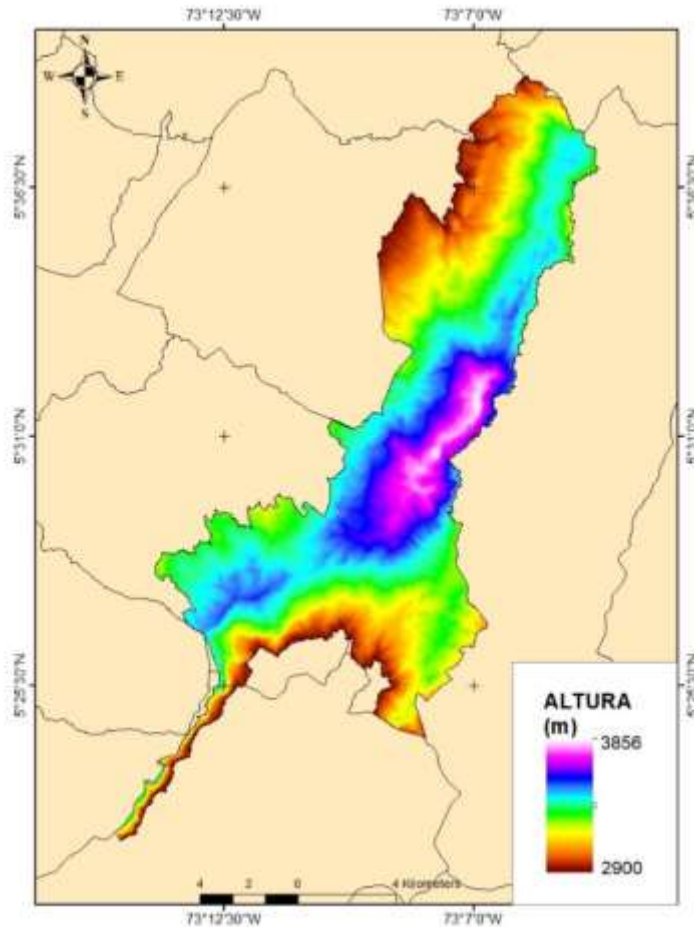


Figura 10 Variación espacial de altura zona de estudio.

Para esta investigación, una de las covariables que resultó significativa es la altura, lo que concuerda con Wanga et al. (2018) en el este de Australia, que encontró que la altura tenía efectos en el contenido de COS; frente a otras covariables topográficas como aspecto, pendiente, TWI, entre otros, que no fueron significantes en el COS. Dentro de las covariables más utilizadas para la predicción de COS se encuentra, en primer lugar, el relieve (Chen et al., 2022). Las covariables topográficas son de gran importancia por su papel de formación del suelo (Behrens et al., 2010) y son de los predictores más utilizados específicamente en la estimación del COS (Wang et al., 2018).

Además de la altura, el factor LS es una covariable que también tiene incidencia en el COS, concordando con el estudio de Reyes et al. (2018), donde el factor LS y las covariables topográficas tuvieron una importancia relevante en el COS, en una zona alta

de la cordillera de los Andes, un ecosistema relativamente similar al de la zona de estudio del páramo.

El estudio Davy y Koen (2013) explica que la topografía afecta el almacenamiento de COS mediante sus efectos en la profundidad de suelo y la humedad de suelo, concordando con las covariables significativas resultantes como el TWI, el factor LS, MRVBF y la altura, relacionadas con la retención de agua y nutrientes. Esta alta capacidad de retención deriva en el crecimiento de la vegetación, que termina aportando materia orgánica, componente principal del COS, y es un factor importante para el ciclo de nutrientes, infiltración y la capacidad retención de agua (Wang et al., 2018). En el caso de MRVBF, muestra áreas donde potencialmente se deposita material erosionado que incluyen en la materia orgánica (Chen et al., 2022).

La distribución espacial del índice NDVI se presenta en la figura 11. En la parte norte, presenta los menores valores de este índice, que corresponde a la zona de amortiguamiento del páramo y de mayor intervención agrícola. Los mayores valores se presentan en la zona sur, que corresponde principalmente a matorrales de páramo. La importancia de los índices espectrales radica en que han sido utilizados para determinar la variación espacial de la vegetación (Kunkel et al., 2011), teniendo en cuenta que son utilizados para representar el tipo de vegetación, y esto afecta el almacenamiento del COS (Yang et al., 2018). En donde coberturas como los pastos y cultivos presentan valores bajos de COS, debido a que las actividades antrópicas afectan fuertemente el almacenamiento del COS, especialmente para este ecosistema de páramo (Gutiérrez et al., 2020).

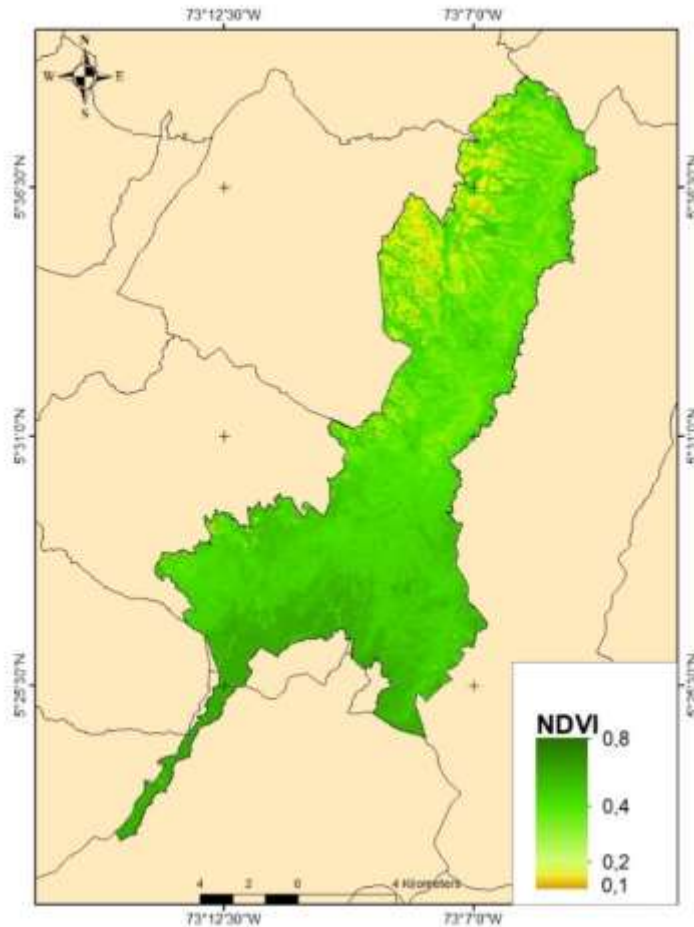


Figura 11 *Distribución espacial del NDVI zona de estudio.*

Los índices espectrales se correlacionan con el COS, especialmente el NDVI con el valor más alto, debido a que no se cuenta en este estudio con una covariable de cobertura de suelo. La utilización del índice NDVI puede sustituir la cobertura de suelo debido a la naturaleza del índice, que representa la relación que tiene, reemplazando en el modelo la cobertura del suelo (Wanga et al., 2017). Teniendo en cuenta su relación con las existencias de COS (Waters et al., 2015 y Orgill et al., 2017). Los avances en los últimos 50 años en cuanto a teledetección del ciclo de carbono, hace énfasis en el avance de los sensores remotos a la hora de cuantificar el carbono en los diferentes ecosistemas, evidenciando la importancia de los índices espectrales en la estimación y cuantificación del COS (Xiao et al., 2019). Esto concuerda con el resultado obtenido entre los datos de COS y su influencia de los índices espectrales, pero no en todos los casos los índices espectrales brindan buena información y dependen mucho de la zona de estudio, pues no

son confiables cuando se tiene un porcentaje mayor del 20 % de suelo desnudo (Sankey y Weber, 2009).

En la figura 12, se puede observar la distribución espacial de la precipitación promedio multianual. Los mayores valores se encuentran en la parte sur con un máximo de 2359 mm y disminuyen hacia el norte hasta un mínimo de 837 mm. La precipitación se relaciona con el COS, ya que influye en el tipo de vegetación y la tasa de descomposición de la materia orgánica (Wang et al., 2014). Los climas con altas precipitaciones y bajas temperaturas, como los páramos, ayudan a una mayor acumulación de COS, debido a la inhibición de los microorganismos para descomponer la materia orgánica (Buol et al., 2011).

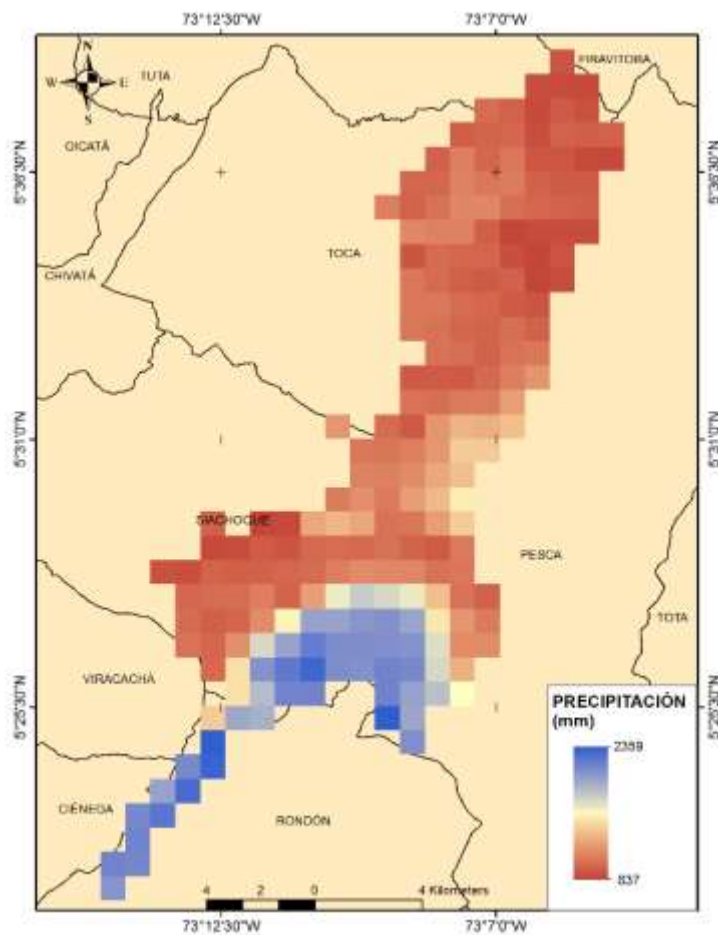


Figura 12 Mapa de distribución espacial de la PPT en la zona de estudio

La temperatura y la precipitación mostraron un grado de significancia con el contenido de COS, pero una limitación de las variables climáticas en su resolución espacial. Aun así, son relevantes para este estudio, concordando con Akpa et. (2016), quienes indicaron que las variables climáticas influyen en la tasa de descomposición de la materia orgánica (MO), en la cantidad de MO y la calidad de MO que circula en los sistemas terrestres. Es por esta razón que se obtiene un valor de significancia que hace que se encuentre entre las covariables a utilizar dentro del modelo en las dos profundidades, y la ppt en el modelo de la profundidad de 15-30 cm.

La covariable VH se observa en la figura 13. Los mayores valores se encuentran en la parte sur, disminuyendo hacia el norte. Los valores de esta variable se asocian con contenidos de humedad del suelo (Hosseini y McNairn, 2017).

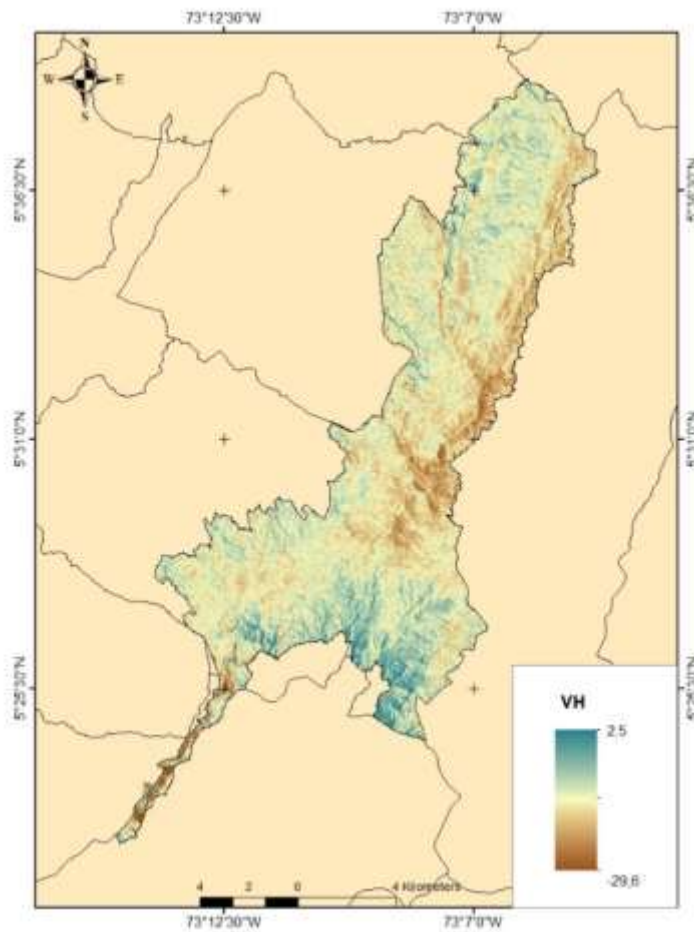


Figura 13 Mapa de la distribución espacial VH zona de estudio.

En el estudio de Zhou (2020), utilizaron la relación VV/VH como covariables que provienen de Sentinel 1, donde se determinó que, en conjunto con las variables provenientes de Sentinel 2 y las variables topográficas, las covariables VV/VH dan buenos resultados en la explicación del COS. Mejoran los resultados en estudios donde se utilizaron solo índices espectrales. Además, muestran mejores resultados con los datos de Sentinel sobre otros datos como Landsat y Modis.

5.1.2 Evaluación de los algoritmos bosques aleatorios y redes neuronales para estimación del COS.

Para estimar el contenido de COS en este estudio, se evaluaron diferentes combinaciones de covariables seleccionadas con base en los factores que inciden en el COS, tales como el clima, vegetación y relieve, que han sido reportadas en otras investigaciones por su potencial predictivo del COS. Igualmente, se tuvo en cuenta la disponibilidad de datos espaciales de fácil acceso. La selección de las covariables tiene gran incidencia en la predicción del COS. Rodríguez-Galiano et al. (2012) Taghizadeh-Mehrjardi et al. (2016); Xie et al. (2015) indican que, para mejorar la capacidad de generalización, acelerar el proceso de aprendizaje automático y aumentar la interpretabilidad de los modelos, se requiere hacer una selección cuidadosa de las covariables.

En esta sección se presenta los resultados de los modelos desarrollados para la estimación del COS para cada una de las profundidades evaluadas.

Para la profundidad de 0-15 cm se evaluaron 13 modelos, cada uno consta de un modelo con bosques aleatorios y otro con redes neuronales, combinando diferentes covariables. En cada caso se excluyeron algunas variables para evaluar su efecto en la predicción del COS.

El primer modelo (Tabla 8) se hizo con bosques aleatorios, incluyendo las covariables que se muestran en la Tabla 7. Se obtuvieron estadísticos de R^2 0.68, el RMSE de 20.7, el RPD de 5.4. Con redes neuronales se obtuvo un R^2 de 0.54, un RMSE de 24.6 y un RPD

de 4.7. Como se puede observar, el modelo con bosques aleatorios mostró ~~unos~~ estadísticos superiores, lo cual indica un mejor desempeño para la estimación del COS a la profundidad de 0-15 cm comparado con el modelo de redes neuronales.

Para el modelo 2 se utilizaron las covariables que mayor correlación presentaron con el COS (Altura, Factor LS, MRVBF, NDVI, Temperatura, TWI, VH). Con el algoritmo de bosques aleatorios se obtuvo un R^2 de 0.67, un RMSE de 21.1 y un RPD de 5.3. mientras que con redes neuronales se encontró un R^2 de 0.58, un RMSE de 23.5 y un RPD de 4.8 (Tabla 8). Al igual que el caso 1, el algoritmo de bosques aleatorios mostró estadísticas superiores a las de redes neuronales.

El modelo 3 se realizó con las covariables Altura, MRVBF, NDVI, Temperatura, TWI y VH, excluyendo el factor LS. Para boques aleatorios, se obtuvo un R^2 de 0.66, un RMSE de 21.3 y un RPD de 5.4, y para redes neuronales, el R^2 fue de 0.55, el RMSE de 23.5 y el RPD de 4.8 (Tabla 8). Al igual que en los modelos 1 y 2, el algoritmo de bosques aleatorios obtuvo mejor desempeño que el de redes neuronales para la predicción del COS.

En el modelo 4 se utilizaron las covariables Altura, Factor LS, NDVI, Temperatura, TWI y VH, se excluyó el MRVBF. Con bosques aleatorios se obtuvo un R^2 de 0.7, un RMSE de 19.9 y un RPD de 5.3, mientras que para las redes neuronales el R^2 fue de 0.6, el RMSE de 22.9 y el RPD de 4.7 (Tabla 8). Como se observa, continúa la tendencia de un mejor desempeño con el algoritmo de bosques aleatorios en relación con las redes neuronales.

En el modelo 5, las covariables utilizadas fueron Altura, NDVI, Temperatura, TWI, y VH, y se excluyeron el factor LS y MRVBF. Con bosques aleatorios se obtuvo un R^2 de 0.65, un RMSE de 21.5 y un RPD de 5.3, y para redes neuronales se obtuvo un R^2 de 0.59, un RMSE de 23.5 y un RPD de 4.7 (Tabla 8). Esto indica que el algoritmo de bosques aleatorios presentó un mejor desempeño que el de las redes neuronales involucradas en este caso.

Para el modelo 6 se efectuó con las covariables de Altura, Factor LS, MRVBF, NDWI, Temperatura, TWI y VH. Se obtuvo para bosques aleatorios un R^2 de 0.62, un RMSE de 22.5 y un RPD de 5.1, mientras que para redes neuronales los valores disminuyeron a un R^2 de 0.27, un RMSE de 31.2 y un RPD de 4.5 (Tabla 8).

Para el modelo 7 se utilizaron las covariables de Altura, Factor LS, MRVBF, MSI, Temperatura, TWI y VH. Con bosques aleatorios se obtuvo un R^2 de 0.63, un RMSE de 22.2 y un RPD de 5.2. En el caso de redes neuronales se obtuvo un R^2 de 0.46, un RMSE de 26.8 y un RPD de 4.7 (Tabla 8).

El modelo 8 incluyó las covariables de Altura, Factor LS, MRVBF, NBRI, Temperatura, TWI y VH. Con bosques aleatorios se obtuvo un R^2 de 0.68, un RMSE de 20.7 y un RPD de 5.4, mientras que las de redes neuronales presentaron un desempeño similar a bosques aleatorios, con un R^2 de 0.67, un RMSE de 21 y un RPD de 4.7 (Tabla 8).

Para el modelo 9 se incluyeron los índices espectrales que se correlacionaron con el COS (NDVI, BSI, EVI, GNDVI, MSI, NBRI, NDMI, NDWI), y la Altura, Factor LS, MRVBF, Temperatura, TWI y VH. En el modelo de bosques aleatorios se obtuvo R^2 de 0.67, un RMSE de 21.1 y un RPD de 5.2, igualmente, en redes neuronales se obtuvo un R^2 de 0.62, un RMSE de 22.4 y un RPD de 4.7 (Tabla 8).

El modelo 10 se hizo con Altura, NDVI, TWI y VH. Para bosques aleatorios se obtuvo un R^2 de 0.64, un RMSE de 22.1 y un RPD de 5.2, mientras que para redes neuronales se obtuvo un R^2 de 0.48, un RMSE de 26.4 y un RPD de 4.8.

El modelo 11 se realizó con NDVI, Temperatura, TWI y VH. Para bosques aleatorios se obtuvo un R^2 de 0.71, un RMSE de 19.6 y un RPD de 5.4, para redes neuronales se obtuvo R^2 de 0.36, un RMSE de 29.2 y un RPD de 4.3.

El modelo 12 se evaluó con el Factor LS, MRVBF, NDVI, Temperatura, TWI y VH. Para el modelo con bosques aleatorios se obtuvo un R^2 de 0.7, un RMSE de 19.9 y un RPD de 5.5, lo mismo que para redes neuronales se obtuvo R^2 de 0.34, un RMSE de 29.6 y un RPD de 4.2.

El modelo 13 se realizó con Altura, Factor LS, MRVBF, NDVI, TWI y VH. Para el modelo con bosques aleatorios se obtuvo un R^2 de 0.66, un RMSE de 21.4 y un RPD de 5.2, mientras que para redes neuronales se obtuvo R^2 de 0.35, un RMSE de 29.3 y un RPD de 4.2.

Como se puede observar en la tabla 8, al comparar los modelos desarrollados, independiente de las covariables utilizadas, se obtuvo un mejor desempeño con los modelos de bosques aleatorios que con los de redes neuronales. Para los modelos 6,11,12 y 13, la diferencia entre estos dos algoritmos fue notoria. Solo en los casos 8 y 9, los modelos de redes neuronales presentaron valores de R^2 similares a los bosques aleatorios, pero los valores de RMSE y RPD fueron mayores. Los tres mejores modelos para estimar el COS, según las métricas estadísticas evaluadas, fueron los 4, 11 y 12 con bosques aleatorios. Estos tres modelos tuvieron en común las covariables temperatura, NDVI, TWI y VH, que corresponden a parámetros climáticos, de relieve, de vegetación y de humedad del suelo. A continuación, se presentan los resultados de la importancia relativa de las covariables en la estimación del COS para los tres modelos.

Tabla 8 Desempeño de cada caso evaluado por bosques aleatorios y redes neuronales para la profundidad 0-15 cm.

Métodos	Bosques aleatorios			Redes neuronales		
	R ²	RMSE	RPD	R ²	RMSE	RPD
Modelo 1 (RF,RN, Todas las variables)	0.68	20.7	5.4	0.54	24.6	4.7
Modelo 2 (RF, RN, Altura, Factor LS MRVBF, NDVI, Temperatura, TWI, VH)	0.67	21.1	5.3	0.58	23.5	4.8
Modelo 3 (Altura, MRVBF, NDVI, Temperatura, TWI, VH)	0.66	21.3	5.4	0.55	24.5	4.8
Modelo 4 (Altura, Factor LS, NDVI, Temperatura, TWI, VH)	0.7	19.9	5.3	0.6	22.9	4.7
Modelo 5 (Altura, NDVI, Temperatura, TWI, VH)	0.65	21.5	5.3	0.59	23.5	4.7
Modelo 6 (Altura, Factor LS, MRVBF, NDWI, Temperatura, TWI, VH)	0.62	22.5	5.1	0.27	31.2	4.5
Modelo 7 (Altura, Factor LS, MRVBF, MSI, Temperatura, TWI, VH)	0.63	22.2	5.2	0.46	26.8	4.7
Modelo 8 (Altura, Factor LS, MRVBF, NBRI, Temperatura, TWI, VH)	0.68	20.7	5.4	0.67	21	4.7
Modelo 9 (Altura, Factor LS, MRVBF, NDVI, BSI, EVI,GNDVI,MSI,NBRI,NDMI,NDWI Temperatura, TWI, VH)	0.67	21.1	5.4	0.62	22.4	4.7
Modelo 10 (Altura, NDVI, TWI, VH,)	0.64	22.1	5.2	0.48	26.4	4.8
Modelo 11 (NDVI, Temperatura, TWI, VH)	0.71	19.6	5.4	0.36	29.2	4.3
Modelo 12 (Factor LS, MRVBF, NDVI, Temperatura, TWI, VH)	0.7	19.9	5.5	0.34	29.6	4.2
Modelo 13 (Altura, Factor LS, MRVBF, NDVI, TWI, VH)	0.66	21.4	5.2	0.35	29.3	4.2

En la figura 14 se muestra la importancia de las covariables para el modelo 4 de bosques aleatorios. La temperatura tuvo la mayor incidencia (27.9% de importancia relativa) en la estimación del COS, seguida por la altura (21.1%), el NDVI (17%) y TWI (17%). Los de menor importancia fueron el VH (8.3%) y el factor LS (7.8%).

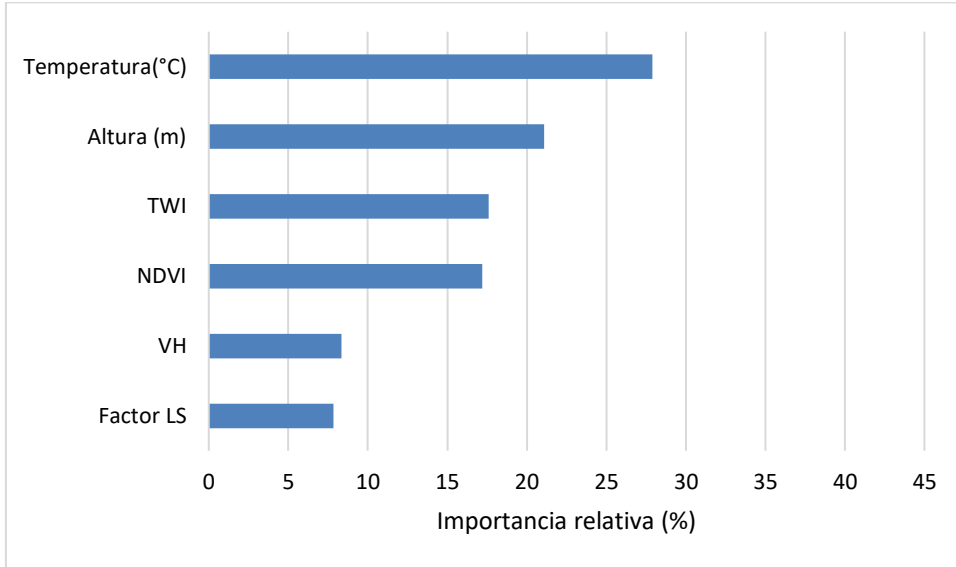


Figura 14 *Importancia relativa de las covariables para el modelo 4 con bosques aleatorios 0-15 cm.*

La importancia de las covariables para el modelo de bosques aleatorios en el modelo 11 (Figura 15) mostró que la temperatura (42.2 %) fue la que tuvo mayor influencia para la estimación del COS, seguida del TWI (24.4%), el NDVI con 21.4% y, por último, el VH (11.9%).

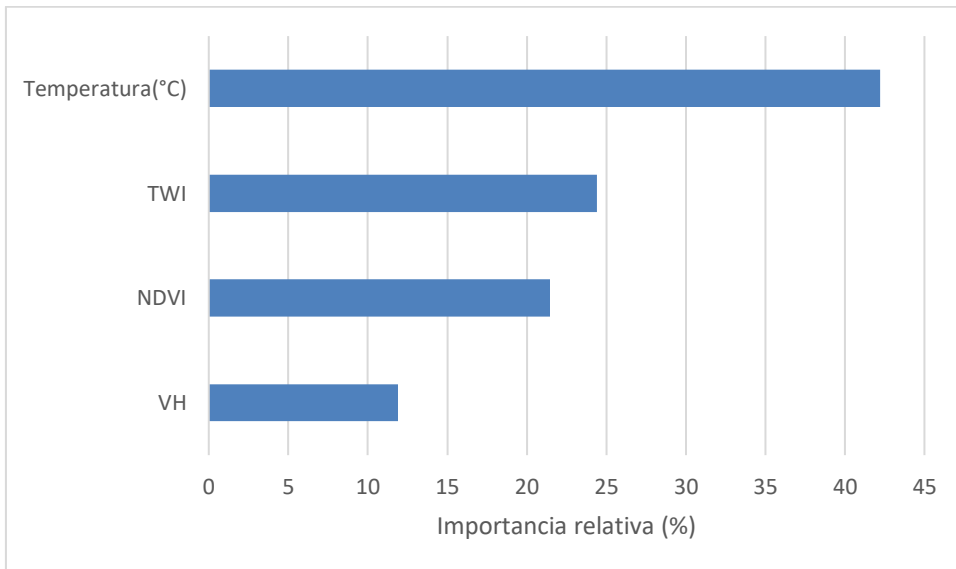


Figura 15 *Importancia covariables en el modelo 11 de bosques aleatorios.*

Para el modelo 12 (Figura 16), la temperatura tuvo mayor influencia para la estimación del COS (39.6%), seguida de TWI (20.2%), el NDVI (17.8%) y con menos del 10 % VH, factor LS y MRVBF.

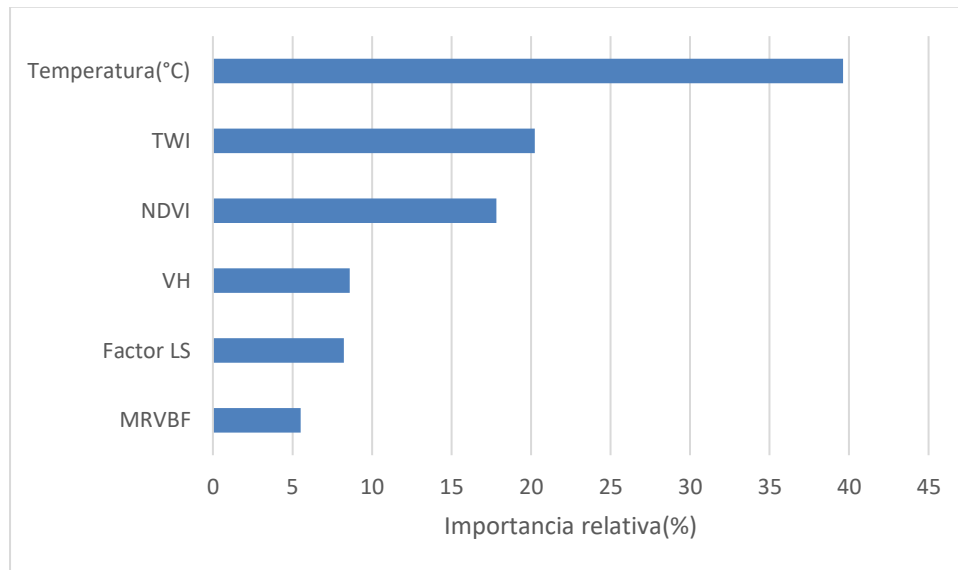


Figura 16 *Importancia covariables modelo 12 bosques aleatorios.*

Con base en los resultados anteriores, se encontró que la temperatura tuvo mayor importancia para la estimación del COS en los 3 modelos considerados como los de mejor desempeño. El TWI y el NDVI tuvieron una importancia relativa intermedia, sin embargo, otras variables presentaron menores grados de importancia relativa. En los modelos 11 y 12, la importancia de la temperatura fue cerca del 40%, y para el modelo 4 disminuyó al 20%, posiblemente debido a la correlación que existe entre la altura y la temperatura, y al excluir la altura, la influencia de la temperatura aumenta.

Con base en los resultados anteriores, se encontró que los tres modelos que mejor desempeño tuvieron permiten hacer una estimación del COS con un patrón de distribución similar para la profundidad de 0-15 cm.

Para la profundidad de 15-30 cm, se evaluaron 12 modelos (Tabla 9), cada uno consta de un modelo con bosques aleatorios y otro con redes neuronales, combinando diferentes covariables.

Para el primer modelo de la profundidad de 15-30 cm, se hizo con bosques aleatorios incluyendo todas las covariables que se muestran en la Tabla 7, obteniendo los

estadísticos de R^2 0.49, el RMSE de 61.9, el RPD de 6.9. Con redes neuronales se obtuvo un R^2 de 0.45, un RMSE de 64.2 y un RPD de 5.7 (Tabla 9).

Para el modelo 2 se utilizaron las covariables Altura, MRVBF, NDVI, Temperatura, PPT, VH. Con el algoritmo de bosques aleatorios se obtuvo un R^2 de 0.57, un RMSE de 56.5 y un RPD de 7, mientras que con redes neuronales se encontró un R^2 de 0.43, un RMSE de 64.2 y un RPD de 5.7 (Tabla 9).

El modelo 3 se realizó con las covariables Altura, NDVI, Temperatura, PPT y VH. Para bosques aleatorios se obtuvo un R^2 de 0.51, un RMSE de 60.8 y un RPD de 6.7, y para redes neuronales, el R^2 fue de 0.49, el RMSE de 62.1 y el RPD de 5.5 (Tabla 9).

En el modelo 4 se utilizaron las covariables Altura, NDVI, Temperatura y VH. Con bosques aleatorios se obtuvo un R^2 de 0.56, un RMSE de 57.4 y un RPD de 6.5, mientras que, para las redes neuronales, el R^2 fue de 0.37, el RMSE de 68.7 y el RPD de 5.5 (Tabla 9).

En el modelo 5, las covariables utilizadas fueron Altura, MRVBF, NDVI, Temperatura, y VH. Con bosques aleatorios se obtuvo un R^2 de 0.59, un RMSE de 55.3 y un RPD de 7, y para redes neuronales se obtuvo un R^2 de 0.33, un RMSE de 70.5 y un RPD de 5.7 (Tabla 9).

Para el modelo 6 se utilizaron las covariables de Altura, MRVBF, GNDVI, Temperatura, PPT y VH, obteniendo con bosques aleatorios un R^2 de 0.57, un RMSE de 56.7 y un RPD de 7.1, mientras que para redes neuronales disminuyeron los valores con R^2 de 0.4, un RMSE de 66.9 y un RPD de 5.8 (Tabla 9).

Para el modelo 7 se incluyeron los índices espectrales, se utilizaron las covariables Altura, MRVBF, NDWI, Temperatura, PPT y VH. Con bosques aleatorias se obtuvo un R^2 de 0.56 un RMSE de 57.1 y un RPD de 6.9. En el caso de redes neuronales se obtuvo un R^2 de 0.32, un RMSE de 71.4 y un RPD de 5.3 (Tabla 9).

El modelo 8 se efectuó con las covariables de Altura, MRVBF, NDVI, NDWI, GNDVI, Temperatura, PPT y VH. Con bosques aleatorios se obtuvo un R^2 de 0.6, un RMSE de 54.7 y un RPD de 7.1, de igual manera las de redes neuronales presentaron un desempeño

menor a los bosques aleatorios, con un R^2 de 0.5, un RMSE de 61 y un RPD de 5.8 (Tabla 9).

El modelo 9 se realizó con las covariables Altura, MRVBF, NDVI y VH. Para bosques aleatorios se obtuvo R^2 de 0.59, un RMSE de 55.6 y un RPD de 6.7, para redes neuronales se obtuvo un R^2 de 0.42, un RMSE de 66.1 y un RPD de 6 (Tabla 9).

El modelo 10 se hizo con MRVBF, NDVI, Temperatura y VH, para bosques aleatorios se obtuvo un R^2 de 0.57, un RMSE de 57.1 y un RPD de 7, para redes neuronales se obtuvo R^2 de 0.1, un RMSE de 83.7 y un RPD de 5.1.

El modelo 11 se realizó con MRVBF, NDVI, Temperatura, PPT y VH, para bosques aleatorios se obtuvo un R^2 de 0.53, un RMSE de 59.6 y un RPD de 7, para redes neuronales se obtuvo R^2 de 0.16, un RMSE de 79.6 y un RPD de 5.2.

El modelo 12 se evaluó con la Altura, MRVBF, NDVI, PPT y VH, para el modelo con bosques aleatorios se obtuvo un R^2 de 0.58, un RMSE de 56.3 y un RPD de 6.7, para redes neuronales se obtuvo R^2 de 0.46, un RMSE de 63.6 y un RPD de 6.

Para los modelos evaluados para la profundidad de 15-30 cm, con bosques aleatorios los modelos fueron los de mejor desempeño. El de menor desempeño fue el caso 1, que utilizó todas las variables, con R^2 de 0.49, un RMSE de 61.9 y un RPD de 6.9 (Tabla 8).

Los tres mejores modelos para estimar el COS para la profundidad de 15-30 cm, según las métricas estadísticas, fueron los modelos 5, 8 y 9 con bosques aleatorios. Estos tres modelos tuvieron en común las covariables alturas, MRVBF, NDVI y VH, relacionadas con parámetros de relieve, de vegetación y de humedad del suelo. Al igual que ocurrió para la profundidad de 0-15 cm, las estimaciones basadas en bosques aleatorios presentaron mejor desempeño que aquellas con redes neuronales.

Tabla 9 Desempeño de cada modelo evaluado por bosques aleatorios y redes neuronales para la profundidad 15-30 cm.

Métodos	Bosques aleatorios			Redes neuronales		
	R ²	RMSE	RPD	R ²	RMSE	RPD
Modelo 1 (Todas las variables)	0.49	61.9	6.9	0.45	64.2	5.7
Modelo 2 (Altura, MRVBF, NDVI, Temperatura, PPT, VH)	0.57	56.5	7	0.43	65	5.8
Modelo 3 (Altura, NDVI, Temperatura, PPT, VH)	0.51	60.8	6.7	0.49	62.1	5.5
Modelo 4 (Altura, NDVI, Temperatura, VH)	0.56	57.4	6.5	0.37	68.7	5.5
Modelo 5 (Altura, MRVBF, NDVI, Temperatura, VH)	0.59	55.3	7	0.33	70.5	5.7
Modelo 6 (Altura, MRVBF, GNDVI, Temperatura, PPT, VH)	0.57	56.7	7.1	0.4	66.9	5.8
Modelo 7 (Altura, MRVBF, NDWI, Temperatura, PPT, VH)	0.56	57.1	6.9	0.32	71.4	5.3
Modelo 8 (Altura, MRVBF, NDVI, GNDVI, NDWI, Temperatura, PPT, VH)	0.6	54.7	7.1	0.5	61	5.8
Modelo 9 (Altura,MRVBF, NDVI, VH)	0.59	55.6	6.7	0.42	66.1	6
Modelo 10 (MRVBF, NDVI, Temperatura, VH)	0.57	57.1	7	0.1	83.7	5.1
Modelo 11 (MRVBF, NDVI, Temperatura, PPT, VH)	0.53	59.6	7	0.16	79.6	5.2
Modelo 12 (Altura, MRVBF, NDVI, PPT, VH)	0.58	56.3	6.7	0.46	63.6	6

En esta investigación, se encontró que el algoritmo de bosques aleatorios fue el de mejor desempeño con base en los estadísticos R², RMSE y RPD para estimar el contenido del COS, independientemente de la combinación de las covariables. Según la revisión de Chen et al. (2020), el 80.3 % de los estudios para estimar alguna propiedad del suelo utilizaron el R², un indicador adimensional y apropiado del desempeño de los modelos. También se ha empleado el RMSE y el RPD como indicadores del desempeño de los modelos. El algoritmo de bosques aleatorios logró una estimación del COS explicando entre el 60 y 70%, resultados similares a lo reportado han sido corroborados por Dong et al. (2019) y Gomes et al. (2019), en comparación con las redes neuronales, que obtuvieron una estimación entre el 40 y el 50%.

En otros estudios (Mouazen et al., 2010), indican que para la estimación del COS los algoritmos de aprendizaje automático tienen mejor desempeño que los métodos estadísticos como la regresión lineal paso a paso, la regresión de componentes principales y la regresión de mínimos cuadrados parciales, debido a las complejas relaciones no lineales. Bou et al. (2010) utilizaron con éxito árboles de clasificación y regresión (CART) para modelar patrones de distribución de COS con variables de teledetección en

Dinamarca, mientras que Were et al. (2015) compararon el rendimiento de la regresión de vectores de soporte (SVR), las redes neuronales artificiales (ANN) y el bosque aleatorio para predecir y mapear el COS.

Wang et al. (2017) obtuvieron el 39-65% de predicción del COS en el noreste de China utilizando un modelo basado en Bootstrap Random Forest (BRT), y Wang et al. (2018) utilizando bosques aleatorios explicaron el 48 %, siendo mayor a lo obtenido en modelo de BRT. Estos valores contrastan con el estudio de Roman et al. (2018), que solo lograron explicar el 18% utilizando bosques aleatorios. En el estudio de Wang et al. (2018) se concluyó que los modelos con bosques aleatorios y BRT eran superiores a los demás modelos como SVM, al igual que en Siewert (2018); Forkour et al. (2107); Yang et al. (2016) y Akpa et al. (2016), en todos estos estudios se compraron varios métodos de aprendizaje automatizado siendo los bosques aleatorios los de mejor desempeño. (Lamichane et al. 2019) concluyen que los bosques aleatorios superan a otros métodos principalmente cuando es un área de estudio grande y una densidad de muestreo baja y Forkour et al. (2017) indican que ningún algoritmo de aprendizaje automatizado podría servir para todos los casos y que por lo tanto los modelos se deben calibrar y evaluar para identificar el más apropiado para cada situación en especial.

Las redes neuronales han sido utilizadas con éxito en estudios como Taghizadeh et al. (2016); Were et al. (2015) y Mirzaee et al. (2016) comparado únicamente con la regresión lineal y no con otros métodos de aprendizaje automatizado.

En la figura 17 se muestra la importancia de las covariables para el modelo de bosques aleatorios del modelo 5. La temperatura tuvo la mayor incidencia (52.5% importancia relativa) en la estimación del COS, seguida de la altura (16.7%), el MRVBF (12.6%), el NDVI (10.5%) y por último VH (7.8).

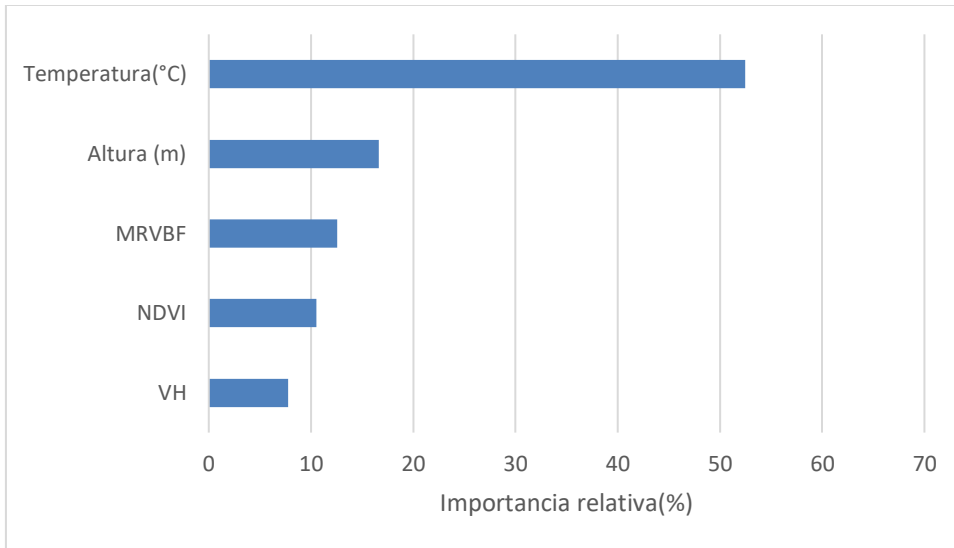


Figura 17 *Importancia relativa de las covariables para el modelo 5 con bosques aleatorios 15-30 cm.*

Para el modelo 8 (figura 18), la temperatura tuvo mayor influencia (51.8%), seguida de la altura (13.2%), el MRVBF (10.6), el NDVI (7.1%), y con menos del 10 % VH, NDWI, GNDVI y PPT para la estimación del COS.

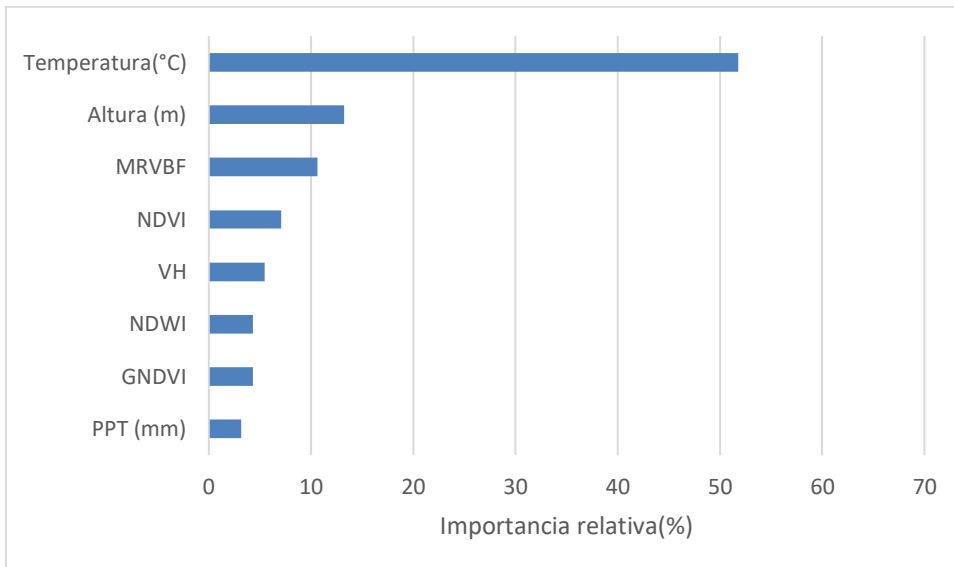


Figura 18 *Importancia relativa de las covariables para el modelo 8 con bosques aleatorios 15-30 cm.*

La importancia de las covariables para el modelo de bosques aleatorios en el modelo 9 (figura 19), mostró que la altura fue la covariabile que mayor incidencia tuvo (62%) en la estimación del COS, seguida del NDVI (15.6%), el MRVBF (13.1%), y por último VH (9.3%).

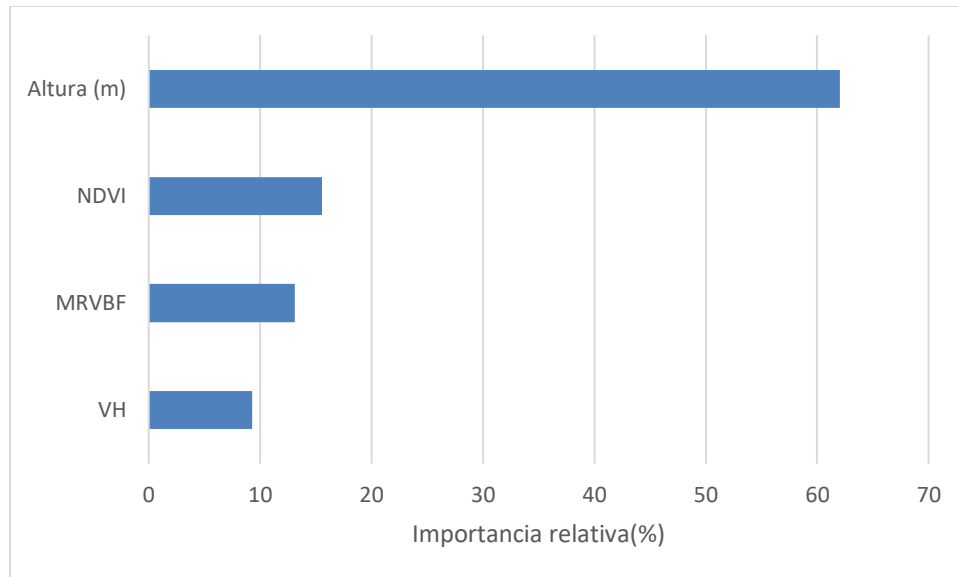


Figura 19 *Importancia relativa de las covariables para el caso 5 con bosques aleatorios 15-30 cm.*

La importancia relativa de las covariables para los modelos de mejor desempeño fueron la temperatura y la altura, seguidas del TWI, MRVBF, el NDVI y en menor medida VH. La incidencia de los parámetros morfométricos y la altura, concuerdan con lo reportado en Davy y Koen (2013); Reyes et al. (2018); Siewert (2018) y Zhou et al. (2020), que señalaron que la topografía afecta el almacenamiento de COS mediante sus efectos en la profundidad y humedad del suelo, que a su vez se relacionan con la retención de agua y nutrientes, que inciden en el crecimiento y tipo de vegetación que es la fuente de materia orgánica y por lo tanto determinan el contenido de COS (Wang et al., 2018). El efecto de la altura como variable predictora del COS ha sido reportada en otros estudios como Adhikari et al. (2014); Obu et al. (2015); Ramifehiarivo et al. (2017); Wang et al. (2018); Hinge et al. (2018) y Chen et al. (2022), quienes además de la altura encontraron incidencia del TWI y el MRVBF.

Los climas con altas precipitaciones y bajas temperaturas, como los páramos, ayudan a una mayor acumulación de COS, debido a la inhibición de los microorganismos para descomponer la materia orgánica (Buol et al. (2011), esto posiblemente explica la alta importancia relativa de la temperatura concordando con Gray et al. (2015); Hobbey et al. (2015); Ramifehiarvo et al. (2016); Sayao y Dematte (2018) y Wang et al. (2018). Por su parte Akpa et. (2016) indicaron que las variables climáticas inciden en la tasa de descomposición, en la cantidad y calidad de la materia orgánica que circula en los sistemas

terrestres. La altura y la temperatura fueron las dos covariables que más influyeron en la estimación del COS y la dos inciden en la descomposición y mineralización de la materia orgánica (Quesada et al.,2020; Huamán et al., 2021; Hunt et al., 2020).

El NDVI tuvo una importancia media entre el 15-20% en los modelos. Wang et al. (2018); Were et al. (2015); Zhou et al. (2020), concluyeron que el NDVI era de los índices más utilizados para la explicación del contenido COS y que la combinación de diferentes imágenes tanto espectrales como de radar mejoraban el desempeño de los modelos predictivos. En algunos casos los índices espectrales no brindan buena información para predecir el COS y depende mucho de la zona de estudio, algunos autores como Sankey y Weber, (2009) consideran que no son confiables cuando se tiene un porcentaje mayor al 20 % de suelo desnudo, Wang et al. (2018) encontraron que para un ambiente semiárido el NDVI no fue importante para explicar el COS. Yang et al. (2016); Mahmudabadi et al. (2017) y Schialli et al. (2017) afirman que el NDVI fue importante a la hora de predecir el COS cuando se utilizó con parámetros topográficos y variables climáticas.

La covariable VH se relaciona con la humedad en los suelos, lo cual influye en el contenido del COS y en conjunto con los índices espectrales y los parámetros morfométricos, VH influye en la explicación del COS, mejorando los resultados en estudios donde se utilizaron solo índices espectrales según lo reportado por Zhou et al. (2020).

5.1.3 Validación del método

De acuerdo a los tres modelos que tuvieron mejor desempeño, se realizó la validación primero para la profundidad de 0-15 cm. Como se muestra en la tabla 10, el resultado de las métricas que permiten evaluar el desempeño tanto para el conjunto de prueba, como para el de entrenamiento y la validación cruzada para los modelos en los 3 casos de mejor desempeño de los bosques aleatorios. Se observó que las diferencias entre cada una no son significativas, concordando con los resultados obtenidos en las métricas de desempeño de la tabla 10.

Tabla 10 Estadísticas COS predicho 0-15 cm.

	R ² prueba	R ² cv	R ² entrenamiento	RMSE prueba	RMSE cv	RMSE entrenamiento
Caso 4	0.7	0.31	0.84	19.92	36.7	18.31
Caso 11	0.71	0.32	0.84	19.61	36.16	18.54
Caso 12	0.7	0.36	0.85	19.9	35.33	18.02

En la figura 20, se puede observar el gráfico de dispersión entre los valores predichos y los valores observados para el conjunto de prueba (n=51) para el modelo del caso 4, ajustándose mejor en los valores bajos-medios del COS 0-15 cm y disminuyendo este ajuste en los valores más altos.

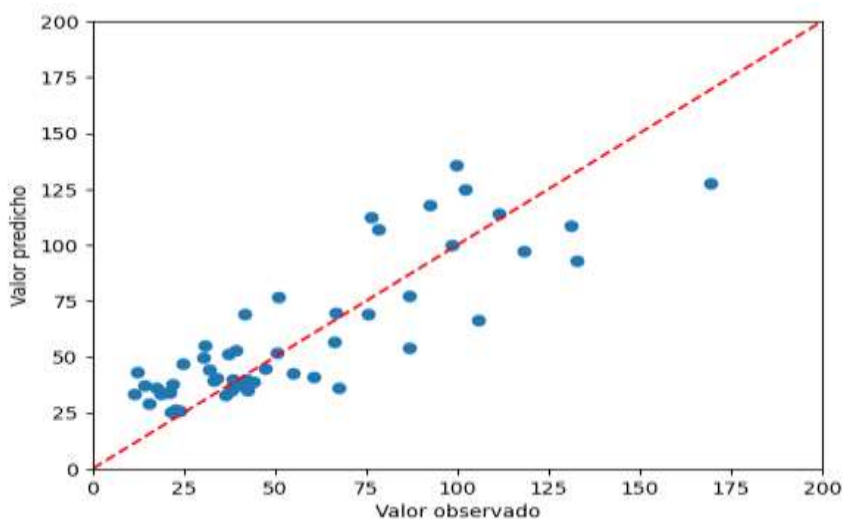


Figura 20 Gráfico de dispersión de los valores predichos/observados para el caso 4 bosques aleatorios 0-15 cm.

En la figura 21, se observa el gráfico de dispersión entre los valores predichos y los valores observados para el conjunto de prueba (n=51) para el caso 11, para este modelo al igual que en el caso 4 se ajustó mejor en los valores bajos-medios del COS 0-15 cm y disminuyendo este ajuste en los valores más altos, mejorando en los valores medios en comparación del caso 4.

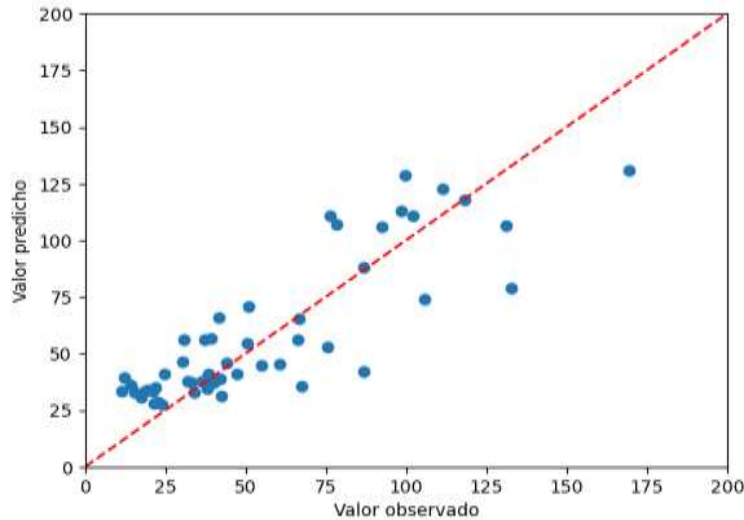


Figura 21 *Gráfico de dispersión de los valores predichos/observados para el caso 11 bosques aleatorios 0-15 cm.*

Para el modelo del caso 12, se muestra en la figura 22 el gráfico de dispersión entre los valores predichos y los valores observados para el conjunto de prueba (n=51), para este caso solo se observa un buen ajuste en los valores bajos ya que en los valores medios y altos los puntos de dispersión están más alejados del valor ideal, lo cual se diferencia en este caso en comparación con los modelos de los dos casos anteriores.

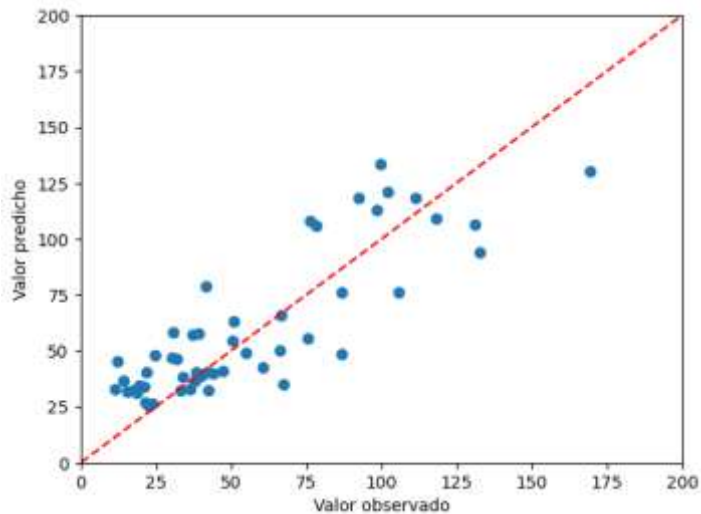


Figura 22 *Gráfico de dispersión de los valores predichos/observados para el caso 12 bosques aleatorios 0-15 cm.*

Para la profundidad de 15-30 cm se realizó la tabla 11, que permite evaluar el desempeño tanto para el conjunto de prueba, como para el de entrenamiento y la validación cruzada

para los modelos en los 3 casos de mejor desempeño de los bosques aleatorios. Se observó que las diferencias entre cada una no son significativas, concordando con los resultados obtenidos en las métricas de desempeño de la tabla 8.

Tabla 11 Estadísticas COS predicho 15-30 cm.

	R ² prueba	R ² cv	R ² entrenamiento	RMSE prueba	RMSE cv	RMSE entrenamiento
Caso 5	0.59	0.45	0.87	55.35	65.46	32.61
Caso 8	0.60	0.47	0.92	54.70	64.21	24.94
Caso 9	0.59	0.39	0.87	55.6	68.37	32.82

En la figura 23, se puede observar el gráfico de dispersión entre los valores predichos y los valores observados para el conjunto de prueba (n=51) para el modelo del caso 5, con una dispersión uniforme tanto en los valores bajos como en los valores altos.

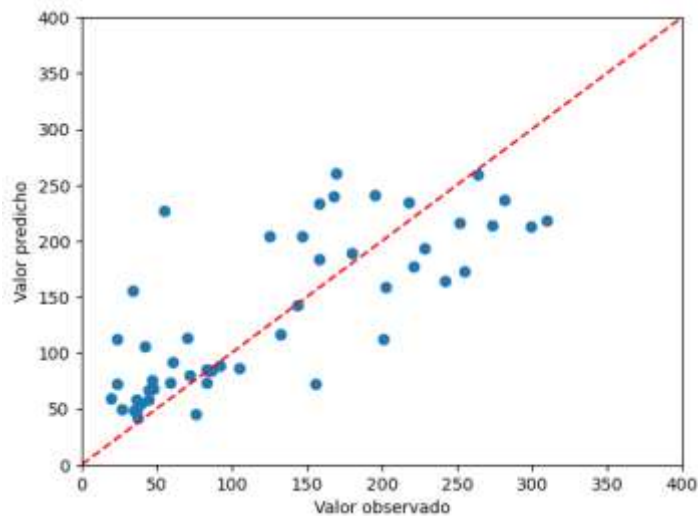


Figura 23 Gráfico de dispersión de los valores predichos/observados para el caso 5 bosques aleatorios 15-30 cm.

En la figura 24, se observa el gráfico de dispersión entre los valores predichos y los valores observados para el conjunto de prueba (n=51) para el caso 8, en este modelo se observó un leve mejor ajuste en algunos de los valores altos en comparación con el caso 5.

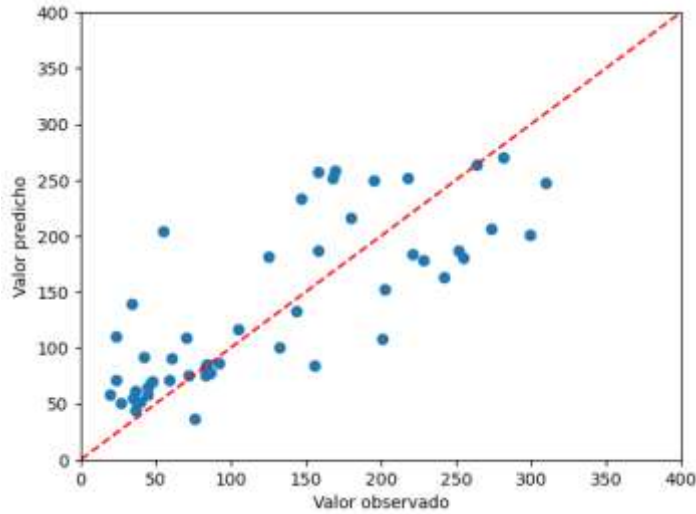


Figura 24 *Gráfico de dispersión de los valores predichos/observados para el caso 8 bosques aleatorios 15-30 cm.*

Para el modelo del caso 9, se muestra en la figura 25 el gráfico de dispersión entre los valores predichos y los valores observados para el conjunto de prueba (n=51), se observó un ajuste similar al del caso 5.

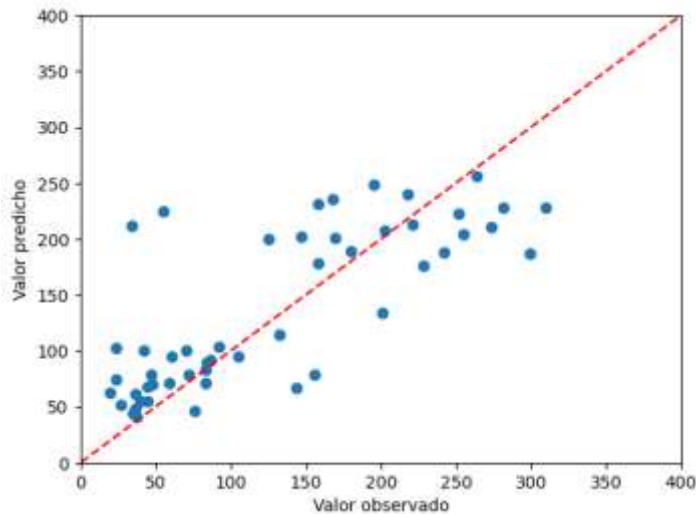


Figura 25 *Gráfico de dispersión de los valores predichos/observados para el caso 9 bosques aleatorios 15-30 cm.*

La validación del método mediante división de datos mostró buenos resultados debido a que como se planteó en este estudio, se predijo el COS una gran cantidad de puntos que contenían datos de las covariables por lo cual la validación cruzada por su naturaleza no se tomó como método principal, esto concuerda con lo expuesto por Lamichane et al.

(2019), que concluyeron que la validación por división de datos es un progreso en el contexto de la distribución espacial del COS, además, de generar estimaciones imparciales y validas cuando el conjunto de datos de prueba hace parte del muestreo (Brus et al. 2011).

5.1.4 Estimación del método desarrollado

En las figuras 26, 27 y 28 se presentan la estimación de la distribución espacial del COS para la profundidad de 0-15 cm, correspondiente a los modelos 4,11 y 12, respectivamente. Los valores estimados variaron entre 24 -170 t ha⁻¹, siendo similares para los tres modelos. Los menores valores se presentaron en la zona de mayor intervención del páramo, que corresponde a la parte norte con gran actividad antrópica. Los contenidos intermedios y altos de COS se encontraron en la zona de menor intervención y mayor altura, posiblemente debido a una mayor mineralización del COS por el efecto del uso y la temperatura.

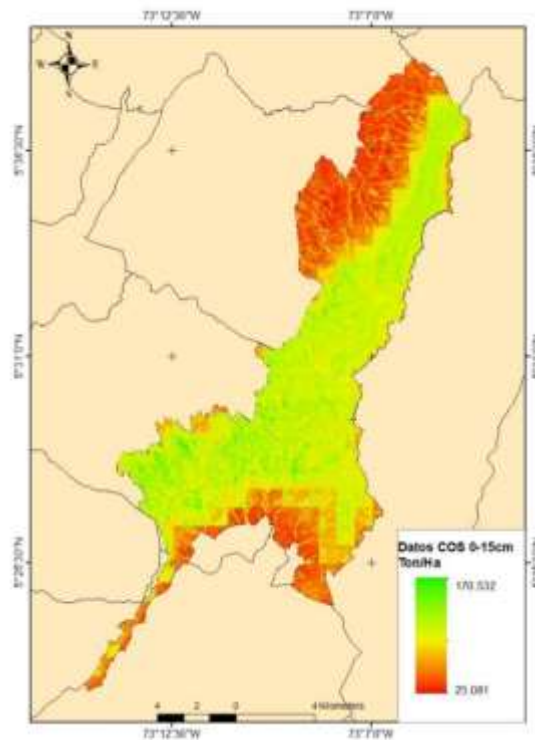


Figura 26 Distribución espacial del COS predicho caso 4 con bosques aleatorios 0-15 cm.

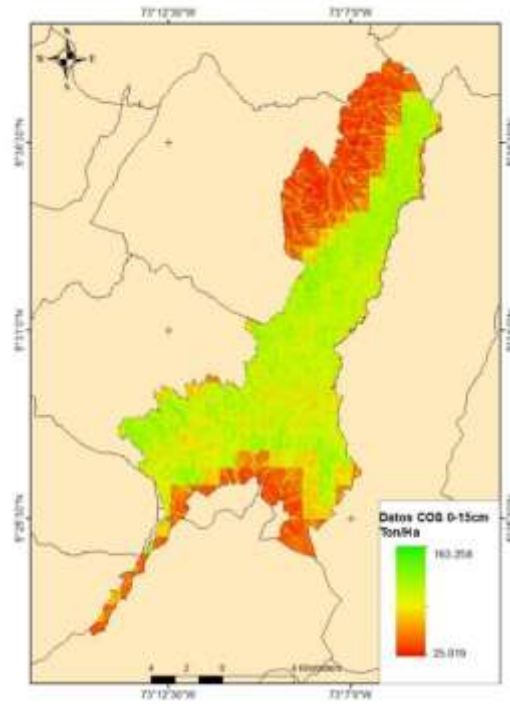


Figura 27 Distribución espacial del COS predicho modelo 11 con bosques aleatorios 0-15 cm.

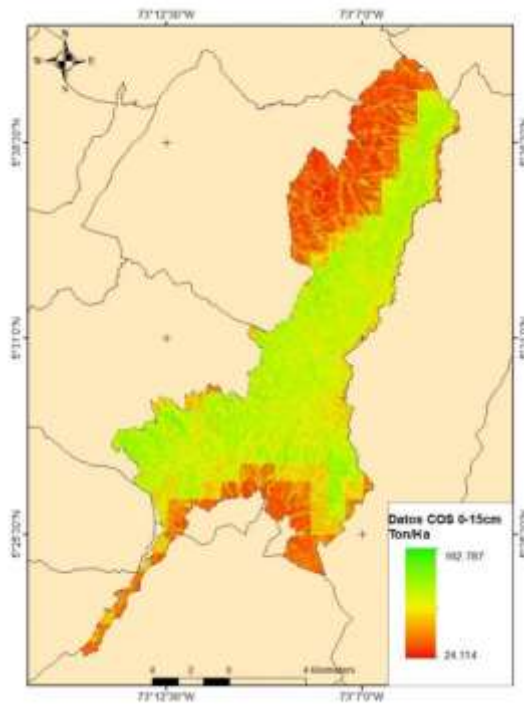


Figura 28 Distribución espacial del COS predicho modelo 12 con bosques aleatorios 0-15 cm.

La estimación de distribución espacial para los modelos 5, 8 y 9 se muestran en las figuras 29, 30, y 31, para la profundidad de 15-30 cm. Se presentó la misma tendencia que en las estimaciones de la primera profundidad (0-15 cm), se mantuvo la tendencia para la estimación de esta segunda profundidad que presentó los menores valores del COS en la zona de amortiguación del páramo y los mayores valores en las partes más altas y menos intervenidas.

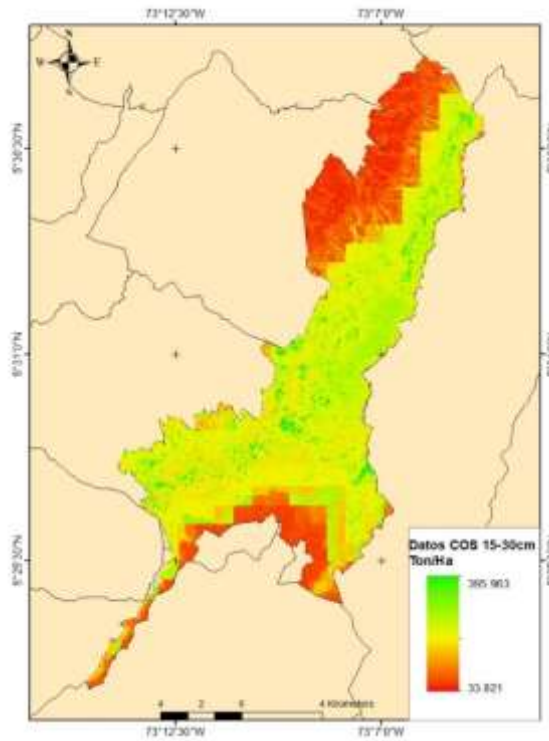


Figura 29 Distribución espacial del COS predicho modelo 5 con bosques aleatorios 15-30 cm.

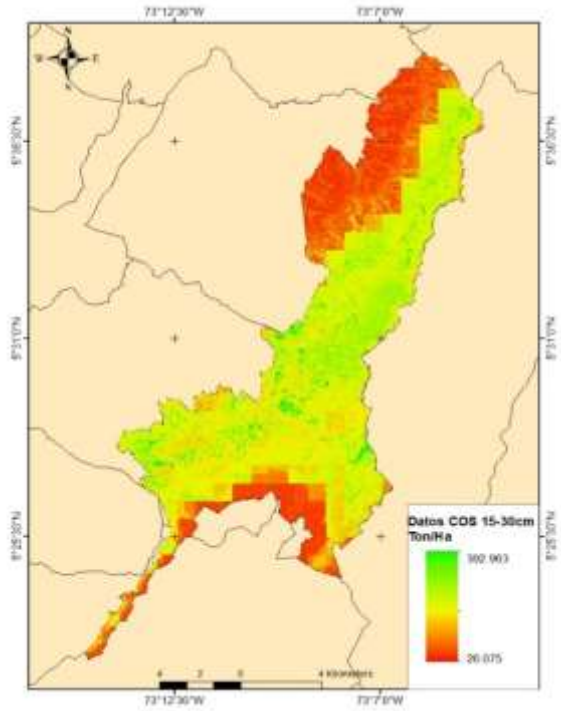


Figura 30 Distribución espacial del COS predicho modelo 8 con bosques aleatorios 15-30 cm.

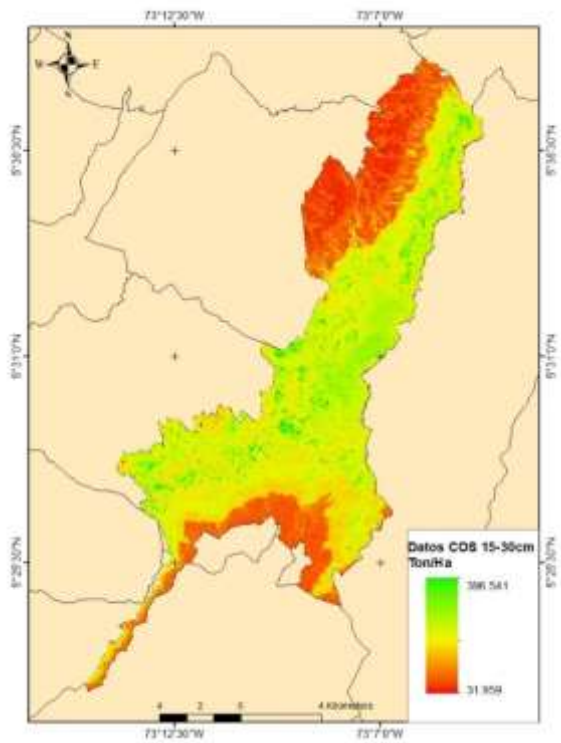


Figura 31 Distribución espacial del COS predicho modelo 9 con bosques aleatorios 15-30 cm.

La estimación de la distribución espacial del COS presentó los mayores contenidos del COS en las zonas de mayor altura, con menores temperaturas y menor grado de intervención del páramo. Griffiths et al. (2009) concluyen que a medida que incrementa la altitud, hay menor temperatura y por lo tanto mayor capacidad de almacenamiento de COS. Las áreas de mayor intervención con cobertura de pastos y cultivos presentaron menores valores de COS que corresponde a la zona de amortiguación del páramo. La cobertura vegetal es uno de los factores que afecta el almacenamiento del COS (Waters et al., 2015), razón por la cual el NDVI se consideró como un buen indicador de la cobertura vegetal lo que coincide con lo reportado por Wang et al. (2018).

Los resultados de los procesos planteados para desarrollar el método de la estimación del COS en el suelo de páramo, mostraron importantes generalidades que se deben destacar, los modelos de mejor desempeño tienen ciertas características en común, que son acordes a lo revisado en otras investigaciones, como los factores climáticos y factores morfométricos que influyen en el COS, especialmente la relación altura-temperatura, es así como es necesario tener en cuenta estos factores en cualquier método que se vaya a desarrollar, independientemente de la fuente de datos, se hace necesario contar con estas covariables.

Además, los factores de uso y cobertura del suelo mostraron incidencia en el contenido del COS, en este caso representados por los índices espectrales, lo cual se hace necesario al desarrollar un método, contar con covariables de índices espectrales, por lo menos con el NDVI por ser el índice más conocido y utilizado, además de ser uno de los más fáciles de calcular, pero se tiene que tener cuidado con todo el proceso, debido a que el éxito de los índices espectrales para el método, está en que no tenga problemas de nubosidad pues puede hacer que los valores estimados estén errados si no se tiene en cuenta estas características propias de los sensores. Caso contrario de lo que ocurre con la covariables de VH-VV de Sentinel 1, las cuales no les afecta la nubosidad y pueden ser de apoyo al método, aunque no han sido muy utilizadas para estimar el COS, la tendencia actual y el factor de humedad en el suelo que representan, vuelve este factor de relevancia para tener en cuenta.

Los factores a tener en cuenta para el método se basaron en factores climáticos, parámetros morfométricos, uso, cobertura y humedad del suelo, las fuentes de datos son

las que convierten al método en un estudio de escala regional por su nivel de detalle, además por la misma extensión del páramo que está presente en varios municipios del departamento. Este método es útil y necesario por su bajo costo, además que se convierte en la base para poder planificar el manejo y la conservación de este ecosistema, pues muestra las zonas críticas de menores valores que están siendo intervenidas, lo que se ve reflejado en la distribución espacial de la estimación.

Como resultado de esta investigación se propone el uso de bosques aleatorios para hacer la predicción del COS. Otros investigadores Adhikari et al. (2014); Hengl et al. (2015); Kuang et al. (2015); Sreenivas et al. (2016); Were et al. (2015) resaltaron como ventajas de los bosques aleatorios que se requieren menos parámetros y que permiten investigar relaciones no lineales entre los predictores y la respuesta. Los métodos de regresión lineal son los más comunes (Thompson et al., 2006; Zhao et al., 2014) por ser simples y sencillos de aplicar, usar e interpretar (Thompson et al., 2006), sin embargo, estos métodos asumen una relación lineal entre el COS y las covariables, aunque en realidad, estas relaciones suelen ser complejas y no lineales. Para el buen desempeño de los modelos la selección de las variables es una etapa muy importante, ya que influye en gran medida en las predicciones. El uso de datos disponibles y de fácil acceso como los utilizados en esta investigación permiten una estimación apropiada del COS.

Sabiendo cuales son estos factores que afectan la estimación espacial del COS, se pueden orientar medidas de mitigación en el suelo de páramo con base justamente en los factores que afectan el servicio ecosistémico que brinda, además de relacionar las dinámicas socioeconómicas que se dan en las zonas de amortiguación del área de estudio que reflejan la realidad de estos ecosistemas que han tenido mucha presión ambiental.

6. Conclusiones

La selección de las variables es muy importante a la hora de establecer el método para estimar la distribución espacial del COS, encontrándose que aquellas relacionadas con los factores que inciden en el contenido del COS, como el relieve, el clima y la vegetación, son las que permiten una mejor predicción de la distribución espacial del COS.

Los bosques aleatorios fue el algoritmo de aprendizaje automatizado de mejor desempeño comparado con redes neuronales para la estimación del COS. Las variables de altura, temperatura, TWI, MRVBF, NDVI y VH, fueron las que mayor importancia relativa presentaron para estimar el COS, éstas representan los factores climáticos, topográficos, uso y cobertura del suelo y humedad del suelo, que son los que inciden en los valores del contenido del COS.

El método desarrollado permitió estimar la distribución espacial del contenido de COS en el suelo de páramo, identificando las zonas de menor y mayor contenido del COS, convirtiéndose en información vital para el ordenamiento y planificación territorial con el fin de mitigar la presión ambiental sobre el páramo.

El método desarrollado es un aporte con el fin de obtener como resultado buenas estimaciones espaciales que sean base para planear y ordenar el territorio y además de esto mitigar algunos efectos del cambio climático en espacial con los GEI.

Bibliografía

- Abuín, J. R. (2007). Regresión lineal múltiple. *IdEyGdM-Ld Estadística, Editor*, 32.
- Adhikari, K., Hartemink, A. E., Minasny, B., Bou Kheir, R., Greve, M. B., & Greve, M. H. (2014). Digital mapping of soil organic carbon contents and stocks in Denmark. *PLoS ONE*, 9(8). <https://doi.org/10.1371/journal.pone.0105519>
- Africano, K., Cely, G., & Serrano, P. (2016). *Potential CO2 Capture Associated with Edaphic Component in Moorlands Guantiva - La Rusia, Department of Boyacá, Colombia*.
- Akpa, S. I. C., Odeh, I. O. A., Bishop, T. F. A., Hartemink, A. E., & Amapu, I. Y. (2016). Total soil organic carbon and carbon sequestration potential in Nigeria. *Geoderma*, 271, 202–215. <https://doi.org/10.1016/j.geoderma.2016.02.021>
- Alberto Quesada, C., Paz, C., Oblitas Mendoza, E., Lawrence Phillips, O., Saiz, G., & Lloyd, J. (2020). Variations in soil chemical and physical properties explain basin-wide Amazon forest soil carbon concentrations. *SOIL*, 6(1), 53–88. <https://doi.org/10.5194/soil-6-53-2020>
- Armas, D., Guevara, M., Alcaraz-Segura, D., Vargas, R., Soriano-Luna, M. A., Durante, P., & Oyonarte, C. (2017). Mapa digital del perfil del carbono orgánico en los suelos de Andalucía, España. *Ecosistemas*, 26(3), 80–88. <https://doi.org/10.7818/ECOS.2017.26-3.10>
- Ayala, J. (2019). *Mapeo digital de carbono orgánico del suelo mediante imágenes satelitales y algoritmos de autoaprendizaje en el ecosistema Herbazal del Páramo, provincia de Chimborazo, Ecuador* (Doctoral dissertation, Universidad Nacional de La Plata).
- Ayala, L., Villa, M., Aguirre, Z., & Aguirre, N. (2014). Quantification of carbon in the moors of the Yacuri National Park, provinces of Loja and Zamora Chinchipe, Ecuador. In *CEDAMAZ 2014* · (Vol. 4, Issue 1).
- Bannari, A., Morin, D., Bonn, F., & Huete, A. R. (1995). A review of vegetation indices. *Remote Sensing Reviews*, 13(1–2), 95–120. <https://doi.org/10.1080/02757259509532298>
- Bernardeau, A., Sarría, F., & Castillo, F. J. (2014). Elaboración de un mapa de carbono orgánico del suelo en la Región de Murcia. *XVI Congreso Nacional de Tecnologías de la Información Geográfica*, Universidad de Alicante.

- Behrens, T., Zhu, A. X., Schmidt, K., & Scholten, T. (2010). Multi-scale digital terrain analysis and feature selection for digital soil mapping. *Geoderma*, 155(3–4), 175–185. <https://doi.org/10.1016/j.geoderma.2009.07.010>
- Blanco, A., Sarría, F., & Castillo, F. (2014). *Elaboración de un mapa de carbono orgánico del suelo en la Región de Murcia*.
- Breiman, L. (2001). Random forests. *Machine learning*, 45 (1), 5-32.
- Brus, D., Kempen, B., & Heuvelink, G. (2011). Sampling for validation of digital soil maps. *European Journal of Soil Science*, 62(3), 394-407.
- Bou Kheir, R., Greve, M. H., Bøcher, P. K., Greve, M. B., Larsen, R., & McCloy, K. (2010). Predictive mapping of soil organic carbon in wet cultivated lands using classification-tree based models: The case study of Denmark. *Journal of Environmental Management*, 91(5), 1150–1160. <https://doi.org/10.1016/j.jenvman.2010.01.001>
- Buol, S. W., Southard, R. J., Graham, R. C., & McDaniel, P. A. (2011). *Soil genesis and classification*. John Wiley & Sons.
- Burbano, H. (2018). El carbono orgánico del suelo y su papel frente al cambio climático. *Revista de Ciencias Agrícolas*, 35(1), 82–86. <https://doi.org/10.22267/rcia.183501.85>
- Campbell, J.B., 2002. Introduction to Remote Sensing. Taylor & Francis, London.
- Chang, K. T., & Tsai, B. W. (1991). The effect of dem resolution on slope and aspect mapping. *Cartography and Geographic Information Systems*, 18(1), 69–77. <https://doi.org/10.1559/152304091783805626>
- Chen, S., Arrouays, D., Leatitia Mulder, V., Poggio, L., Minasny, B., Roudier, P., Libohova, Z., Lagacherie, P., Shi, Z., Hannam, J., Meersmans, J., Richer-de-Forges, A. C., & Walter, C. (2022). Digital mapping of GlobalSoilMap soil properties at a broad scale: A review. *Geoderma*, 409. <https://doi.org/10.1016/j.geoderma.2021.115567>
- Chenu, C., Angers, D. A., Barré, P., Derrien, D., Arrouays, D., & Balesdent, J. (2019). Increasing organic stocks in agricultural soils: Knowledge gaps and potential innovations. *Soil and Tillage Research*, 188, 41–52. <https://doi.org/10.1016/j.still.2018.04.011>
- Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., & Böhner, J. (2015). System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geoscientific Model Development*, 8(7), 1991–2007. <https://doi.org/10.5194/gmd-8-1991-2015>
- Croft, H., Kuhn, N. J., & Anderson, K. (2012). On the use of remote sensing techniques for monitoring spatio-temporal soil organic carbon dynamics in agricultural systems. In *Catena* (Vol. 94, pp. 64–74). <https://doi.org/10.1016/j.catena.2012.01.001>

- Conrad, O. (2006). SAGA—program structure and current state of implementation. *SAGA—Analysis and Modelling Applications*, edited by: Böhner, J., McCloy, KR, and Strobl, J., *Göttinger Geographische Abhandlungen*, Göttingen, 39-52.
- Conforti, M., Pascale, S., Robustelli, G., Sdao, F., 2014. Evaluation of prediction capability of the artificial neural networks for mapping landslide susceptibility in the Turbolo River catchment (northern Calabria, Italy). *Catena* 113, 236–250.
- Cuervo, E., Cely, G., & Moreno, D. (2016). *Determinación de las fracciones de carbono orgánico en el suelo del páramo La Cortadera, Boyacá*. 139–149.
- Davy, M. C., & Koen, T. B. (2013). Variations in soil organic carbon for two soil types and six land uses in the Murray Catchment, New South Wales, Australia. *Soil Research*, 51(7–8), 631–644. <https://doi.org/10.1071/SR12353>
- De Paul Obade, V., & Lal, R. (2013). Assessing land cover and soil quality by remote sensing and geographical information systems (GIS). In *Catena* (Vol. 104, pp. 77–92). <https://doi.org/10.1016/j.catena.2012.10.014>
- Dong, W., Wu, T., Luo, J., Sun, Y., & Xia, L. (2019). Land parcel-based digital soil mapping of soil nutrient properties in an alluvial-diluvia plain agricultural area in China. *Geoderma*, 340, 234-248.
- Elachi, C., & Van Zyl, J. J. (2006). Introduction to the physics and techniques of remote sensing (Vol. 28). John Wiley & Sons.
- Fao 2017. Carbono orgánico del suelo: el potencial oculto. Organización para las naciones unidas para la alimentación. Roma Italia. 90P
- FAO - Organización de las Naciones Unidas para la Alimentación. (2002). *Captura de carbono en los suelos para un mejor manejo de la tierra*. Basado en el trabajo de Michel Robert. Informes sobre recursos mundiales de suelos. FAO. 61p.
- Filippi, N., Gallego, F., Montanarella, L., & Panagos, P. (2005). *Soil Sampling Protocol to Certify the Changes of Organic Carbon Stock in Mineral Soils of European Union*. <https://www.researchgate.net/publication/237440368>
- Forkuor, G., Hounkpatin, O., Welp, G., & Thiel, M. (2017). High resolution mapping of soil properties using remote sensing variables in south-western Burkina Faso: a comparison of machine learning and multiple linear regression models. *PloS one*, 12(1), e0170478.
- Gardi, C., Angelini, M., Barceló, S., Comerma, J., Cruz Gaistardo, C., Encina Rojas, A., Jones A., Krasilnikov, P., Mendonça Santos Brefin, M. L., Montanarella, L., Muñiz Ugarte, O., Schad, P., Vara Rodríguez, M. I. y Vargas, R. (Eds.), (2014). Atlas de suelos de América Latina y el Caribe. Luxembourg: Comisión Europea - Oficina de Publicaciones de la Unión Europea, L-2995.

- Gallant, J., & Dowling, I. (2003). A multiresolution index of valley bottom flatness for mapping depositional areas. *Water Resources Research*, 39(12).
<https://doi.org/10.1029/2002WR001426>
- Gholizadeh, A., Žižala, D., Saberioon, M., & Borůvka, L. (2018). Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging. *Remote Sensing of Environment*, 218, 89–103. <https://doi.org/10.1016/j.rse.2018.09.015>
- Gomez, C., Viscarra Rossel, A., & McBratney, A. (2008). Soil organic carbon prediction by hyperspectral remote sensing and field vis-NIR spectroscopy: An Australian case study. *Geoderma*, 146(3–4), 403–411. <https://doi.org/10.1016/j.geoderma.2008.06.011>
- Gomes, L., Faria, R., de Souza, E., Veloso, V., Schaefer, C., & Fernandes Filho, I. (2019). Modelling and mapping soil organic carbon stocks in Brazil. *Geoderma*, 340, 337-350.
- Gray, J. M., Bishop, T. F. A., & Yang, X. (2015). Pragmatic models for the prediction and digital mapping of soil properties in eastern Australia. *Soil Research*, 53(1), 24–42.
<https://doi.org/10.1071/SR13306>
- Griffiths, R. P., Madritch, M. D., & Swanson, A. K. (2009). The effects of topography on forest soil characteristics in the Oregon Cascade Mountains (USA): Implications for the effects of climate change on soil properties. *Forest Ecology and Management*, 257(1), 1–7.
<https://doi.org/10.1016/j.foreco.2008.08.010>
- Guo, P., Li, F., Luo, W., Tang, Q., Liu, W., & Lin, M. (2015). Digital mapping of soil organic matter for rubber plantation at regional scale: An application of random forest plus residuals kriging approach. *Geoderma*, 237, 49-59.
- Gutiérrez, J., Ordoñez, N., Bolívar, A., Bunning, S., Guevara, M., Medina, E., Olivera, C., Olmedo, G., Rodríguez, L., Sevilla, V., & Vargas, R. (2020). Estimation of organic carbon in paramo ecosystem soils in Colombia. *Ecosistemas*, 29(1).
<https://doi.org/10.7818/ECOS.1855>
- Hinge, G., Surampalli, R., & Goyal, M. (2018). Prediction of soil organic carbon stock using digital mapping approach in humid India. *Environmental Earth Sciences*, 77(5).
<https://doi.org/10.1007/s12665-018-7374-x>
- Hengl, T., Heuvelink, G. B. M., Kempen, B., Leenaars, J. G. B., Walsh, M. G., Shepherd, K. D., Sila, A., MacMillan, R. A., De Jesus, J. M., Tamene, L., & Tondoh, J. E. (2015). Mapping soil properties of Africa at 250 m resolution: Random forests significantly improve current predictions. *PLoS ONE*, 10(6). <https://doi.org/10.1371/journal.pone.0125814>
- Hobley, E., Wilson, B., Wilkie, A., Gray, J., & Koen, T. (2015). Drivers of soil organic carbon storage and vertical distribution in Eastern Australia. *Plant and Soil*, 390(1–2), 111–127.
<https://doi.org/10.1007/s11104-015-2380-1>

- Hosseini, M., & McNairn, H. (2017). Using multi-polarization C-and L-band synthetic aperture radar to estimate biomass and soil moisture of wheat fields. *International Journal of Applied Earth Observation and Geoinformation*, 58, 50-64.
- Huamán-Carrión, M. L., Espinoza-Montes, F., Barrial-Lujan, A. I., & Ponce-Atencio, Y. (2021). Influence of altitude and soil characteristics on organic carbon storage capacity of high Andean natural pastures. *Scientia Agropecuaria*, 12(1), 83–90.
<https://doi.org/10.17268/SCI.AGROPECU.2021.010>
- Hunt, J. R., Celestina, C., & Kirkegaard, J. A. (2020). The realities of climate change, conservation agriculture and soil carbon sequestration. In *Global Change Biology* (Vol. 26, Issue 6, pp. 3188–3189). Blackwell Publishing Ltd. <https://doi.org/10.1111/gcb.15082>
- Keskin, H., Grunwald, S., & Harris, W. G. (2019). Digital mapping of soil carbon fractions with machine learning. *Geoderma*, 339, 40–58.
<https://doi.org/10.1016/j.geoderma.2018.12.037>
- Kuang, B., Tekin, Y., & Mouazen, A. M. (2015). Comparison between artificial neural network and partial least squares for on-line visible and near infrared spectroscopy measurement of soil organic carbon, pH and clay content. *Soil and Tillage Research*, 146(PB), 243–252.
<https://doi.org/10.1016/j.still.2014.11.002>
- Kunkel, M. L., Flores, A. N., Smith, T. J., McNamara, J. P., & Benner, S. G. (2011). A simplified approach for estimating soil carbon and nitrogen stocks in semi-arid complex terrain. *Geoderma*, 165(1), 1–11. <https://doi.org/10.1016/j.geoderma.2011.06.011>
- Lal, R., & Kimble, J. M. (1997). Conservation tillage for carbon sequestration. In *Nutrient Cycling in Agroecosystems* (Vol. 49). Kluwer Academic Publishers.
- Lamichhane, S., Kumar, L., & Wilson, B. (2019). Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review. In *Geoderma* (Vol. 352, pp. 395–413). Elsevier B.V. <https://doi.org/10.1016/j.geoderma.2019.05.031>
- Lee, S., Evangelista, D.G., 2006. Earthquake-induced landslide susceptibility mapping using an artificial neural network. *Nat. Hazards Earth Syst. Sci.* 6, 687–695.
- Lizeth, A., & Blanco, T. (2018). *Estimación de biomasa aérea de eucalipto (Eucalyptus grandis) y pino (Pinus spp) en plantaciones forestales comerciales, usando imágenes satelitales Sentinel.*
- López González, E. (1998). TRATAMIENTO DE LA COLINEALIDAD EN REGRESIÓN MÚLTIPLE. In *Psicothema* (Vol. 10, Issue 2).
- Mahmoudabadi, E., Karimi, A., Haghnia, G. H., & Sepehr, A. (2017). Digital soil mapping using remote sensing indices, terrain attributes, and vegetation features in the rangelands of

northeastern Iran. *Environmental Monitoring and Assessment*, 189(10).
<https://doi.org/10.1007/s10661-017-6197-7>

- Martínez, E., Fuentes, J. P., & Acevedo, E. (2008). *Soil organic carbon and soil properties*.
- Minasny, B., & McBratney, A. B. (2016a). Digital soil mapping: A brief history and some lessons. *Geoderma*, 264, 301–311. <https://doi.org/10.1016/j.geoderma.2015.07.017>
- Minasny, B., & McBratney, A. B. (2016b). Digital soil mapping: A brief history and some lessons. *Geoderma*, 264, 301–311. <https://doi.org/10.1016/j.geoderma.2015.07.017>
- Mirzaee, S., Ghorbani-Dashtaki, S., Mohammadi, J., Asadi, H., & Asadzadeh, F. (2016). Spatial variability of soil organic matter using remote sensing data. *Catena*, 145, 118–127. <https://doi.org/10.1016/j.catena.2016.05.023>
- Mouazen, A. M., Kuang, B., De Baerdemaeker, J., & Ramon, H. (2010). Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy. *Geoderma*, 158(1–2), 23–31. <https://doi.org/10.1016/j.geoderma.2010.03.001>
- Obu, J., Lantuit, H., Myers-Smith, I., Heim, B., Wolter, J., & Fritz, M. (2017). Effect of Terrain Characteristics on Soil Organic Carbon and Total Nitrogen Stocks in Soils of Herschel Island, Western Canadian Arctic. *Permafrost and Periglacial Processes*, 28(1), 92–107. <https://doi.org/10.1002/ppp.1881>
- Panagos, P., Borrelli, P., Meusburger, K., Alewell, C., Lugato, E., & Montanarella, L. (2015). Estimating the soil erosion cover-management factor at the European scale. *Land use policy*, 48, 38-50.
- Pla Sentis, I. (1994). Labranza, propiedades físicas y producción de maíz en los llanos occidentales. *SVCS N*, 46, 32-42.
- Qin, C., Zhu, A. X., Pei, T., Li, B., Zhou, C., & Yang, L. (2007). An adaptive approach to selecting a flow-partition exponent for a multiple-flow-direction algorithm. *International Journal of Geographical Information Science*, 21(4), 443-458.
- Ramifehiarivo, N., Brossard, M., Grinand, C., Andriamananjara, A., Razafimbelo, T., Rasolohery, A., Razafimahatratra, H., Seyler, F., Ranaivoson, N., Rabenarivo, M., Albrecht, A., Razafindrabe, F., & Razakamanarivo, H. (2017). Mapping soil organic carbon on a national scale: Towards an improved and updated map of Madagascar. *Geoderma Regional*, 9, 29–38. <https://doi.org/10.1016/j.geodrs.2016.12.002>
- Reyes, L., Adhikari, K., & Ventura, S. (2018). Projecting Soil Organic Carbon Distribution in Central Chile under Future Climate Scenarios. *Journal of Environmental Quality*, 47(4), 735–745. <https://doi.org/10.2134/jeq2017.08.0329>

- Ricardo, J., & Florez, M. (2019). *Evaluación de imágenes de radar Sentinel-1A e imágenes multiespectrales Sentinel-2A en la clasificación de cobertura del suelo en diferentes niveles de detalle.*
- Rivas, M., Otero, J., Van der Hammen, T., Torres, A., Cadena, C., Pedraza, C., Rodríguez, N., Franco, C., Betancourth, J., Olaya, É., Posada, E., & Cárdenas, L. (2007). *Atlas de páramos de Colombia.*
- Roa Lobo, J., & Kamp, U. (2012). Use of the topographic wetness index (TWI) for the diagnosis of the river overflow threat, Trujillo State-Venezuela. *Revista Geográfica Venezolana*, 53(1), 109-126.
- Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67(1), 93–104. <https://doi.org/10.1016/j.isprsjprs.2011.11.002>
- Román-Sánchez, A., Vanwalleghem, T., Peña, A., Laguna, A., & Giráldez, J. V. (2018). Controls on soil carbon storage from topography and vegetation in a rocky, semi-arid landscapes. *Geoderma*, 311, 159–166. <https://doi.org/10.1016/j.geoderma.2016.10.013>
- Rügnitz, M., M. Chacón & R. Porro. (2009). Guía para la determinación de carbono en pequeñas propiedades rurales. Lima, Perú.: Centro Mundial Agroflorestal (ICRAF) / Consorcio Iniciativa Amazónica (IA).
- Sandoval, M., Stolpe, N., Venegas, E., Mardones, M., & Montano, J. (2003). El secuestro de carbono en la agricultura y su importancia en el calentamiento global. In *Theoria* (Vol. 12).
- Sankey, T. T., & Weber, K. T. (2009). Rangeland assessments using remote sensing: is NDVI useful. *Final Report: Comparing Effects of Management Practices on Rangeland Health with Geospatial Technologies (NNX06AE47G)*, 168. Sankey, T. T., & Weber, K. T. (2009). Rangeland assessments using remote sensing: is NDVI useful. *Final Report: Comparing Effects of Management Practices on Rangeland Health with Geospatial Technologies (NNX06AE47G)*, 168.
- Sayão, V. M., & Demattê, J. A. M. (2018). Soil texture and organic carbon mapping using surface temperature and reflectance spectra in Southeast Brazil. *Geoderma Regional*, 14. <https://doi.org/10.1016/j.geodrs.2018.e00174>
- Schillaci, C., Acutis, M., Lombardo, L., Lipani, A., Fantappiè, M., Märker, M., & Saia, S. (2017). Spatio-temporal topsoil organic carbon mapping of a semi-arid Mediterranean region: The role of land use, soil texture, topographic indices and the influence of remote sensing data to modelling. *Science of the Total Environment*, 601–602, 821–832. <https://doi.org/10.1016/j.scitotenv.2017.05.239>

- Siewert, M. (2018). High-resolution digital mapping of soil organic carbon in permafrost terrain using machine learning: A case study in a sub-Arctic peatland environment. *Biogeosciences*, 15(6), 1663–1682. <https://doi.org/10.5194/bg-15-1663-2018>
- Sreenivas, K., Dadhwal, V. K., Kumar, S., Harsha, G. S., Mitran, T., Sujatha, G., Suresh, G. J. R., Fyzee, M. A., & Ravisankar, T. (2016). Digital mapping of soil organic and inorganic carbon status in India. *Geoderma*, 269, 160–173. <https://doi.org/10.1016/j.geoderma.2016.02.002>
- Stolbovoy, V., Montanarella, L., Filippi, N., Selvaradjou, S., Panagos, P., & Gallego, J. (2005). Soil sampling protocol to certify the changes of organic carbon stock in mineral soils of European Union. *vol EUR, 21576*. Sumfleth, K., & Duttmann, R. (2008). Prediction of soil property distribution in paddy soil landscapes using terrain data and satellite information as indicators. *Ecological Indicators*, 8(5), 485–501. <https://doi.org/10.1016/j.ecolind.2007.05.005>
- Thompson, J. A., Pena-Yewtukhiw, E. M., & Grove, J. H. (2006). Soil-landscape modeling across a physiographic region: Topographic patterns and model transportability. *Geoderma*, 133(1–2), 57–70. <https://doi.org/10.1016/j.geoderma.2006.03.037>
- Taghizadeh, R., Nabiollahi, K., & Kerry, R. (2016). Digital mapping of soil organic carbon at multiple depths using different data mining techniques in Baneh region, Iran. *Geoderma*, 266, 98–110. <https://doi.org/10.1016/j.geoderma.2015.12.003>
- Torres, J., Valdez, J., Pérez, G., De los Santos, H., & Aguirre, C. (2017). Inventory and mapping of a pine forest under timber management using data obtained with a SPOT 6 sensor. In *Revista Mexicana de Ciencias Forestales* (Vol. 8, Issue 39).
- Wang, S., Zhuang, Q., Wang, Q., Jin, X., & Han, C. (2017). Mapping stocks of soil organic carbon and soil total nitrogen in Liaoning Province of China. *Geoderma*, 305, 250–263.
- Wang, B., Waters, C., Orgill, S., Gray, J., Cowie, A., Clark, A., & Liu, D. L. (2018). High resolution mapping of soil organic carbon stocks using remote sensing variables in the semi-arid rangelands of eastern Australia. *Science of the Total Environment*, 630, 367–378. <https://doi.org/10.1016/j.scitotenv.2018.02.204>
- Wang, X., Cammeraat, E. L. H., Cerli, C., & Kalbitz, K. (2014). Soil aggregation and the stabilization of organic carbon as affected by erosion and deposition. *Soil Biology and Biochemistry*, 72, 55–65. <https://doi.org/10.1016/j.soilbio.2014.01.018>
- Ward, A., Dargusch, P., Grussu, G., & Romeo, R. (2015). Using carbon finance to support climate policy objectives in high mountain ecosystems. *Climate Policy*, 16(6), 732–751. <https://doi.org/10.1080/14693062.2015.1046413>
- Waters, C. M., Melville, G. J., Orgill, S. E., & Alemseged, Y. (2015). The relationship between soil organic carbon and soil surface characteristics in the semi-arid rangelands of southern Australia. *Rangeland Journal*, 37(3), 297–307. <https://doi.org/10.1071/RJ14119>

- Were, K., Bui, D. T., Dick, Ø. B., & Singh, B. R. (2015). A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. *Ecological Indicators*, 52, 394–403. <https://doi.org/10.1016/j.ecolind.2014.12.028>
- Wilson, J. P., & Gallant, J. C. (Eds.). (2000). *Terrain analysis: principles and applications*. John Wiley & Sons.
- Wilson, J. P. (2012). Digital terrain modeling. *Geomorphology*, 137(1), 107–121. <https://doi.org/10.1016/j.geomorph.2011.03.012>
- Wischmeier, W. H., & Smith, D. D. (1978). *Predicting rainfall erosion losses: a guide to conservation planning* (No. 537). Department of Agriculture, Science and Education Administration.
- Xie, H., Zhao, J., Wang, Q., Sui, Y., Wang, J., Yang, X., ... & Liang, C. (2015). Soil type recognition as improved by genetic algorithm-based variable selection using near infrared spectroscopy and partial least squares discriminant analysis. *Scientific reports*, 5(1), 10930.
- Xiao, J., Chevallier, F., Gomez, C., Guanter, L., Hicke, J. A., Huete, A. R., Ichii, K., Ni, W., Pang, Y., Rahman, A. F., Sun, G., Yuan, W., Zhang, L., & Zhang, X. (2019). Remote sensing of the terrestrial carbon cycle: A review of advances over 50 years. *Remote Sensing of Environment*, 233. <https://doi.org/10.1016/j.rse.2019.111383>
- Yang, Y., Fang, J., Tang, Y., Ji, C., Zheng, C., He, J., & Zhu, B. (2008). Storage, patterns and controls of soil organic carbon in the Tibetan grasslands. *Global Change Biology*, 14(7), 1592–1599. <https://doi.org/10.1111/j.1365-2486.2008.01591.x>
- Yang, R. M., Zhang, G. L., Yang, F., Zhi, J. J., Yang, F., Liu, F., Zhao, Y. G., & Li, D. C. (2016). Precise estimation of soil organic carbon stocks in the northeast Tibetan Plateau. *Scientific Reports*, 6. <https://doi.org/10.1038/srep21842>
- Yang, S., Sheng, D., Adamowski, J., Gong, Y., Zhang, J., & Cao, J. (2018). Effect of land use change on soil carbon storage over the last 40 years in the Shi Yang River Basin, China. *Land*, 7(1). <https://doi.org/10.3390/land7010011>
- Zeraatpisheh, M., Ayoubi, S., Jafari, A., Tajik, S., & Finke, P. (2019). Digital mapping of soil properties using multiple machine learning in a semi-arid region, central Iran. *Geoderma*, 338, 445-452.
- Zhao, M. S., Rossiter, D. G., Li, D. C., Zhao, Y. G., Liu, F., & Zhang, G. L. (2014). Mapping soil organic matter in low-relief areas based on land surface diurnal temperature difference and a vegetation index. *Ecological Indicators*, 39, 120–133. <https://doi.org/10.1016/j.ecolind.2013.12.015>

- Zhou, Q., & Liu, X. (2004). Analysis of errors of derived slope and aspect related to DEM data properties. *Computers & Geosciences*, 30(4), 369-378.
- Zhou, T., Geng, Y., Chen, J., Pan, J., Haase, D., & Lausch, A. (2020). High-resolution digital mapping of soil organic carbon and soil total nitrogen using DEM derivatives, Sentinel-1 and Sentinel-2 data based on machine learning algorithms. *Science of the Total Environment*, 729. <https://doi.org/10.1016/j.scitotenv.2020.138244>
- Zimmermann, M., Meir, P., Silman, M. R., Fedders, A., Gibbon, A., Malhi, Y., Urrego, D. H., Bush, M. B., Feeley, K. J., Garcia, K. C., Dargie, G. C., Farfan, W. R., Goetz, B. P., Johnson, W. T., Kline, K. M., Modi, A. T., Rurau, N. M. Q., Staudt, B. T., & Zamora, F. (2010). No differences in soil carbon stocks across the tree line in the Peruvian Andes. *Ecosystems*, 13(1), 62–74. <https://doi.org/10.1007/s10021-009-9300-2>
- Zomer, R. J., Bossio, D. A., Sommer, R., & Verchot, L. V. (2017). Global Sequestration Potential of Increased Organic Carbon in Cropland Soils. *Scientific Reports*, 7(1). <https://doi.org/10.1038/s41598-017-15794-8>
- Zuur, A. F., Ieno, E. N., & Smith, G. M. (2007). *Analysing ecological data* (Vol. 680). New York: Springer.