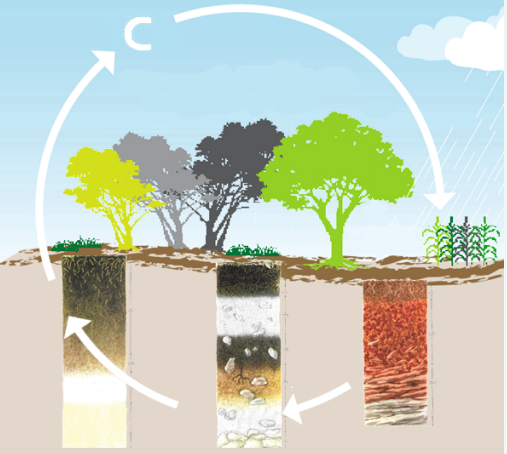
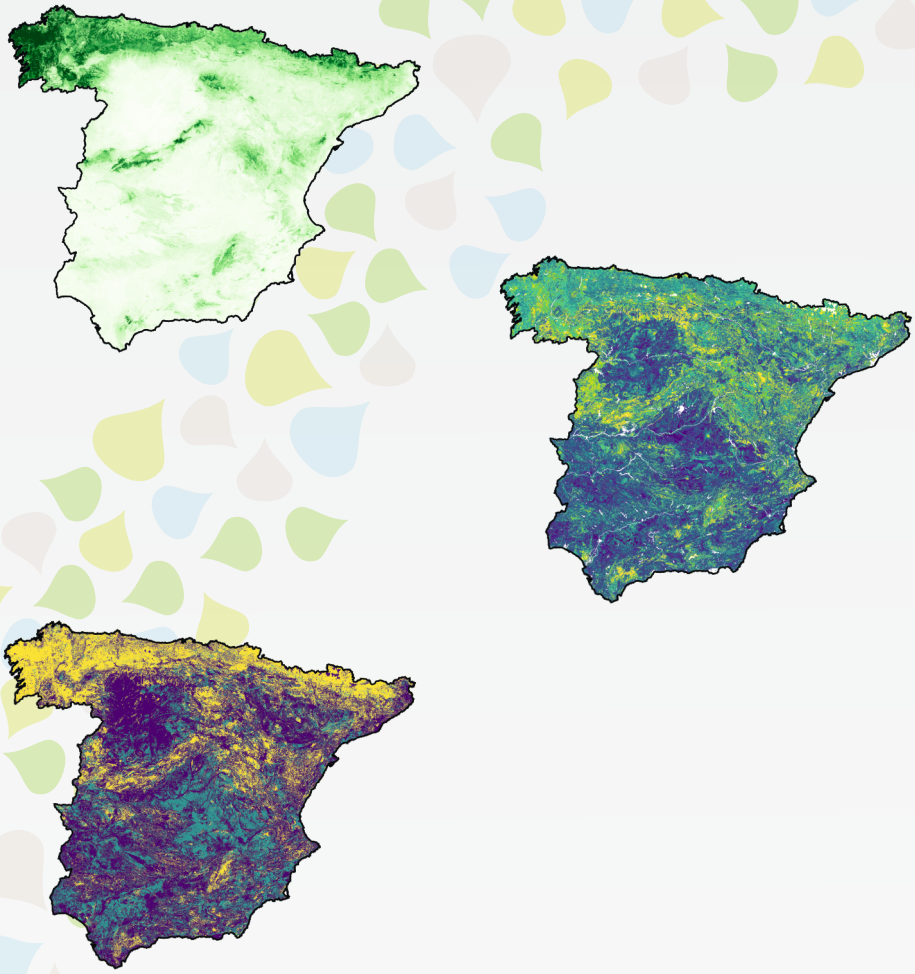
CAESCG

Modelización del carbono orgánico para la cuantificación de sus reservorios en los ecosistemas terrestres a escala nacional

Modeling organic carbon for quantification of reservoirs in terrestrial ecosystems at the national level





Pilar Durante Hernández

Tesis Doctoral Diciembre 2021

UNIVERSIDAD DE ALMERÍA

PROGRAMA DE DOCTORADO CIENCIAS APLICADAS AL MEDIO AMBIENTE

UNIVERSIDAD DE ALMERÍA. DEPARTAMENTO DE AGRONOMÍA

UNIVERSIDAD DE ALMERÍA

MODELIZACIÓN DEL CARBONO ORGÁNICO PARA LA CUANTIFICACIÓN DE SUS RESERVORIOS EN LOS ECOSISTEMAS TERRESTRES A ESCALA NACIONAL

MODELING ORGANIC CARBON FOR QUANTIFICATION OF RESERVOIRS IN TERRESTRIAL ECOSYSTEMS AT THE NATIONAL LEVEL

TESIS DOCTORAL

Pilar Durante Hernández

Ingeniera de Montes

DIRECTORES

Cecilio Oyonarte Gutiérrez

Doctor Ciencias Biológicas

Nur Algeet Abarquero

Doctora Ingeniera de Montes

DICIEMBRE 2021

MENCIÓN DE DOCTORADO INTERNACIONAL

INTERNATIONAL DOCTORATE MENTION

Esta tesis ha sido informada positivamente para su defensa en exposición pública por los siguientes investigadores:

This Ph.D. Thesis has been positively evaluated for its public defense by the following external reviewrs:

Dr. Mario Guevara

Centro De Geociencias

Universidad Nacional Autónoma de México

Campus Juriquilla (México)

Dr Juan Miguel Requena Mullor

University of Michigan (USA)

Este tesis ha sido parcialmente financiación por el Subprograma Torres Quevedo del antiguo Ministerio de Economía y Competitividad, con número de expediente DI-15-08093, así como por Agresta S.Coop. El trabajo ha sido desarrollado en el Departamento de Agronomía y en el Centro Andaluz para la Evaluación y Seguimiento del Cambio Global (CAESCG) de la Universidad de Almería. El mismo programa de financiación permitió la realización de la estancia en el Laboratorio del Departamento de Ciencias Vegetales y del Suelo de la Universidad de Delaware (USA), gracias al apoyo de su director Prof. Dr. Rodrigo Vargas. Parte de los análisis de este estudio han sido realizados gracias a los servicios de supercomputación facilitados por el Centro Informático Científico de Andalucía (CICA).

AGRADECIMIENTOS

Recapitulando desde que comencé con este trabajo, me he dado cuenta de la gran influencia que ha tenido este período especialmente en la parte emocional. Para mí, este apartado es un resumen de la tesis desde ese punto de vista. Y es que esta etapa ha sido excepcional en todos los sentidos. Tanto por lo aprendido como por lo vivido, con una pandemia por medio que no nos dejará indiferentes.

He aprendido mucho sobre los ecosistemas, y no sólo por la temática de la tesis. También del ecosistema que se ha generado alrededor este trabajo, donde todas las personas que han influido se han intercontectando desde diferentes enfoques. Y así lo he querido plasmar en la imagen del árbol que acompaña este texto, dibujado por mi hijo en un intento de poner en valor los bosques.

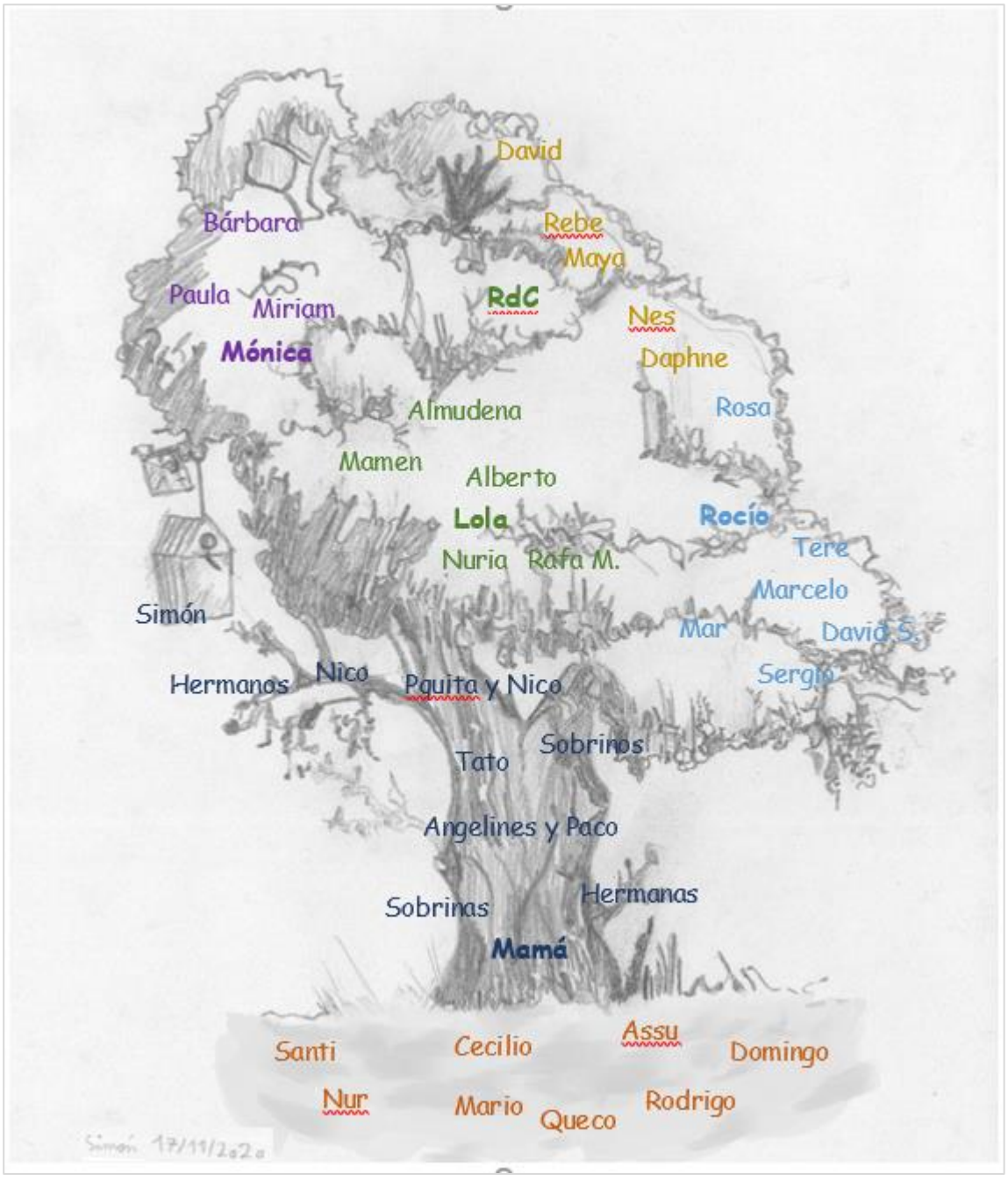
Si remontamos al momento de la semilla, indudablemente mis directores de tesis, Cecilio y Nur, contribuyeron a plantarla. A Cecilio le estoy especialmente agradecida por su implicación, conocimiento y buena sintonía. Nos conocemos desde hace tiempo y este trabajo ha afianzado nuestra amistad. Me gustaría que todo este esfuerzo en conjunto sirviera de homenaje a tu carrera profesional, a tu entusiasmo en la materia y a tu disposición por transferir el conocimiento. También dar especialmente las gracias a Nur, por su plena confianza en mí, y por trasmitirme su sabiduría, sensatez y tranquilidad. Y por abrirme las puertas a Agresta, lo cual le estaré siempre agradecida. Me ha permitido trabajar con un pedazo de equipo tanto técnico como humano. En esta parte relacionada con el sustento de la tesis, también me gustaría incluir a Mario Guevara que, sin ser mi director oficial, no ha dudado en colaborar y compartir ciencia, confiando en mis capacidades incluso cuando yo dudaba. Y a Rodrigo Vargas, por su hospitalidad, cercanía y colaboración, haciéndome sentir como una más del equipo de investigación al llegar a Delaware. Siempre dispuesto a ayudarme. También a Domingo Alcaráz, por su colaboración con los índices verdes, con el que he creado un vínculo a prueba del paso del tiempo.

En el tronco y ramas, sosteniéndome, a mi gran familia (la biológica y la felizmente adoptada). Resistente y flexible para darme apoyo infinito. Esto no hubiera sido posible sin vuestra ayuda.

En la copa del árbol, a todas las personas que me habéis acompañado y escuchado pacientemente, dándome ánimos y empatizando con mi proyecto. Especial mención a mi grupo de Almería, tan importante en la caída de las hojas como en el resurgir de la primavera. A mi querido grupo de Murcia, mi segundo hogar. A mi especial grupo de Cuenca, compartiendo esos frutos maduros y construyendo sororidad. Y a las que lleváis a mi lado desde los inicios, compartiendo otra etapa más de nuestras vidas.

Cada persona ha hecho que este ecosistema haya funcionado. Gracias de verdad.

Quiero acabar con una mención especial a Tato et al., es decir, también a Simón y a Nico. Por aportarme ánimo, coraje y espacio durante este desarrollo. Pero especialmente a Tato, por su mirada serena y transparente, y por enseñarme a que no necesitamos la vista para percibir las cosas más importantes porque también "*se ven con los ojos cerrados. Por eso será que los cerramos cuando besamos, lloramos y soñamos*".


David
Bárbara
Rebe
Maya
Paula
Miriam
RdC
Nes
Mónica
Daphne
Almudena
Rosa
Mamen
Alberto
Lola
Rocío
Nuria
Rafa M.
Tere
Marcelo
Simón
Mar
David S.
Sergio
Hermanos
Nico
Paula y Nico
Sobrinos
Tato
Angelines y Paco
Sobrinas
Hermanas
Mamá
Santi
Cecilio
Assu
Domingo
Nur
Mario
Queco
Rodrigo
Simón 17/11/2020

X

Cada cual posee la fuerza para intentar aquello que se cree incapaz de lograr. Encuentra una nueva fuente de energía y de voluntad, y después… continúa avanzando.

Scott Jurek

A Nico y a Simón, por el tiempo robado

ÍNDICE FIGURAS

ÍNDICE TABLAS

RESUMEN

Bajo el actual contexto de cambio de los patrones climáticos, los ecosistemas terrestres se encuentran gravemente alterados respecto a su biodiversidad, estructura o funcionamiento. Concretamente, la región Mediterránea representa una de las zonas más afectadas por el impacto del cambio climático. En base a las nuevas políticas ambientales derivadas de las graves amenazas provocadas por el cambio climático, se deben establecer medidas prácticas para disminuir el CO_2 atmosférico. El secuestro de carbono es una de las medidas principales para reducir las concentraciones de CO_2 atmosférico a corto y medio plazo, en el que los ecosistemas terrestres juegan un papel fundamental como sumideros de carbono. Es necesario cuantificar y monitorear los reservorios de carbono orgánico que sirvan de base para la gestión ambiental, adaptar las políticas locales y evaluar los impactos potenciales.

El objetivo general de esta tesis es avanzar en la estrategia metodológica para la cuantificación de los reservorios de carbono orgánico en ecosistemas terrestres, basada en técnicas estandarizadas a diferentes escalas espaciales, así como establecer modelos dinámicos capaces de predecir los resultados bajo diferentes escenarios de gestión. Esta tesis se centra en dos aspectos clave de la modelización de los stocks de carbono a nivel nacional: (1) la estimación del carbono de la biomasa aérea, y (2) la cuantificación del carbono almacenado en el suelo, así como su potencial secuestro bajo diferentes escenarios de gestión de usos de suelo; en todos los casos mediante modelos espacialmente explícitos.

Tras la introducción y métodos generales (capítulo 1 y 2), en el capítulo 3 se integran dos tecnologías de teledetección complementarias para elaborar información detallada de la distribución espacial del carbono almacenado en la biomasa. La distribución multitemporal y global de índices de resolución moderada (MODIS) compensan la limitación temporal de los datos del escáner láser aerotransportado (ALS) de alta precisión. Como caso de estudio, esta metodología se aplicó en una región mediterránea semiárida en el sureste de la Península Ibérica (Región de Murcia). Los resultados muestran resultados robustos en el modelado de datos ALS, calibrados con medidas a nivel de parcela, y variables espectrales biogeofísicas (8 índices derivados de MODIS), lo que confirma su aplicabilidad a tamaños mayores de píxel.

En el capítulo 4 se emplean diferentes técnicas de mapeo digital de suelos (DSM) para generar un mapa local de carbono orgánico del suelo (SOC) y testarlo con estimaciones derivadas de productos disponibles tanto regionales como globales. El objetivo de este capítulo es definir un marco metodológico para la generación de un mapa de SOC de alta resolución, analizando diversos aspectos. Estos aspectos se refieren a diferentes variables de carbono (concentración de SOC -g / kg-, y stock de SOC –Tc / ha-) utilizando distintos métodos de interpolación espacial (linear model, quantile regression forest –QRF-; random forest and support vector machine) en tres resoluciones diferentes (100 m, 250 m, 1000 m). Considerando de nuevo la 'Región de Murcia' como caso de estudio, este estudio muestra que el marco metodológico con resultados más robustos está basado en datos locales del suelo, covariables ambientales (incluidos índices de teledetección simples y / o

multitemporales), técnicas de DMS y cuantificación de la incertidumbre espacialmente explícita. Específicamente, el enfoque QRF parametrizado con datos de concentración de SOC, a una resolución espacial de 100 m, muestra la mayor concordancia en la modelización de los datos, así como el mejor balance entre la precisión, la validación externa y la interpretabilidad de los resultados. Este estudio proporciona una mejor comprensión del almacenamiento de SOC en una zona con asociaciones de suelo complejas y con información muestral limitada.

A la vista de los resultados del capítulo 4, en el capítulo 5 se generan diferentes mapas de la distribución espacial del SOC a nivel nacional, con una resolución de 90 m y sus incertidumbres asociadas espacialmente explícitas. La modelización de la base de datos patrimonial (8,361 muestras de perfil) y las covariables ambientales previamente seleccionadas, se basó en tres enfoques de aprendizaje supervisado: quantile regression forest, ensemble machine learning and auto-machine learning. Los mapas han sido estimados para las profundidades de 0-30 cm, 30-100 cm y para la profundidad efectiva del perfil. Para la generación de estos mapas finales de distribución espacial del SOC se ha empleado una metodología novedosa mediante diferenetes ensambles de modelos y su posterior combinación espacial. Dicha combinación se genera mediante la asignación del ensamble de modelos más preciso a cada píxel, es decir, el que presenta menor incertidumbre. Los mapas resultantes mostraron valores medios de concentración del SOC de 15.7 g/kg, almacenando el 25% aproximadamente en los horizontes subsuperficiales (> 30 cm). El stock total de SOC se ha estimado en 3,8 Pg C a su profundidad efectiva, de los cuales 2,82 Pg C se almacenan en los 30 cm superiores (74% del total).

En el capítulo 6 se predice el mapa de secuestro potencial de SOC con el objetivo de detectar usos de suelo, áreas y regiones con mayor potencial de absorción del SOC en España peninsular bajo diferentes escenarios de gestión. En este estudio se utiliza el modelo RothC a nivel nacional (grid de 1 km) para la proyección en el período 2020-2040. Los resultados muestran que el secuestro de SOC en España Peninsular, suponiendo que las condiciones ambientales actuales permanecen constantes durante los próximos 20 años, disminuirá en 430 Gg C/año. Sin embargo, se estima que se puede esperar un potencial secuestro de 1.977 Gg C/año si se adaptan prácticas de gestión sostenible que aumenten la tasa de entrada de carbono en el suelo al menos un 5% en los próximos 20 años.

En resumen, los avances que proporciona esta tesis contribuyen a mejorar la calidad y precisión de las actuales metodologías disponibles del carbono ecosistémico a escala nacional, con información novedosa sobre su cantidad almacenada, sirviendo como referencia para las estrategias y políticas de adaptación/mitigación al cambio climático.

ABSTRACT

Due to current climate change patterns, terrestrial ecosystems are being seriously affected with respect to their biodiversity, structure, and function. Specifically, the Mediterranean region is one of the areas most sensitive to climate change effects. Given the recent environmental policies derived from these serious threats caused by global climate change, practical measures to decrease net CO_2 emissions must be established. Carbon sequestration is a major measure to reduce atmospheric CO_2 concentrations within the short and medium term, in which terrestrial ecosystems play an essential role as carbon sinks. Quantification and monitoring of organic carbon reservoirs is needed to inform environmental management, adapt local policies and assess potential impacts.

The main aim of this thesis is to provide the methodological strategy for the quantification of organic carbon reservoirs in terrestrial ecosystems, based on standardized techniques at different spatial and management scales, in addition to establishing dynamic models to predict results under different management scenarios. This thesis is focused on two critical aspects of carbon stock modeling at the national level: (1) estimation of the carbon in aboveground biomass, and (2) quantification of the soil carbon storage, as well as its potential sequestration under different land use management scenarios; using spatially explicit models in both cases.

After the general introduction and methods (chapter 1 and 2), chapter 3 integrates two complementary remote sensing technologies to detail information about the spatial distribution of carbon stored in biomass. The multitemporal and global distribution of moderate resolution indexes (MODIS) are balanced with the temporal limitation of the high-precision airborne laser scanning (ALS) data. As a case study, this methodology was applied in a Mediterranean semiarid region in the southeastern Iberian Peninsula (specifically the region of Murcia). The results shows a robust performance in modeling of ALS data calibrated with plot-level ground-based measures, and bio-geophysical spectral variables (8 different indexes derived from MODIS), confirming its applicability at coarser resolutions.

Chapter 4 improves different digital soil mapping (DSM) techniques to develop a local soil organic carbon (SOC) map and test it against estimates derived from available regional-to-global carbon products. The aim of this chapter is to define a high-resolution SOC map framework, analyzing diverse aspects. These aspects are refered to different carbon variables (SOC concentration -g/kg-, and SOC stock -tC/ha-) using different spatial interpolation methods (linear model, quantile regression forest –QRF-; random forest and support vector machine) at three spatial resolutions (100m, 250m, 1000m). Considering again the 'Region de Murcia' as a case study, the results show an optimal framework based on local soil data, environmental covariates (including single and/or multitemporal remote sensing indices), DMS modeling and spatially explicit uncertainty quantification. Specifically, the QRF approach parameterized with SOC concentration data at 100 m spatial resolution confirms the best data-model agreement and the best balance for accuracy, external validation, and interpretability of results. This study provides a better understanding of SOC storage across complex soil-forming environments with limited soil samples.

In view of the results of chapter 4, chapter 5 maps the SOC spatial distribution at the national level at 90 m and the associated spatially explicit uncertainties. Modeling of the legacy soil database (8,361 profile samples) and a selection of environmental data-driven covariates was based on three supervised learning approaches: quantile regression forest, ensemble machine learning and auto-machine learning. The maps are estimated at 0-30 cm, 30-100cm and the effective soil depth. For the final SOC spatial distribution maps, a novel methodology is used. It is based on a combination of different predictive ensemble models where each pixel is assigned the prediction from the most accurate model, i.e. lowest uncertainty. The mean value of the SOC concentration map is 15.7 g/kg, storing approximately 25% in subsoils (>30cm). The total SOC stock at its effective depth is 3.8 Pg C, of which 2.82 Pg C are stored in the upper 30 cm (74% of the total).

Chapter 6 predicts SOC potential sequestration map to detect land uses, sites and regions with greater potential to absorb SOC stocks for peninsular Spain under different management scenarios. In this study, the RothC model at the national level (1 km grids) is used for the projection of the 2020-2040 period. The results shows that the SOC sequestration in peninsular Spain, supposing the current environmental conditions remain constant for the next 20 years, will decrease at a rate of 430 Gg C/yr approximately. However, a potential sequestration of 1,980 Gg C/yr can be expected approximately under the adoption of sustainable management practices that increase the carbon input rate into soils by at least 5% in the next 20 years.

In summary, the advances provided by this thesis contribute to improving the quality and precision of current methodologies available at the national level for the carbon in terrestrial ecosystem, with novel information on its carbon storage, and serve as a reference for climate change mitigation strategies and policies.

Capítulo 1. Introducción General y Objetivos

1.1. Antecedentes

El proceso de fotosíntesis captura la energía solar y la almacena en carbohidratos, materia orgánica, que son los que proporcionan la energía que impulsa las reacciones bioquímicas de la vida. La acumulación de biomasa en un ecosistema (producción neta del ecosistema, NEP) ocurre principalmente debido al crecimiento de la biomasa aérea, transfiriéndose al suelo con su consecuente incremento de materia orgánica (Chapin et al., 2011). La mayoría de los restos de biomasa, litter, se depositan en las capas superficiales del suelo donde están sometidos a procesos de descomposición controlados por la microfauna, las bacterias y los hongos. La descomposición conduce a la liberación de CO_2, H_2O y elementos nutritivos al medio, y a la generación de compuestos orgánicos ligados al suelo altamente resistentes conocidos como humus (Schlesinger and Bernhardt, 2013).

Estos compuestos húmicos se acumulan en el perfil del suelo y componen la mayor parte de la materia orgánica edáfica. La dinámica de la reserva de carbono en los suelos incluye dos etapas: procesos que conducen a un rápido recambio de la mayor parte del litter en la superficie, y procesos que conducen a una producción, acumulación de humus en profundidad con un recambio metabólico (turnover) muy bajo (Rumpel and Kögel-Knabner, 2011). El humus es muy resistente al proceso de descomposición, y se estima que los materiales húmicos extraídos del suelo tienen una edad media entre 250 a 940 años (Campbell et al., 1967). Se ha comprobado que incluso los ecosistemas estables continúan almacenando algo de materia orgánica en los suelos (Zhou et al., 2006), manteniéndose durante prolongados periodos su capacidad de secuestro de carbono.

Ante las nuevas medidas y políticas medioambientales, resultado del compromiso de estabilizar las concentraciones de gases de efecto invernadero (GEI) acordado en diversas convenciones y tratados de cambio climático, iniciados con la Convención Marco de las Naciones Unidas sobre el Cambio Climático –UNFCCC- y el Protocolo de Kioto, y actualizados con la XXI Conferencia Internacional sobre Cambio Climático 2015, surge la necesidad de reducir las emisiones netas de CO_2 (utilizado en adelante como un subrogado de los GEI en general). Dada la escala y gravedad de las amenazas derivadas del cambio climático, en la Cumbre del Clima de París 2015 (COP21) se insta a los Gobiernos a que tomen medidas concretas vinculantes. Dichas medidas están encaminadas a alcanzar el objetivo de limitar el aumento del calentamiento global, mediante la reducción de las emisiones de gases de efecto invernadero.

Junto a la disminución de emisiones, se propone la fijación de carbono como la vía para reducir las concentraciones de CO_2 en la atmósfera a corto y medio plazo, donde los ecosistemas terrestres juegan un papel fundamental como sumideros de carbono. Esta fijación derivada de actividades agrícolas y forestales puede suponer una importante ayuda a la hora de cumplir con los compromisos de reducción de gases acordados para cada país (Camps Arbestain and Pinto, 2004; FAO, 2017b).

El marco político relacionado con las emisiones de CO_2 ha evolucionado recientemente a nivel europeo y mundial, desarrollando directrices tanto para la recuperación de ecosistemas con mayor capacidad de almacenamiento de carbono, como para mejorar tales capacidades. Así mismo, para la contabilidad actual del stock de carbono

ecosistémico se incluye no solo el derivado por la vegetación o cambios de usos de suelo, sino específicamente por la propia gestión de los suelos (Bispo et al., 2017).

El desarrollo de metodologías de la cuantificación de los stocks de carbono han de permitir también la gestión de los reservorios de carbono terrestre a nivel espacial y su seguimiento temporal. Esto resulta esencial para contribuir al cumplimiento de los compromisos acordados a nivel estatal para la mitigación del cambio climático mediante una mejor comprensión de los procesos espacio-temporales relacionados con el ciclo del carbono.

1.2. El carbono en los ecosistemas terrestres

El secuestro de carbono en los ecosistemas terrestres se define como el proceso por el cual el CO_2 atmosférico es absorbido mediante la fotosíntesis almacenándose en sus diferentes reservorios mediante diferentes procesos. Estos reservorios incluyen principalmente a la vegetación y al suelo (Figure 1.1). En la vegetación, el C absorbido es almacenado en la biomasa (troncos, ramas, hojas y raíces) y en los suelos se almacena en forma de materia orgánica (IPCC, 2014). A gran escala, la relación promedio entre el carbono almacenado en los suelos y el almacenado en la vegetación varía según el tipo de ecosistema, la composición vegetal y la latitud (Dixon et al., 1994). Por esto, la eficiencia de los ecosistemas para secuestrar carbono está limitada, en parte, por variaciones en la composición vegetal, en su tasa de crecimiento y mortalidad, en la composición química del material biológico vegetal, así como en las propiedades del suelo, la topografía o el clima (Lorenz and Lal, 2010).

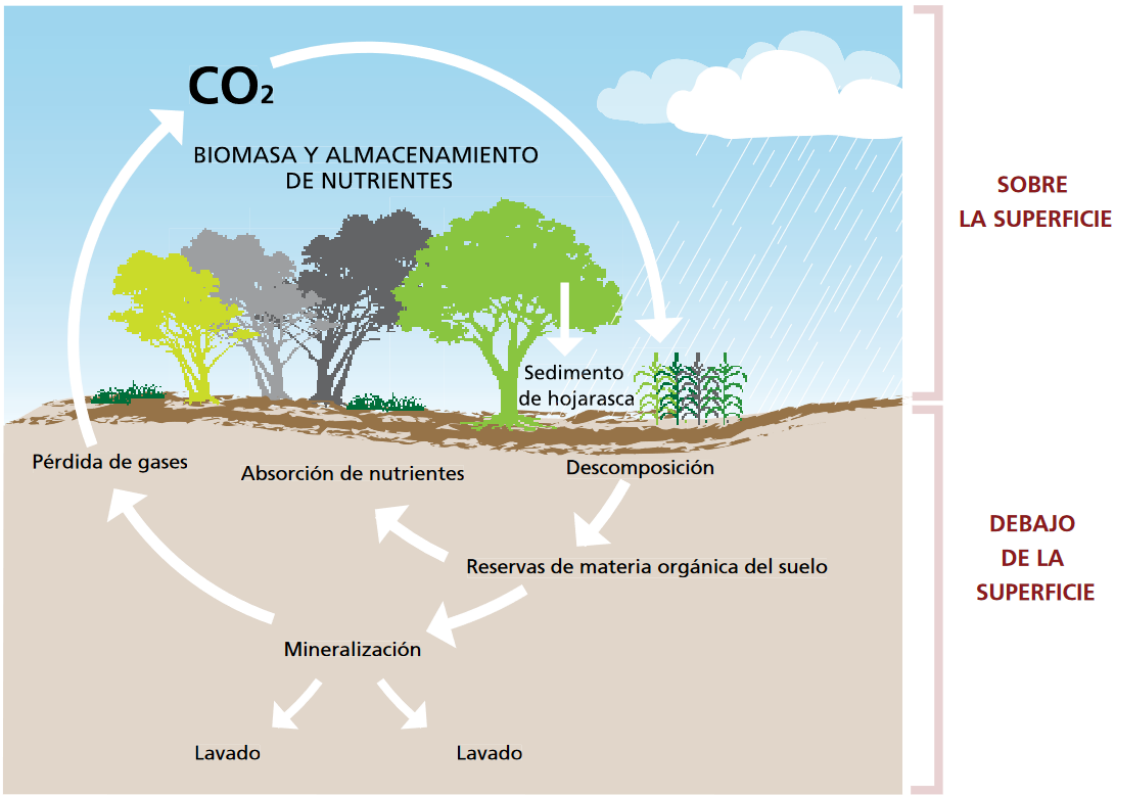


Figure 1.1 Esquema simplificado del almacenamiento de carbono en los ecosistemas terrestres debajo y sobre la superficie. *Fuente*: FAO 2007

El balance de carbono de un ecosistema se considera como la diferencia ente la pérdidas y ganancias de carbono, determinando si está actuando como una fuente (emisión de C a la atmósfera) o como un sumidero (absorción de C de la atmósfera) (Lal, 2004). La cuantificación y predicción de este balance resultan complejas debido a falta de información espacialmente explicita a escalas adecuadas, así como la limitación del conocimiento en los procesos implicados de intercambio de C con la atmósfera, en particular en los suelos (Heimann and Reichstein, 2008).

En las últimas décadas, se han desarrollado una gran cantidad de métodos para la cuantificación, modelización y mapeo del stock de carbono ecosistémico, los cuales son esenciales en el diseño e implementación de políticas regionales y nacionales (Crossman et al., 2011). Sin embargo, muchos de ellos adolecen de procesos ecológicos claves, y los resultados disponibles se basan en tablas de datos aplicados a diferentes clases de usos de suelo (Sun and Liu, 2020). Actualmente, la estimación estándar del stock de C en estos modelos está más ampliamente descrita para la biomasa aérea (AGB, por sus siglas en inglés) (450–650 PgC a nivel global), empleando estimaciones más generales para el 70% del C global almacenado en los suelos (1500–2400 PgC; Ciais et al., 2014). Los avances en estos modelos y mapas dependerán de la integración de los procesos ecológicos más relevantes en la caracterización de la variación espacial y temporal del carbono, como son los relacionados con el stock de C en suelos.

1.3. Cuantificación del carbono ecosistémico

Entre los ecosistemas terrestres, los bosques representan una posición central en el ciclo del carbono global, siendo la mayor reserva con cerca del 45% de todo el carbono secuestrado a nivel global (Miquelajauregui, 2013; Zhao et al., 2019). El stock de carbono en los ecosistemas forestales principalmente incluye el almacenamiento en la vegetación, el suelo y el litter. Este último supone apenas un 5% del carbono total forestal (Lafleur et al., 2018). Para la estimación de estos almacenamientos o reservorios se han empleado métodos diferentes basados principalmente en diferentes tipos de ecosistemas y escalas espaciales.

Estimación del almacenamiento del carbono en la vegetación

Los métodos para la estimación del carbono en la vegetación se pueden dividir básicamente en tres tipos: estimación basada en inventarios, basada en información satelital o basada en procesos.

La estimación del carbono de la vegetación basada en inventarios es un método clásico empleado principalmente para muestreos de superficies limitadas, como escalas regionales. Estos métodos están basados en la estimación de determinadas variables de la masa vegetal, tales como densidad, volúmenes, tipos de vegetación, altura, etc., con la consecuente conversión a biomasa y carbono mediante el uso de ecuaciones alométricas. Uno de los problemas detectados en estos métodos es su tendencia a sobreestimar la cantidad de carbono en la vegetación al asignar su valor de muestreo a superficies mayores. Esto es debido a que los resultados de muestreo en campo suelen ser mayores a

la media estimada a nivel regional o nacional (Dixon et al., 1994; Fernández-Landa et al., 2018).

La teledetección o percepción de información de elementos de la superficie terrestre desde sensores remotos ha convertido en una técnica alternativa ampliamente utilizada de cuantificación de biomasa aérea y almacenamiento de carbono (Xiao and Moody, 2004). Esta técnica está basada en la construcción de relaciones empíricas entre la información obtenida de la radiación electromagnética y determinados atributos de masa vegetal, integrando mediciones alométricas y datos de campo. Básicamente hay tres tipos de sensores que se utilizan para la cuantificación del carbono: sensores ópticos, radar y lidar. La precisión de la información obtenida de los diferentes sensores varía en función de la escala espacial y temporal de los mismos. Por lo que la estimación del stock carbono en la vegetación está muy condicionada a estas características (Matasci et al., 2018).

Mediante la parametrización de modelos en base a las principales variables ecosistémicas que influyen en el almacenamiento de carbono, se construyen modelos estadísticos de interpolación espacial que estiman la distribución de la biomasa forestal. Estos modelos hacen referencia principalmente a métodos geoestadíticos, de aprendizaje supervisado o basados en mecanismos, los cuales describen cuantitativamente el proceso del ciclo del carbono forestal. Para este tipo de métodos es necesario disponer de datos de vegetación de alta calidad (Tian et al., 2015).

Estimación del almacenamiento del carbono en el suelo

Los métodos utilizados en la cuantificación de la distribución espacial del carbono en suelo pueden agruparse en cuatro: estimación basada en tipos de suelo, en unidades de mapeo, en modelización empírica y en modelización basada en procesos.

La estimación del almacenamiento del carbono orgánico del suelo (SOC) para una unidad taxonómica requiere datos de muestreo de diferentes tipos de suelo. Posteriormente, se estima el almacenamiento de carbono a nivel regional en base al mapa de tipos de suelo (Bohn, 1982; Muñoz-Rojas et al., 2012).

Otros métodos se basan en la caracterización de unidades de densidad de carbono orgánico del suelo, de acuerdo a la distribución geográfica de muestreos de perfiles edáficos y factores ambientales, como la vegetación y los factores climáticos. El resultado final es el producto de la densidad del carbono por el área de cada unidad (Post and Mann, 2005).

Los modelos empíricos se basan en la relación entre el SOC y los numerosos factores ambientales que afectan a su almacenamiento, tales como factores climáticos, topográficos, vegetación, propiedades de suelo. Para establecer estas relaciones estadísticas se precisa de mediciones repetidas en campo y distribuidas espacialmente tanto del contenido del SOC como de los factores ambientales a relacionar (Sun and Liu, 2020).

Finalmente, los métodos basados en procesos compartimentan la estructura del suelo en diferentes reservorios de carbono para simular el proceso del ciclo del carbono del suelo

en función de su tasa de descomposición. Si se obtienen los datos necesarios para alimentar dicha simulación, estos modelos pueden ser empleados para estimar el carbono en diferentes escenarios (pasado, presente y futuro). RothC, SOMM, Yasso, CENTURY y ROMUL son una buena representación de estos modelos (Xu et al., 2018).

Sin embargo, todos estos modelos precisan de datos de mediciones directas de carbono en suelo. Dada su gran variabilidad, la alta densidad de puntos de muestreo necesaria para obtener estimaciones precisas y actualizadas hace que los inventarios resulten caros y operacionalmente complejos (Smith et al., 2020).

1.4. Necesidad de estandarización de procesos en la modelización del carbono.

Existen diversas iniciativas, tanto a nivel europeo como global, que promueven la reducción del CO_2 mediante el secuestro de C ecosistémico a través de la implantación de programas prácticos (Minasny et al., 2017) demostrándose que estos sistemas son una oportunidad para la regulación climática. La finalidad de muchas de estas iniciativas es poder evaluar y monitorear el impacto de las medidas de mitigación y documentar dicha información. Esto resulta más relevante en aquellas que incluyen estos resultados en los mecanismos financieros donde se realizará un pago por reducir las emisiones, compensarlas o aumentar las existencias de C ecosistémico (Thamo and Pannell, 2016). Sin embargo, dentro de estos mecanismos de reducción y mitigación existe una disparidad referente a los métodos de cuantificación, monitoreo y verificación del carbono tanto a nivel nacional como global. Los métodos de estimación de biomasa están más desarrollados y estandarizados que los del carbono en suelo (Gibbs et al., 2007). Concretamente, desde la Oficina de Cambio Climático en España (MITECO), existe un interés manifiesto por incorporar la cuantificación tanto del stock actual del C en suelo como de su secuestro potencial, en sus mecanismos regulados para el cálculo, reducción y compensación de emisiones. Dicho sistema está basado actualmente en las absorciones de la cubierta forestal exclusivamente.

Además de la disparidad de métodos para los distintos stocks, algo que se evidencia en estas iniciativas de reducción de carbono es la complejidad de las interacciones entre el carbono aéreo y el carbono del suelo con la atmósfera. Esto se debe principalmente, a que estas interacciones se rigen por otros factores también complejos tales como el clima, la topografía, las características del suelo o las prácticas de manejo (forestales o agrícolas). Los efectos de las interacciones entre todos estos factores implican grandes incertidumbres en su cuantificación (Fujisaki et al., 2015).

Dada la gran variedad de aproximaciones para cuantificar el stock de carbono en los ecosistemas, es necesario la estandarización de metodologías previamente testadas que aseguren su interoperabilidad a dominios espaciales diversos, desde lo local a lo global, para facilitar el intercambio y la comparación de resultados (Bispo et al., 2017; Montanarella, 2015; Vargas et al., 2017).

1.5. Objetivos y preguntas de investigación

El objetivo principal de esta tesis es avanzar en la estrategia metodológica para la cuantificación del carbono orgánico en ecosistemas terrestres, basada en técnicas estandarizadas a diferentes escalas espaciales y de gestión, así como establecer modelos dinámicos capaces de predecir los resultados bajo diferentes escenarios de gestión.

Consideramos que la incorporación de metodologías que combinen el uso de recientes métodos estadísticos de aprendizaje automático con técnicas digitales y de teledetección, así como la automatización de tareas, permitirá mejorar los actuales modelos el stock del carbono orgánico ecosistémico. Esto facilitará su aplicación a cualquier escala espacial de trabajo y la realización de proyecciones temporales acerca de los cambios en el stock.

Para la consecución de dicho objetivo, los objetivos parciales a abordar son:

- Desarrollar un enfoque metodológico para mejorar la evaluación y el mapeo de la biomasa aérea (AGB) a nivel nacional mediante la combinación de datos LiDAR de alta resolución con imágenes de satélite hipertemporales y de resolución moderada en los bosques mediterráneos.
- Evaluar diferentes estrategias metodológicas basadas en técnicas de mapeo digital de suelo (DSM) y datos de satélite para la elaboración de mapas locales de COS que mejoren las estimaciones actualmente disponibles, derivadas de bases de datos tanto regionales como globales.

  Establecer el marco metodológico más robusto para la generación de mapas de COS.
- Generar mapas de almacenamiento y distribución espacial del COS a nivel nacional basados en las mejores estimaciones posibles y los recursos disponibles mediante técnicas estandarizadas que permitan la interoperabilidad de los resultados.
- Identificar áreas potenciales de fuentes y sumideros de carbono del suelo mediante la proyección de la distribución espacial de las tasas de secuestro de COS a nivel nacional, que sirvan de apoyo en la toma de decisiones de prácticas de gestión sostenible o la conservación del suelo.

Capítulo 2. Metodología general

2.1. Área de estudio

Esta tesis se enmarca en el territorio de España Peninsular, como representación de un caso de estudio de la zona Mediterránea. La zona mediterránea es considerada particularmente vulnerable al cambio climático, ya que los escenarios de predicción acentúan las sequías, las olas de calor y la escasez de agua a la vez que inundaciones más frecuentes (Abd-Elmabod et al., 2020).

Concretamente, los escenarios climáticos futuros para España Peninsular (Figure 2.1), situada en la región Mediterránea Europea, prevén serias limitaciones respecto a la disponibilidad de agua y un incremento del riesgo de desertificación, variables íntimamente ligadas a la variación del contenido de carbono orgánico de los ecosistemas (Ferreira et al., 2022). Por lo que la gestión de los usos agrícolas y forestales, los cuales ocupan un 33% y un 59 % de la superficie total, respectivamente (MAPA, 2021), resulta crucial tanto para la conservación como para el fomento del secuestro de carbono ecosistémico.

Las características específicas de las áreas tanto local como nacional abordadas en esta tesis se describen en sus capítulos correspondientes.

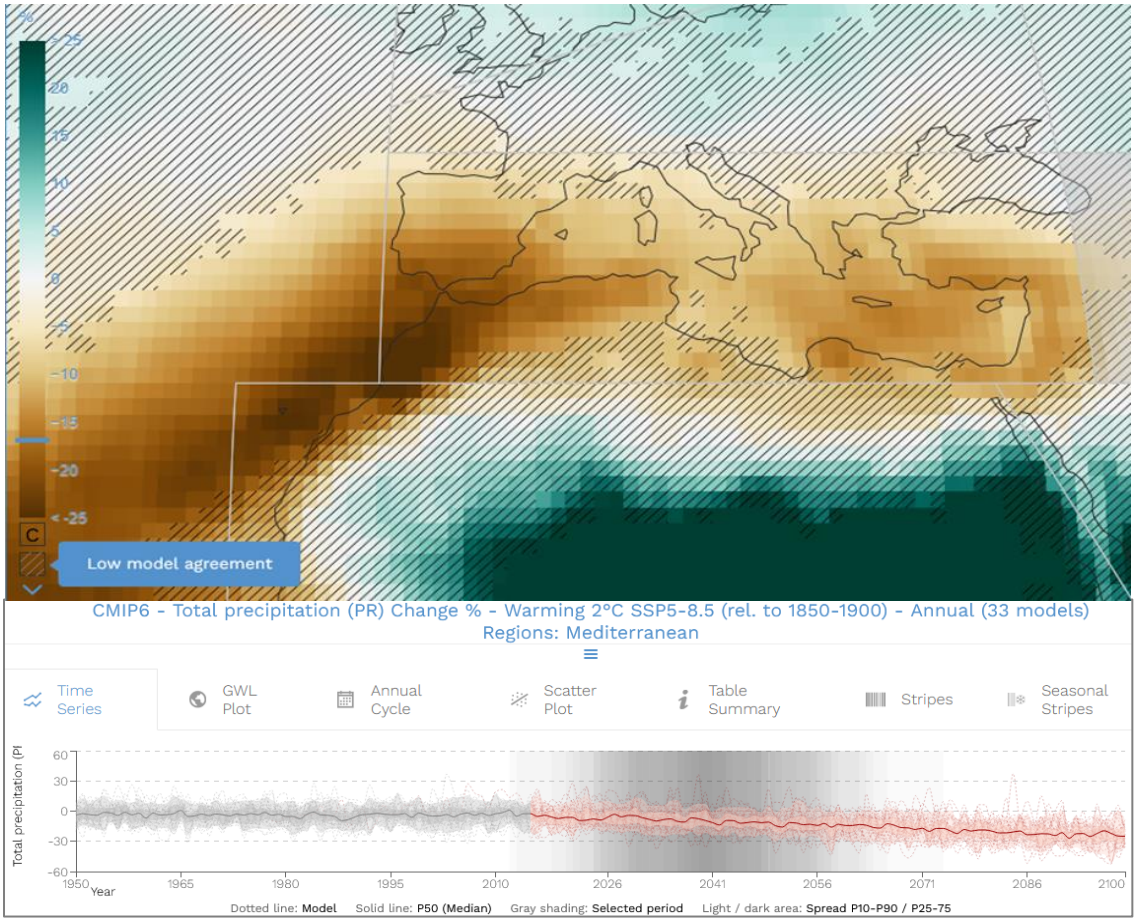


Figure 2.1 Escenario climático futuro para la región Mediterránea basado el porcentaje de cambio de la precipitación total (PR). Realizado a partir de las proyecciones globales del Sexto Informe de Evaluación del IPCC (2021). Fuente: Atlas interactivo del IPCC: síntesis regional.

2.2. Marco metodológico

Para el desarrollo del objetivo general propuesto es imprescindible profundizar en el conocimiento detallado de la dinámica del carbono en los ecosistemas, y poder obtener las mejores estimaciones dentro de los límites establecidos por el conocimiento científico actual y los recursos disponibles a nivel nacional, reduciendo la incertidumbre en lo posible.

El marco metodológico general que encuadra esta tesis se centra en la estimación del carbono de los dos reservorios ecosistémicos principales: la vegetación y el suelo (Figure 2.2). Para alcanzar los objetivos descritos anteriormente, esta metodología se ha desglosa en función de las siguientes fases:

1- Fase 1: Generación de algoritmos para la estimación de biomasa aérea de especies forestales a partir de datos LiDAR para su extrapolación a escala nacional.

2- Fase 2: Desarrollo de metodologías mediante técnicas digitales y de inferencia espacial para la elaboración de mapas de carbono en suelo a escala nacional.

    2.1 Optimización de una base de datos nacional de información edáfica patrimonial. Armonización y filtrado de la información patrimonial.

    2.2 Modelización del stock de carbono orgánico en el perfil del suelo en base a sus características analíticas y variables ambientales.

    2.3 Producción de mapas digitales de distribución del stock de carbono en suelo a diferentes escalas de detalle: de local a nacional.

3- Fase 3: Evaluación de la dinámica del secuestro de carbono orgánico del suelo bajo diferentes escenarios de gestión, para la generación del mapa potencial de secuestro de carbono.

Fase 1: Generación de algoritmos para la estimación de biomasa aérea de especies forestales a partir de datos LiDAR para su extrapolación a escala nacional.

El LiDAR (Light Detection And Ranging) es un sistema activo de detección remota basado en un sensor láser que permite obtener, en tiempo real, mayor densidad de medidas geolocalizadas de las superficies que cualquier otro sistema conocido. El LiDAR presenta la ventaja frente a otros sensores de poder penetrar en la cubierta vegetal y capturar de esta forma información de diferentes estratos de vegetación y del suelo.

Esto hace que la tecnología LiDAR esté siendo ampliamente utilizada en el sector forestal (Hollaus et al., 2007; Holmgren et al., 2003) tanto por su capacidad para obtener variables dasométricas con alta precisión como por su economía para la inventariación de grandes superficies.

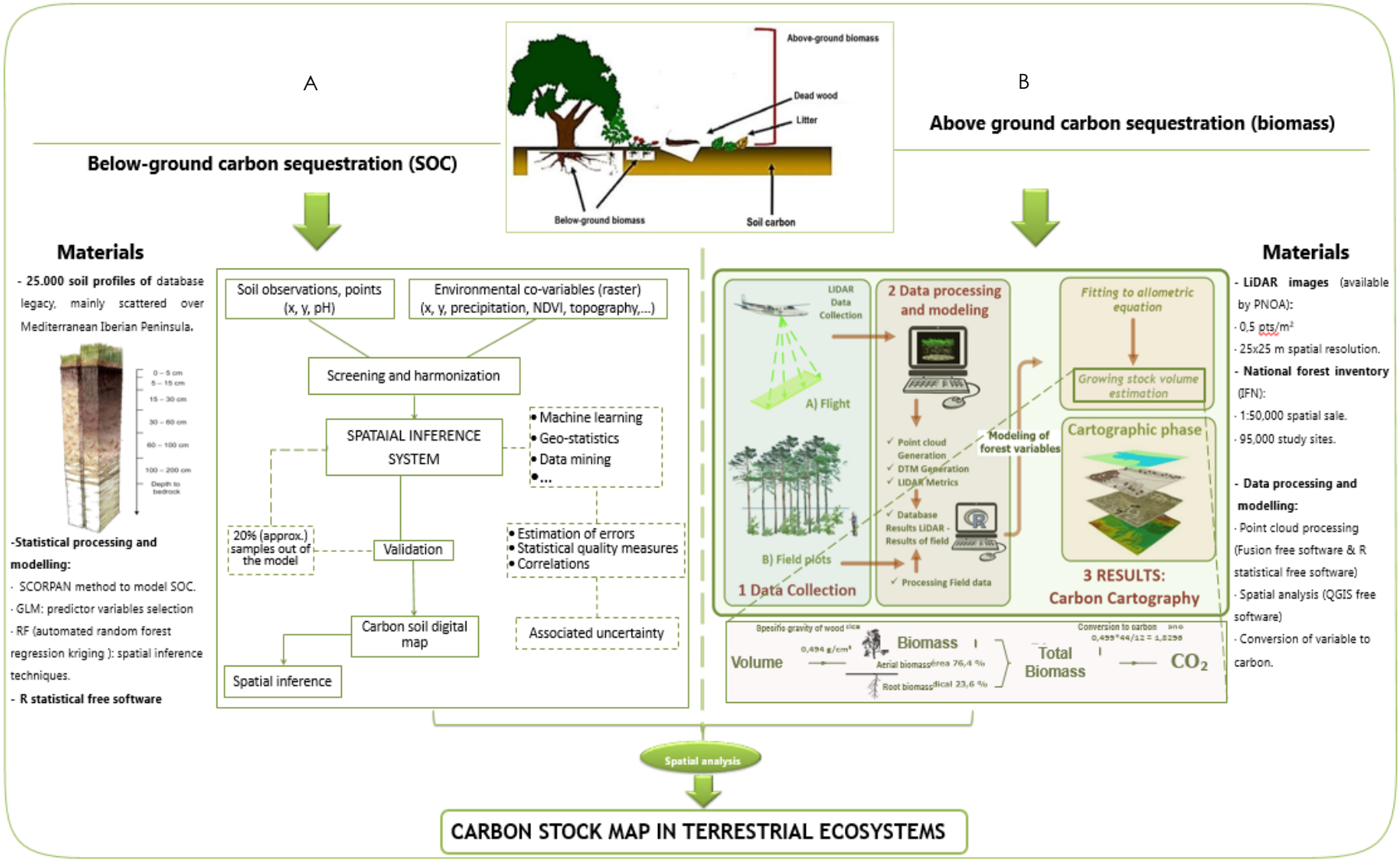


Figure 2.2 Marco metodológico para la estimación de la distribución espacial del carbono ecosistémico.

España tiene una situación privilegiada en cuanto a la disponibilidad de datos LiDAR ya que dentro del Plan Nacional de Ortofotografía Aérea (PNOA), que se realiza en cooperación y cofinanciado entre la Administración General del Estado y las Comunidades Autónomas, hay un producto LiDAR. La primera cobertura nacional de LiDAR se finalizó en el año 2015 con una densidad de 0,5 puntos/m2. Si las condiciones económicas lo permiten, las coberturas nacionales tendrán una periodicidad sexenal.

La metodología propuesta para la estimación de biomasa aérea (Figure 2.2 B) se basa en la utilización de datos LiDAR mediante el ajuste de regresiones entre los estadísticos que describen la nube de puntos LiDAR, con las variables obtenidas de inventarios por métodos de masa, esto es, altura dominante, número de árboles, área basimétrica, volumen, biomasa, etc. (Hyyppä et al., 2008; Rodríguez et al., 2014). Aplicando estas regresiones a los estadísticos LiDAR procesados para la zona de estudio se generó una cartografía continua para cada una de las variables modelizadas, cuyo resultado fue una imagen digital con información espacialmente continua de 25 m x 25 m de tamaño de celda (Asner et al., 2011; Asner and Mascaro, 2014).

La red de parcelas de datos LiDAR resultantes de inventarios de masa fue completada con parcelas del Inventario Nacional Forestal (INF). De cara a poder asignar correctamente los modelos generados para las distintas especies, se utilizará el Mapa Forestal de España, trabajando sobre los estratos forestales definidos en este producto del Ministerio de Agricultura y Medio Ambiente.

La extrapolación de datos LIDAR locales a escala nacional se realizó mediante la integración de los datos locales LiDAR con variables espectrales (MODIS) de alta temporalidad y cobertura espacial, mediante técnicas de aprendizaje automático. Esta modelización se describe más detalladamente en el capítulo 3.

Fase 2. Desarrollo de metodologías mediante técnicas digitales y de inferencia espacial para la elaboración de mapas de carbono en suelo.

Para obtener la distribución espacial del stock de carbono a nivel nacional, en función de variables biofísicas, se siguieron las siguientes subfases: la generación de una base de datos edáfica, la modelización del carbono orgánico y su extrapolación espacial (Figure 2.2 A).

Fase 2.1. Optimización de una base de datos nacional de información edáfica. Armonización y filtrado de la información patrimonial.

Como paso previo a la modelización, dada la gran variedad de fuentes de información edáfica a nivel nacional, se realizó una recopilación, depuración y homogenización de los datos de perfiles de suelo más representativos. La base de datos final resultó de la estandarización de sus datos con el fin de ser consistente, fiable y asimilable para los modelos. En esta fase inicial para la modelización se contó con una base de datos de perfiles aportados por distintos organismos de la Administración (gestión e investigación) nacionales y autonómicos.

Fase 2.2. Modelización del stock de carbono orgánico en el perfil del suelo en base a sus características analíticas y variables ambientales.

El desarrollo de esta fase se basó en técnicas del mapeo digital de suelos, dentro del marco de referencia del modelo conceptual SCORPAN (McBratney et al., 2003). Según este modelo algunas propiedades edafológicas, como el carbono orgánico, pueden estimarse con aceptable precisión mediante su correlación con otros parámetros considerados como formadores del suelo medidos en la misma localización. Por tanto, para modelizar la distribución espacial del stock de carbono se definieron las principales variables ambientales características del entorno, como factores formadores edáficos responsables de las propiedades actuales de los suelos. Desde el punto de vista de la inferencia espacial, estos factores deben ampliarse con la inclusión de la posición geográfica.

Dada la gran variedad de fuentes de información digital, las variables medioambientales incluidas en el modelo como factores formadores del suelo han variado en base a las características de cada estudio. Dichas variables han sido descritas en los capítulos correspondientes. De forma general, las variables utilizadas hacen alusión a los factores morfológicos del terreno, climáticos, litológicos y de la vegetación. La parametrización de este último se realizó mediante información referente a usos de suelo, mapas de biomasa o índices de vegetación derivados de información de teledetección. Para evitar la redundancia de las variables y mejorar el carácter predictivo del modelo (Gregorutti et al., 2017), se procedió a una selección previa de variables.

La modelización de la distribución espacial del carbono en base a las variables seleccionadas se realizó mediante la aplicación de un amplio abanico de funciones de predicción para explicar las relaciones entre las variables edáficas y ambientales. Estos modelos se basaron tanto en métodos lineales como no-lineales, haciendo uso de métodos de aprendizaje automático. La validación de los modelos se basó en la precisión de la capacidad predictiva de los mismos.

Fase 2.3. Producción de mapas digitales de distribución del stock de carbono en suelo a diferentes escalas de detalle.

Tras la modelización, la extrapolación de la información puntual a una superficie espacialmente continua dio lugar a la generación del mapa de la estimación espacial del carbono y el del valor de incertidumbre asociada. Dichos mapas contemplan la distribución de las variables de carbono concentración (g/kg) y stock (TC/ha), a diferentes escalas espaciales, mediante la aplicación de modelos predictivos diferentes en función de las circunstancias específicas de cada estudio.

Fase 3. Evaluación de la dinámica del secuestro de carbono orgánico del suelo, bajo diferentes escenarios de gestión, para la generación del mapa potencial de secuestro de carbono.

Para la determinación de las áreas potenciales de acumulación de carbono orgánico del suelo a nivel nacional, se estimó la cantidad de carbono secuestrado en un período de 20 años (2020-2040) mediante el modelo basado en procesos de RothC. Para la fase inicial de este proceso de modelización se utilizó el mapa de stock previamente estimado (T/ha). Esta modelización se realizó bajo diferentes supuestos de gestión y siguiendo métodos

estandarizados (Figure 2.3) cuyos resultados resulten comparables e interoperables con otros países (FAO, 2020a).

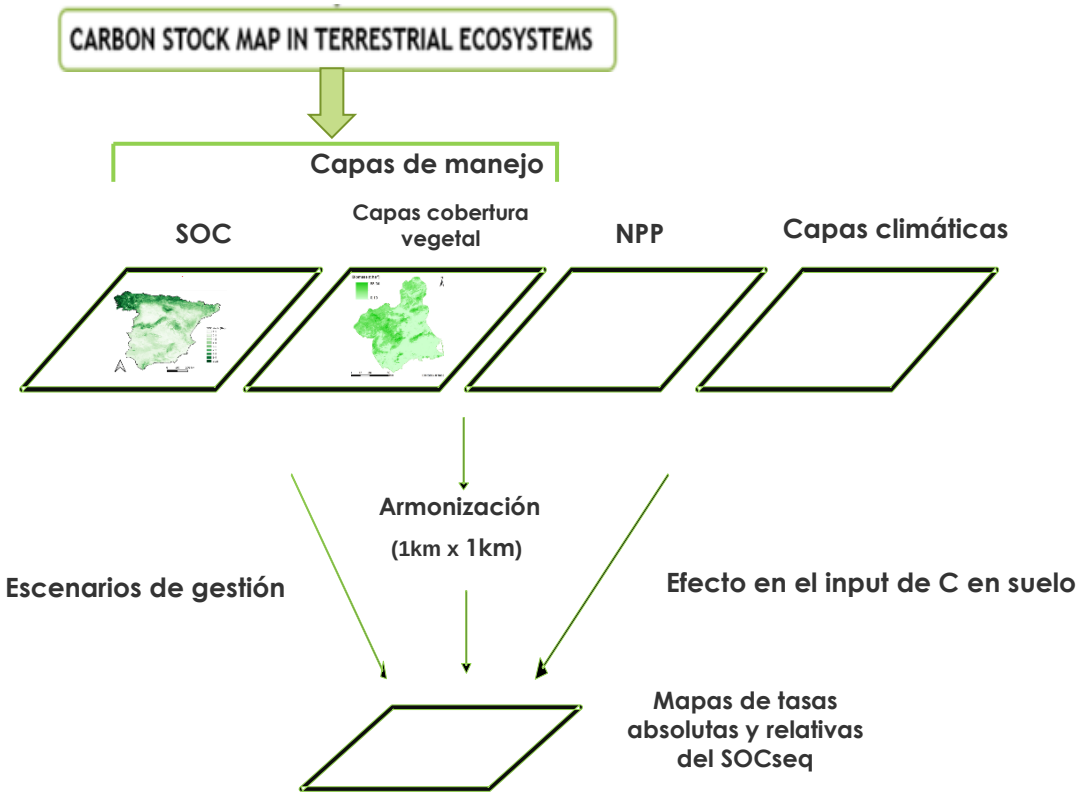


Figure 2.3 Esquema metodológico de la estimación del secuestro potencial del carbono orgánico del suelo para España peninsular a 1 km de resolución espacial.

De forma generalizada para esta tesis, en todos los procesos y modelizaciones contenidos en el desarrollo metodológico, los programas informáticos (software) utilizados fueron de código abierto y dominio libre, principalmente. De entre los principales recursos técnicos utilizados se encuentran:

- Software R para cálculos estadísticos y modelización.
- Software SAGA y QGIS para el procesado de la información georreferenciada.
- Software FUSION para los procesados LiDAR.

Así mismo, debido al gran volumen de datos utilizados, una parte importante de los procesos han sido realizados gracias a los servicios de supercomputación facilitados por el Centro Informático Científico de Andalucía (CICA).

2.3. Estructura de la tesis

Esta tesis está estructurada en 8 capítulos. El capítulo 1 y 2 hacen referencia a la introducción y metodología generales de la tesis. Los capítulos 3, 4, 5 y 6 redactados en inglés responden los contenidos principales abordados en esta tesis, correspondientes a artículos originales de investigación que se encuentran publicados, o pendientes de envío, en revistas que figuran en los listados de Journal Citatioin Reports (JRC). Finalmente, en el capítulo 7 se presenta una discusión general de la tesis y en el capítulo 8 se enumeran las conclusiones generales.

Los contenidos principales abordados en cuatro bloques quedan resumidos en la Table 2.1

Table 2.1 Resumen general de los contenidos de los capítulos 3, 4, 5 y 6 de la tesis

Capítulo		Objetivos generales (1)	Productos finales (2)	Área de estudio
Capítulo 3	Improving Aboveground Forest Biomass Maps: From High-Resolution to National Scale	Desarrollar un enfoque metodológico para mejorar la evaluación y el mapeo AGB a nivel nacional, combinando datos LiDAR de alta resolución con imágenes de satélite hipertemporales y de resolución moderada	Mapa AGB (25 m) basado en datos LIDAR-PNOA Mapa AGB (250 m), y su incertidumbre asociada basados en MODIS	Región de Murcia (España)
Capítulo 4	Testing Modeling Strategies for Soil Organic Carbon at Different Spatial Scales in a Mediterranean Area (Southern Iberian Peninsula)	Evaluar diferentes estrategias metodológicas basadas en técnicas DSM y datos de satélite para la elaboración de mapas locales de COS. Establecer el marco metodológico más robusto para la generación de mapas de COS.	Mapas de concentración de COS (g/kg) a 0-30 cm de profundidad. Mapas de stock de COS (tC/ha) a 0-30 cm de profundidad. Resolución espacial 100 m	Región de Murcia (España)
Capítulo 5	Spatial Distribution Modeling and Quantification of Soil Organic Carbon in Peninsular Spain	Generar mapas de almacenamiento y distribución espacial del COS a nivel nacional basados en las mejores estimaciones posibles y los recursos disponibles mediante técnicas estandarizadas que permitan la interoperabilidad de los resultados	Mapa de concentración de COS (g/kg) a profundidades de 0-30 cm y 30-100 cm. Mapa de stock de COS (tC/ha) a 0-30 cm y a su profundidad efectiva. Resolución espacial 90 m	España peninsular

Capítulo	Objetivos generales [1]	Productos finales [2]	Área de estudio
Capítulo 6 Soil Organic Carbon Sequestration Potential National Map under Future Scenario Projections	Identificar áreas potenciales de fuentes y sumideros de carbono del suelo mediante la proyección de la distribución espacial de las tasas de secuestro de COS a nivel nacional, que sirvan de apoyo en la toma de decisiones de prácticas de gestión sostenible o la conservación del suelo.	Mapas de secuestro de COS (tC/ha) a 30 cm para escenarios BAU y SM Mapa de tasas de secuestro absoluto de COS para escenarios BAU y SM Mapa de tasas de secuestro relativo de COS (tC/ha) para escenario SM Resolución espacial 1 km	España peninsular

(1) COS= carbono orgánico del suelo; AGB= aboveground biomass/biomasa aérea; DSM= digital soil mapping/mapeo digital del suelo; **(2)** BAU= business as usual/escenario sin gestión; SM= soil management/escenario con gestión sostenible

Capítulo 3. Improving Aboveground Forest Biomass Maps: From High-Resolution to National Scale

Este capítulo reproduce el texto del siguiente manuscrito:

Durante, P.; Martín-Alcón, S.; Gil-Tena, A.; Algeet, N.; Tomé, J.L.; Recuero, L.; Palacios-Orueta, A.; Oyonarte, C. Improving Aboveground Forest Biomass Maps: From High-Resolution to National Scale. **Remote Sens. 2019, 11, 795**

ABSTRACT

Forest aboveground biomass (AGB) estimation over large extents and high temporal resolution is crucial in managing Mediterranean forest ecosystems, which have been predicted to be very sensitive to climate change effects. Although many modeling procedures have been tested to assess forest AGB, most of them cover small areas and attain high accuracy in evaluations that are difficult to update and extrapolate without large uncertainties. In this study, focusing on the Region of Murcia in Spain (11,313 km2), we integrated forest AGB estimations, obtained from high-precision airborne laser scanning (ALS) data calibrated with plot-level ground-based measures and bio-geophysical spectral variables (eight different indices derived from MODIS computed at different temporal resolutions), as well as topographic factors as predictors. We used a quantile regression forest (QRF) to spatially predict biomass and the associated uncertainty. The fitted model produced a satisfactory performance (R^2 0.71 and RMSE 9.99 t·ha−1) with the normalized difference vegetation index (NDVI) as the main vegetation index, in combination with topographic variables as environmental drivers. An independent validation carried out over the final predicted biomass map showed a satisfactory statistically-robust model (R^2 0.70 and RMSE 10.25 t·ha−1), confirming its applicability at coarser resolutions.

Keywords: Mediterranean forest; climate change; ALS; MODIS; quantile regression forest; uncertainty.

3.1. INTRODUCTION

The Mediterranean basin represents a hotspot of biological diversity, being a socioecological system very sensitive to climate change effects (Barredo et al., 2018). A keystone concern is how to develop adaptive planning for the assessment, monitoring, and management of organic carbon in ecosystems, which is essential to achieve climate change commitments at a national level. Particularly in terrestrial ecosystems, forests play a basic role as carbon sinks containing about 80% of global terrestrial aboveground biomass (AGB). However, such information is difficult to produce, and the uncertainties about magnitude and location are often large (Ometto et al., 2014; Saatchi et al., 2007). Much of this uncertainty is due to the lack of detailed information about the spatial distribution of carbon stored in biomass. Thus, forest biomass is an important measure for environmental management to provide more insights into the amount and spatial distribution of carbon storage for supporting future climate change mitigation actions (Canadell and Raupach, 2008).

The use of light detection and ranging (LiDAR) technology aimed at studying forested ecosystems started in the 1970s, especially airborne laser scanning (ALS) systems. Airborne laser-based acquisitions have been extensively used in forestry inventories following area-based approaches, such as in AGB mapping (González-Jaramillo et al., 2018; Næsset and Gobakken, 2008). This is due to ALS's high sensitivity to structural features (Bottalico et al., 2017; White et al., 2016) and high precision ability to measure stand parameters in three dimensions previously calibrated with plot-level ground-based measures (Kauranne et al., 2017; Molina et al., 2015). Moreover, several previous studies have proven the use of existing data sets, particularly national forest inventory plots, to be a feasible option to provide training data for stand-level ALS inventories (e.g., Maltamo et al., 2009; Nelson et al., 2012). ALS data combined with machine learning modeling techniques are able to find substantial accuracy in forest biomass estimates at local scale (Fassnacht et al., 2018; Hudak et al., 2012). Applications linking local and national—even global—scales allow the design of dynamic assessment and management planning, therefore increasing the accuracy and predictive capacity of global models predicting forest AGB. Despite technological advancement, the high cost of acquisition of ALS data results in much higher revisit time compared to satellite data, which reduces their continuous temporal and spatial availability, therefore making their use at national scale inoperative. Thus, ongoing efforts are needed to solve the challenges involved in integrating environmental assessment and management at concordant scales in order to address climate change at a national level (Cash and Moser, 2000).

In the last few years, different studies have integrated optical satellite data into different spatial and temporal resolutions for land surface monitoring in a highly operative way (Lu, 2006). Satellite remote sensing technology is considered as the most effective approach to estimate biomass, since it provides adequate methods to produce a high-temporal characterization of vegetation attributes over spatially continuous large areas (Xiao and Moody, 2004). The information provided by the moderate resolution imaging spectroradiometer (MODIS) has been extensively applied to monitor forests at regional and large scales since 2010, therefore enabling the derivation of biophysical parameters which are highly valuable for forest management such as forest biomass (Blackard et al., 2008; Chi

et al., 2015). MODIS offers a short revisit time, and the information acquired in several spectral bands can be summarized in different vegetation indices (VIs) that can be used as proxies of photosynthetic activity of vegetation (Jin et al., 2014). The theoretical basis for empirical-based VIs is derived from typical spectral reflectance signatures of contrast between blue-red and near-infrared wavelengths' response. This contrast can be quantified through the use of ratios, differences, weighted differences or linear band combinations (Didan, 2015). Hence, a MODIS-based hypertemporal series of VIs allow to extract valuable information about vegetation phenology (Reed et al., 2009) and to estimate vegetation biomass (Li et al., 2015). On the one hand, the normalized difference vegetation index (NDVI) (Tucker, 1979) has been widely used in many forestry applications. The NDVI is the ratio of contrasting reflectance between maximum absorption of red wavelength due to chlorophyll pigments and maximum reflectance of infrared wavelength owing to leaf cellular structure (Paruelo et al., 1997; Piao et al., 2007). These characteristics have been highly correlated with green biomass, green leaf area index (LAI), and percent green vegetation cover (Saatchi et al., 2007; Yan et al., 2015). Since forest biomass is conditioned by temporal dynamics, the annual profile of the NDVI time series can be considered as an indicator of seasonal dynamics of forest biomass. Furthermore, NDVI annual mean (as net primary production estimator) and its variation coefficient (as indicator of vegetation seasonality) are commonly used in forest biomass assessment (Alcaraz-Segura et al., 2009). On the other hand, NDVI shows saturation in dense and multilayered forest canopy (Thenkabail et al., 2000). Therefore, substantial efforts have been made to improve and develop new vegetation indices, such as the enhanced vegetation index (EVI), which is considered a modified NDVI. EVI improves sensitivity across regions with high biomass and shows a high vegetation monitoring capability through a decoupling of the canopy background signal with a reduction in atmospheric influences (Hui Qing Liu and Huete, 1995). The soil adjusted vegetation index (SAVI) (Huete, 1988) was generated to minimize the soil reflectance effect. Additionally, Mutanga and Skidmore (2004) noted that NDVI is limited by its nonlinear response, making it insensitive to differences at very low and high densities. Hence, other vegetation indices are considered more linearly related to biophysical parameters of vegetation, such as the renormalized difference vegetation index (RDVI) (Roujean and Breon, 1995), which was developed to take combined advantage of the difference vegetation index (DVI = NIR – Red) (Jordan, 1969) and the NDVI; and the modified simple ratio (Chen, 1996), proposed over the RDVI because it shows more sensitivity to various biophysical parameters.

In combination with dynamic and easily updated remotely-sensed VIs, terrain variables have also been used in many studies predicting AGB as supplemental information aimed at increasing predictive power and statistical discrimination (Matasci et al., 2018). Since topography controls the main landscape processes related to structural characteristics of vegetation, such as water distribution and potential solar radiation (Moore et al., 1991), the digital elevation model (DEM) and the derivation of DEM-based predictor variables partially explain spatial patterns in AGB (Mendoza-Ponce et al., 2018).

On the basis of the above, the use of high-resolution ALS data is a widely proven tool to collect and subsequently characterize structural forest stand attributes, with some limitations due to cost, revisit time, logistics, and data volumes involved in monitoring and mapping large areas (Wulder et al., 2012). Optical remote sensing data provide abundant and highly frequent observations to monitor extensive forest areas. However, coarse-resolution satellite

data are often limited by the ground-based validation and nature of the measures, which can be largely insensitive to vertically distributed attributes (Navarro et al., 2019). The main hypothesis of this study is that there is a robust correlation between ALS-based forest AGB and VIs derived from MODIS data. Therefore, forest AGB spatial distribution can be estimated in those places were ALS information is not available. This study aimed to develop a methodological approach to improve nationwide assessment and mapping of AGB through combining high-resolution ALS data with hypertemporal and moderate-resolution satellite images in Mediterranean forests. A two-stage upscaling approach was applied where ALS-based forest AGB assessments, calibrated with plot-level ground-based measures, were used to train a machine learning method (quantile regression forest). Predictor variables consisted of MODIS–derived indices and topographic factors. As an approach of the estimation accuracy in the upscaling processes, the estimated variance of the AGB estimator was assessed according to Saarela et al. (2016). This methodology was mainly developed with open source software and always considering free available data.

3.2. MATERIALS AND METHODS

3.2.1. Study Area

Our study case was a Mediterranean region in Southeastern Iberian Peninsula (Region of Murcia), which is a particularly vulnerable area to global climate change threats due to its environmental characteristics (Barredo et al., 2018). The area is dominated by a dry, semiarid climate characterized by warm mean annual temperatures (ca. 18 °C), scarce annual rainfall (ca. 300–350 mm), and high mean annual evapotranspiration (ca. 900 mm) (Papadakis, 1966). Climate variation is related to a diverse orography, which combines mountainous areas (up to 2000 m asl), high plateaus (500–1000 m), and badlands. The lithology is mainly represented by Calcisols, Leptosols, Regosols, and Fluvisols.

The combination of these environmental conditions shapes the land uses. Most of the Region of Murcia (11,313 km^2) is highly influenced by human activity and, therefore, it is occupied by cultivated land. Fragmented forest areas represent approximately 45% of the study area. The wooded area is mostly composed of pine forest, *Pinus halepensis* Mill being the main species (ca. 90%) along with *Pinus nigra* Arn *salzmanii* and *Pinus pinaster* Ait, and holm oak (*Quercus ilex* L) to a lesser extent. That is to say, all the main species in this forest community are perennial. The shrubland area covers about of 60% of the forested area in the Region of Murcia.

3.2.2. From Field Plots to ALS Benchmark Map

3.2.2.1. Field Data

ALS plot-level biomass models were calibrated from the field plot database of the Fourth Spanish National Forest Inventory (SNFI) measured in 2010 in the Region of Murcia (MAGRAMA, 2012). The SNFI is an extensive systematic survey carried out every 10 years on Spanish-forested areas on a basis of a regular network of 25 m radius plots at a density of 1 plot per km. From the complete database of the SNFI in Murcia (i.e., 1284 SNFI plots), a training dataset (i.e., 242 SNFI plots) was selected according to a stratified random sampling

done within six predefined strata that covered the entire range of growing stock volume measures.

The main difficulty when using plots from the SNFI in combination with ALS data is the lack of precision in the plot center coordinates. Global navigation satellite system (GNSS) devices used in SNFI data collection, as other studies showed (Mauro et al., 2011; Murgaš et al., 2018), have a nominal accuracy of 5–15 m. Following the methodology and the information sources proposed by Fernández-Landa et al. (2018), the selected plots were relocated manually to improve the accuracy of the plot center coordinates. In this case, plots were shifted an average of 12.5 m in the relocation procedure.

The biomass model was based on four different information sources: (i) 2 m resolution ALS canopy models, (ii) sketches of plot location made by the field teams, (iii) high resolution orthoimagery from the Spanish National Plan for Aerial Orthophotography (PNOA in Spanish), and (iv) total height, species, and location of each measured tree in the field. The plot center relocation was intended to minimize possible errors derived from this lack of precision in the positioning of the plots. The SNFI plot size varies depending on tree diameter at breast height (dbh), that is to say, circular plots with four concentric subplots of a 5, 10, 15, and 25 m radius where trees with a dbh larger than 75, 125, 225, and 425 mm were respectively measured. In the SNFI, percent cover and mean height (in dm) were only estimated for each shrub species in the first plot radius (5 m).

In ALS-based forest inventories, the individual cell size of a grid that covers the entire study area has to be equal to the field plot size (Magnussen and Boudewyn, 1998). In this study, the selected SNFI plots were converted to plots with a maximum radius of 14.1 m (625 m^2), in which we calculated forest AGB and extracted the ALS-derived metrics. Thus, only trees whose distance to the center of the SNFI plot was less than 14.1 m were taken into account to fit a maximum plot area equivalent to a 25 m × 25 m pixel. Therefore, the plots remained with only three concentric subplots of 5, 10, and 14.1 m. We calculated AGB for each selected SNFI plot as the sum of forest AGB related to each shrub species and single tree measured in these three subplots. AGB was scaled up to the hectare level according to the area of each subplot. Tree AGB was calculated through species-specific general models, reporting a goodness-of-fit (adj-R^2) that was always higher than 0.95, as a function of dbh and total height (Montero et al., 2005). To estimate shrub biomass, we applied a general model for Mediterranean shrublands (adj-R^2: 0.63) that considers the shrub cover and its mean height as variables (Montero et al., 2013). Grassland was excluded from the estimation due to ALS data inaccuracy to predict attributes related to this cover type.

Since SNFI data were collected over forested areas exclusively (i.e., forests and shrublands), we applied a forest mask to the study area based on the national forest map. The Spanish forest map (SFM) provides a detailed and updated version for the Region of Murcia at 1:25000 (Murgaš et al., 2018). The SFM is the base cartography for the SNFI and describes the current distribution and seral stage of forest cover, based on structural and ecological descriptive items.

3.2.2.2. ALS Data Acquisition and Processing

ALS data were provided by PNOA. Data from two flight campaigns were used. A recent study demonstrated temporal transferability of pine forest attributes modeling by means of using low-density ALS data, even if data captured with different ALS instruments were considered (Domingo et al., 2019). In our case, data captured in the campaign of 2009 (one year before SNFI plot survey) were used to fit the aboveground biomass ALS model. ALS data captured in 2016–2017 were employed to implement the model and to obtain an updated prediction of forest biomass. Data from 2009 were captured from October to December 2009 using a Leica ALS50 system (Leica Geosystems AG, Heerbrugg, Switzerland) with a mean density of 0.5 pulse m^{-2} and vertical RMSE ≤ 0.20. Data from 2016–2017 were captured from August 2016 to March 2017 using a Leica ALS60 system with a mean density of 0.7 pulse m^{-2} and vertical RMSE ≤ 0.08. In both cases, ALS point cloud data were acquired in 2 × 2 km tiles covering the whole region, which were processed using the FUSION software (Mcgaughey and Carson, 2003).

Firstly, a 2 m resolution DEM was generated to subtract the orthometric elevation of the DEM from the z-coordinate of each ALS return aimed at normalizing the ALS point cloud. This procedure was applied separately to the ALS point cloud from each flight. Then, a normalized ALS point cloud of 2009 was used to compute a total of 15 metrics (Table 3.1) for each field plot using a threshold height of 2 m to separate trees from understory vegetation (Næsset, 2002).

Amongst these considered metrics, tree canopy cover (TCC) refers to the ratio between the number of first returns above 2 m and the total number of first returns, whereas the canopy relief ratio (CRR) describes the relative canopy shape from altimetric observation (Pike and Wilson, 1971). CRR reflects the degree to which canopy surfaces are in the upper (>0.5) or lower (<0.5) portions of the height range (Parker and Russ, 2004). Two other metrics of normalized ALS point cloud for each field plot were computed through filtering from the point cloud those points with a height lower than 2 m and using a height threshold of 0.5 m to separate shrubs from bare soil or herbaceous vegetation. These metrics correspond to mean height of the low stratum (Hmean_LS) and canopy cover of the low stratum (TCC_LS; i.e., the ratio between the number of first returns above 0.5 m and the total number of first returns).

Finally, the normalized ALS point cloud from 2016–2017 was used to compute the same set of metrics for the forest area in the entire region in 25 × 25 m pixels. We used this pixel resolution to reduce the impact on forest biomass scaling-up estimation and because this resolution is compatible with SNFI plot size.

Table 3.1 Summary statistics of the response variable and the airborne laser scanning (ALS) metrics used for modeling in the set of Fourth Spanish National Forest Inventory (SNFI) plots (n = 242; 14.1 m from the plot center).

Group	Variable	Min	Max	Mean	SD
Response	Aboveground forest biomass ($t \cdot ha^{-1}$)	6.88	143.51	57.20	28.03
*ALS metrics	Hmean	2.75	13.34	6.38	1.98
	Hsd	0.55	4.86	1.99	0.73
	Hvar	0.30	23.61	4.49	3.55
	Hcv	0.16	0.50	0.31	0.06
	Hiq	0.76	8.13	2.90	1.15
	Hkur	1.43	4.77	2.51	0.56
	Hp01	2.02	5.00	2.39	0.52
	Hp05	2.04	6.97	3.07	1.00
	Hp10	2.06	8.06	3.66	1.28
	Hp25	2.27	10.94	4.93	1.73
	Hp50	2.48	14.00	6.44	2.17
	Hp75	3.10	17.64	7.83	2.45
	Hp90	3.51	19.24	8.94	2.67
	Hp95	3.77	19.59	9.50	2.78
	Hp99	4.24	20.15	10.25	2.94
	CRR	0.29	0.73	0.48	0.09
	TCC	1.88	93.86	44.55	19.87
	Hmean_LS	0.55	1.77	1.07	0.21
	TCC_LS	0.00	95.49	50.40	20.64

*ALS metrics: H = height of total returns above 2 m, sd = standard deviation, var = variance, cv = coefficient of variation, iq = interquartile range, kur = kurtosis, p = percentiles of total returns above 2 m, CRR = canopy relief ratio, TCC = tree canopy cover, Hmean_LS (low stratum) = mean height of the low stratum (total returns above 0.5 m and below 2 m), and TCC_LS = canopy cover of the low stratum.

3.2.2.3. ALS Benchmark Map: Model, Spatial Prediction, and Resampling

The nonparametric regression algorithm random forest (Breiman, 2001) was used to model the relationship between ALS-derived metrics with the ground-based measure of forest AGB computed from the SNFI plots (14.1 m radius). Analyses were carried out using the packages randomforest (Liaw and Wiener, 2002), VSURF (Genuer et al., 2019), and rfUtilities (Evans and Murphy, 2018) implemented in the R statistical software (R Team, 2013). The VSURF package was applied to identify the combination of ALS metrics that best predicted forest AGB. This package provides an automated method to select predictor variables based on their importance scores while minimizing the redundancy among them. The out-of-bag error statistic provided in RF reports the goodness of model fit but not necessarily the predictive performance. For this reason, an independent data withholding of 10% was performed for a more precise model validation. A bootstrapping validation technique was followed by randomizing the independent data and executing the RF model a thousand times. Withheld

data were predicted at each replicate. This cross-validation procedure was used by applying the random forest regression model cross-validation tool implemented in the rfUtilities package. Cross-validated estimates of root-mean-square error (RMSE) were calculated between the observed and predicted values using the training data. RMSE was normalized by the observations mean (RMSE%). The median percentage of explained variation (pseudo-R^2) and RMSE% were reported for 1000 cross-validations.

The selected model was applied to the ALS metrics computed over the 2016–2017 point cloud covering forest area in 25 × 25 m pixel resolution. Therefore, a continuous 25 m ALS map of forest AGB in tons per hectare ($t \cdot ha^{-1}$) was obtained. The resulting spatial cover was resampled to 250 m moderate resolution (MODIS variables resolution, see next subsection) using the bilinear method implemented in the resample package in R (Hijmans et al., 2018). This ALS-based AGB map at 250 m pixel resolution was used as the forest AGB benchmark of the study area, henceforth ALS benchmark map.

3.2.3. From ALS Benchmark Map to MODIS Map

3.2.3.1. Optical Remote Sensing and Topographic Variables

A stack based on 1060 environmental predictors was generated according to dynamic and static variables. The dynamic variables refer to 8 VIs calculated using the newly release collection 6 of MOD09Q1 (Vermote, 2015) and MOD13Q1 (Didan, 2015) MODIS products. These products respectively consist of 8-day and 16-day composites, both at 250 meters spatial resolution and corresponding to the closest dates to the 2016–2017 ALS data acquisition. Therefore, we obtained data for the years 2015, 2016, and 2017 from the Land Processes Distributed Active Archive Center (LP DAAC) (Golon, 2016). Images were re-projected to the UTM zone 30, datum WGS-84 using the "MODIS Reprojection Tool" (MRT), and quality flags were decoded using the "Land Data Operational Products Evaluation" tool (LDOPE). Both tools were provided by the LP DAAC.

Several vegetation spectral indices were proposed as potential AGB estimators. We used red and near-infrared reflectance bands provided by the MOD09Q1 product to calculate the NDVI, MSR, DVI, RDVI, SAVI, and MSAVI, whereas near-infrared and short infrared reflectance values were used to calculate NDI7 (see equations in Table 3.2). Time series were filtered in order to avoid abnormal values caused by clouds and certain atmospheric conditions. Low-quality pixel values were removed based on MODIS quality data bands and replaced by the average of the previous and subsequent date values with good quality in the time series. A smooth filter was also applied to remove outliers that fall outside the mean plus/minus two times the standard deviation within a five-date period window. The anomalous values were also replaced by the average of the previous and subsequent date values in the time series. In addition, time series from the MOD09Q1 product were smoothed with a Savitzky–Golay filter using a window width of 11 observations.

Reflectance bands were compiled, and indices were computed using the Environment for Visualizing Images (ENVI 4.5) software.

Table 3.2 Equations of vegetation indices derived from moderate resolution imaging spectroradiometer (MODIS) data.

Vegetation Index	Equation [1]	Ref [2]
Normalized difference vegetation index (NDVI)	$\mathrm{NDVI} = \frac{\rho_{\mathrm{NIR}} - \rho_{\mathrm{RED}}}{\rho_{\mathrm{NIR}} + \rho_{\mathrm{RED}}}$	(Tucker, 1979)
Enhanced Vegetation Index (EVI)	$\mathrm{EVI} = \mathrm{G} \cdot \frac{\rho_{\mathrm{NIR}} - \rho_{\mathrm{RED}}}{\rho_{\mathrm{NIR}} + \mathrm{C}_1 \cdot \rho_{\mathrm{RED}} - \mathrm{C}_2 \cdot \rho_{\mathrm{BLUE}} + \mathrm{L}}$	(Hui Qing Liu and Huete, 1995)
Normalized Difference Index 7 (NDI7)	$\mathrm{NDI7} = \frac{\rho_{\mathrm{NIR}} - \rho_{\mathrm{SWIR3}}}{\rho_{\mathrm{NIR}} + \rho_{\mathrm{SWIR3}}}$	(McNairn and Protz, 1993)
Modified Simple Ratio (MSR)	$\mathrm{MSR} = \frac{\left(\rho_{\mathrm{NIR}}/\rho_{\mathrm{RED}}\right) - 1}{\sqrt{\rho_{\mathrm{NIR}}/\rho_{\mathrm{RED}}} + 1}$	(Chen, 1996)
Difference vegetation index (DVI)	$\mathrm{DVI} = \rho_{\mathrm{NIR}} - \rho_{\mathrm{RED}}$	(Jordan, 1969)
Renormalized difference vegetation index (RDVI)	$\mathrm{RDVI} = \frac{\rho_{\mathrm{NIR}} - \rho_{\mathrm{RED}}}{\sqrt{\rho_{\mathrm{NIR}} + \rho_{\mathrm{RED}}}}$	(Roujean and Breon, 1995)
Soil adjusted vegetation index (SAVI)	$\mathrm{SAVI} = \frac{(\rho_{\mathrm{NIR}} - \rho_{\mathrm{RED}}) \cdot (1 + \mathrm{L})}{\rho_{\mathrm{NIR}} + \rho_{\mathrm{RED}} + \mathrm{L}}$	(Huete, 1988)
Modified soil adjusted vegetation index (MSAVI)	$\mathrm{MSAVI} = 0{,}5 \cdot \left(2\rho_{\mathrm{NIR}} + 2 - \sqrt{(2\rho_{\mathrm{NIR}} + 1)^2 - 8(\rho_{\mathrm{NIR}} - \rho_{\mathrm{RED}})}\right)$	(Qi et al., 1994)

[1] ρR, ρNIR and ρSWIR3 respectively correspond to the reflectance in the red (620–670 nm), near-infrared (841–876 nm), and shortwave infrared (2105–2155 nm) wavelengths. In SAVI, the accommodation coefficient of the soil background is L = 0.5, related to moderate canopy cover areas. In EVI, the canopy background adjustment for correcting the nonlinear is L = 1; the coefficients of the aerosol resistance term—which use the blue band to correct for aerosol influences in the red band—are C1 = 6 and C2 = 7.5, and the gain or scaling factor is G = 2.5. [2] References.

The properties of the VIs time-series can be summarized as ecologically relevant measures in a variety of indices (Pettorelli et al., 2005). Different variables derived from MODIS VIs time series were estimated to characterize seasonal dynamics of forest biomass in different time frames. To obtain the annual mean profile of each VIs, the 8-day and 16-day composites were layer-stacked to respectively create a 46- and a 23-band image for each year. We calculated the annual profile as the mean corresponding values of the three years considered (2015–2017).

In order to assess the maximum VIs values related to the growing season, for the 8-day composite of the 2015–2017 time period, we generated aggregates for the computed VIs (Table 3.2) time series using the maximum-value composite (MVC) approach (Holben, 1986). The assumption behind the MVC technique is that the maximum NDVI value of a set of images will correspond to the ideal atmospheric and vegetation conditions. For the remaining VIs, the value corresponding to the date of the maximum NDVI value was selected.

As long-term ecosystem functioning descriptors, we calculated two variables from the 2001–2016 NDVI time series of the MOD13Q1 product: NDVI mean, as annual primary production prediction factor, and its standard deviation (NDVI sd) as the coefficient of seasonal variation, the latter being an indicator of vegetation seasonality (Cash and Moser, 2000)

(Thenkabail et al., 2000). Both variables were calculated according to the method described in (Alcaraz-Segura et al., 2009).

According to (Moore et al., 1991) (Holben, 1986; Moore et al., 1991), a local digital elevation model (DEM) was used to derive the static variables describing the topography to inform models with locally relevant information on landscape position. These static variables were 17 topographic parameters (Table 3.3) derived from a 25 m DEM, which is available at the Centre of Spanish National Geographic Institute (IGN). The current DEM version was generated by the IGN from ALS data interpolation. We used the SAGA GIS software to derivate the parameters, which were included in the terrain analyst functions (Wilson et al., 2012). Previously to the terrain parameters computation, the 25 m DEM was aggregated to 250 m through the same method used for the ALS benchmark map.

All the dynamic and static variables were stacked to build the covariate space which matches the same projection, extent, and pixel size. All the statistical and geo-information analyses in this section were performed using the R software.

Table 3.3. Initial covariates used in the aboveground biomass predictive model from the airborne laser scanning (ALS) benchmark map to the MODIS map.

Covariate		Description (Time Period)	Product
Dynamic[1]	NDVI	8-day composite time series (2015–2017)	MOD09Q1
	MSR DVI RDVI SAVI MSAVI	8-day composite mean annual profile (2015–2017) Maximum-value composite (2015–2017)	
	NDVI EVI NDI7	16-day composite time series (2015–2017) 16-day composite mean annual profile (2015–2017)	MOD13Q1
	NDVI mean	Mean value (2001–2016)	MOD13Q1
	NDVI sd	Standard deviation (2001–2016)	
Static	Topographic	Digital elevation model (DEM), analytical hill shading, slope, aspect, cross-section curvature, longitudinal curvature, convergence index, closed depressions, catchment area, topographic wetness index, Ls factor, channel network base level, vertical distance to channel network, valley depth, relative slope position, multiresolution valley bottom flatness index, multiresolution ridge top flatness index.	Terrain Analyst function in SAGA GIS software

[1] Normalized difference vegetation index (NDVI), enhanced vegetation index (EVI), normalized difference index 7 (NDI7), modified simple ratio (MSR), difference vegetation index (DVI), renormalized difference vegetation index (RDVI), soil adjusted vegetation index (SAVI), modified soil adjusted vegetation index (MSAVI).

3.2.3.2. NDVI Annual Profile per Vegetation Type

Before correlating the ALS-based forest AGB assessment and MODIS-derived indices, we carried out a comparison among vegetation type profiles aimed at characterizing the influence of seasonal dynamics of the understory captured by satellite images on multilayer covers. This was performed mainly because grassland was excluded from ALS biomass estimation (see Subsection 2.2.1). Due to the relationship of the NDVI annual profile with green biomass (Alcaraz-Segura et al., 2009; Paruelo et al., 1997), we generated the annual mean profile of an NDVI 8-day composite per vegetation type in the study area. Pixels of forest areas were aggregated into the main homogeneous vegetation formations at different percentages of canopy cover to describe their seasonal patterns. We classified the main vegetation in three cover types based on the SFM information: dense tree cover, mixed tree and shrub/herbs cover, and dense shrub/herbs cover. Each cover type was divided according to the vegetation species and canopy cover, which resulted in a set of 17 classes. To obtain the annual mean profile, the 8-day MODIS composite was layer-stacked to create a 46-band image for each year. We calculated the annual profile as the mean of the corresponding values of the three years considered (2015–2017).

3.2.3.3. MODIS Map: Spatial Predictive Model and Associated Uncertainty

The pixel values of the ALS benchmark map (250 m resolution) were considered as forest AGB synthetic data representative of the surface covered by the corresponding MODIS pixel. A systematic sample of pixels (39,647 pixels) was selected from the ALS benchmark map, increasing the distribution and sample size of the target variable (forest AGB). Covariate values were then extracted at each location from the covariate stack (Figure 3.1). The most explicative and low correlated predictors were selected through the VSURF package implemented in the R software (Genuer et al., 2015) in combination with a low variance inflation factor (VIF, car package in R). Prior to the calculation of the VIF criterion, a GLM model was generated through the car package algorithm in R (Conrad et al., 2015).

A machine learning algorithm was applied to model the relationship between the target variable (AGB obtained from the ALS benchmark map) and the covariates (selected MODIS derived indices and topographic factors). Due to the performance problems of RF with high dimensional data (Do et al., 2010; Fox, 2015) and the need to conduct pixel-based uncertainty estimation, the extracted pixels of the ALS benchmark map were used to train a quantile regression forest (QRF). QRF is a generalization of RF used to estimate an accurate approximation of the full conditional distribution of the response variable. The QRF infers conditional quantiles to build predicted intervals interpreted as a surrogate of uncertainty associated with the response variable for each pixel value. The quantregForest package (Xu et al., 2012) was applied in the R software to perform predictive and uncertainty maps. Out-of-bag predictions were used to evaluate the quality of the conditional quantile approximations. We computed 2 maps at 250 m resolution: the predicted values of forest AGB based on MODIS-derived index and topographic factors, henceforth MODIS map, and their associated uncertainty. An independent validation of the predicted MODIS map was performed by testing a random sample of 3000 pixels from the ALS benchmark map. The linear correlation between tested and observed data showed the information values of the final MODIS map validation.

In order to analyze the relationship between estimated values of the final MODIS map and their associated uncertainty, a regression line was plotted to visually depict the forest AGB data and their corresponding relative uncertainty at 250 m pixel resolution.

According to (Saarela et al., 2016), ignoring field-to-LiDAR uncertainty can lead to underestimation of estimator variance in a hierarchical modeling process. To address the analysis of the estimation accuracy in the up-scaling process, a generalized hierarchical model-based (ghmb) estimation was therefore performed based on the previous publication (ghmb, HMB package in R). To make the ghmb model computationally affordable, the dimension of the covariate space has been reduced based on the covariate's importance measures depicted in the random forest analysis (randomForest package in R). Moreover, to reduce the number of input data (n=39,647), a representative zone of the study area was selected (Figure 3.1). The results of the ghmb model are the estimated population mean of AGB and its variance as proxies of the estimation accuracy in the upscaling processes. For the same area, the AGB mean based on ALS map and MODIS map was also assessed to compare with the ghmb model results.

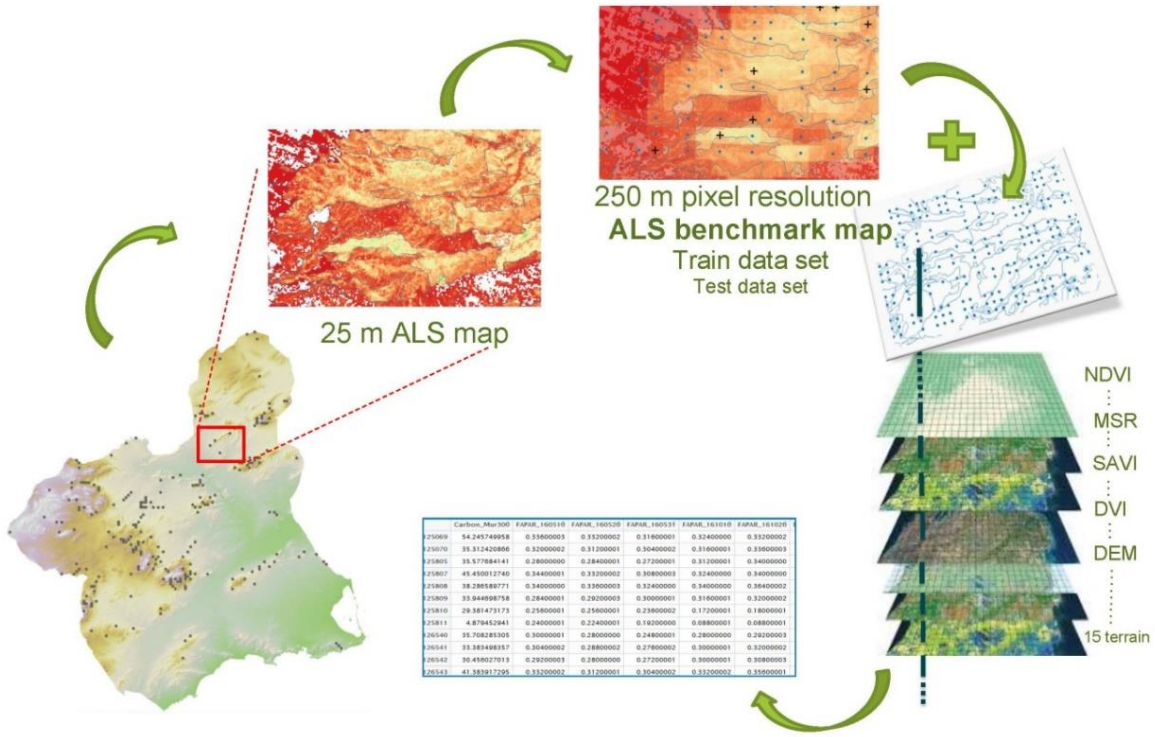


Figure 3.1 Methodological scheme of the target and predictor variables extraction for modeling the final moderate-resolution aboveground biomass.

3.3. RESULTS

3.3.1. ALS-Based Forest AGB Assessment

The VSURF procedure for variable selection has identified as the best combination of ALS-derived variables to predict forest AGB: mean height of total returns above 2 m (Hmean), 25th percentile of total returns above 2 m (Hp25), 50th percentile of total returns above 2 m (Hp50), tree canopy cover (TCC), and canopy cover of the low stratum (TCC_LS). The RF predictive model was calibrated and subsequently used to predict forest AGB at the field plots which were withheld from the model training subset at each cross-validation bootstrap. The RF model fitted with these variables showed a satisfactory performance in terms of variance explained and RMSE% (explained percentage variance: 69.09; median cross-validation RMSE%: 27.18).

The 25 m ALS map (Figure 3.2) showed remarkable spatial heterogeneity with a substantial increase in high elevation areas mainly distributed along the northwestern to southeastern axis of the study area. According to the 25 m ALS map prediction, the maximum values corresponded to the most densely forested areas mainly located in the middle and the northwest parts of the study area. By contrast, the minimum values were associated to shrubland areas. In the 25 m ALS map, the mean AGB value attained in forested areas was 25.44 t·ha^{-1} and 21.7 t·ha^{-1} standard deviation (sd). In tree cover areas (i.e., excluding pure shrubland area), the mean value of AGB was 37.64 t·ha^{-1} (15.64 t·ha^{-1} sd). Since the bilinear interpolation method used to compute the resampling from 25 to 250 m pixel resolution assigns the weighted average of nearest neighboring cells to the output cell value, the biomass mean value decreased at this moderate spatial resolution. Therefore, forest AGB at the 250 m pixel resolution showed a lower mean value and more homogeneous deviation at this coarser spatial resolution: 20.56 t·ha^{-1} (18.57 t·ha^{-1} sd). This resampled ALS biomass map was referred to as the forest ALS benchmark map.

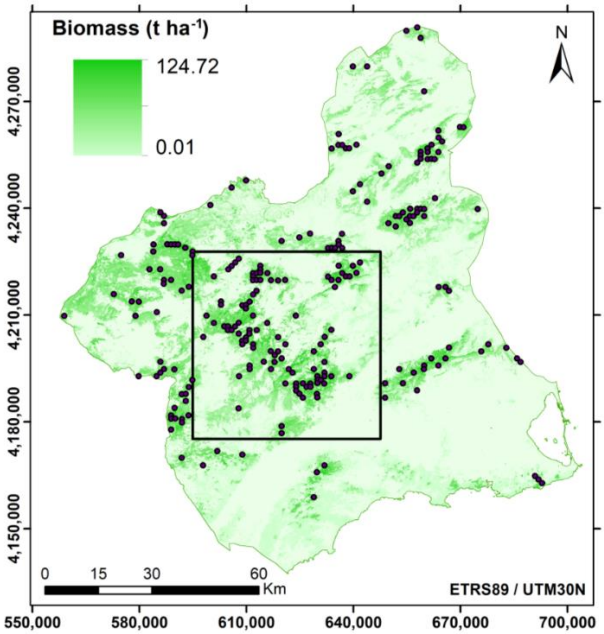


Figure 3.2 25 m airborne laser scanning (ALS) map of aboveground biomass estimation (t ha^{-1}) in the Region of Murcia (Spain), which was calibrated with plot-level ground-based measures (14.1 m radius plots were represented as black dots and were based on the SNFI). The inner rectangle depicts the selected testing area to estimate the global estimator variance through ghmb.

3.3.2. From ALS Benchmark Map to MODIS Map

3.3.2.1. NDVI Annual Mean Profile per Vegetation Type

The signal of forest cover types was extracted from MODIS hypertemporal series of VIs which allowed to analyze valuable information about the vegetation phenology. The differences in seasonal patterns characterized various types of vegetation based on their NDVI annual profile. The comparison of the mean annual profile (8-day composite of the 2015–2017 time series) among the main vegetation formations allowed a discernment of the influence of vegetation different from tree and shrubland depending on the cover density level. The different forest cover types showed a dominant signal signature comparable to the tree canopy cover reflectance pattern (Figure 3.3a) even in combination with other shrub or grass vegetation types (Figure 3.3b,c). Most of the mean seasonal NDVI profiles of the vegetation types were characterized by a bimodal pattern with two maximum values in April–May (0.43 max in pine cover) and December–January (0.47 max in pine cover). The growing seasons started in March/April and September according to spring and autumn in the European Mediterranean basin. The tree-shrubland mixed cover (Figure 3.3b) presented the same profile pattern (i.e., shape) with slightly lower values than the tree cover (Figure 3.3a). The mean annual NDVI profile of pure shrubland of different vegetation formations revealed greater differences regarding the others forest cover types (Figure 3.3c). As stated before, the pure grassland cover was excluded for biomass prediction.

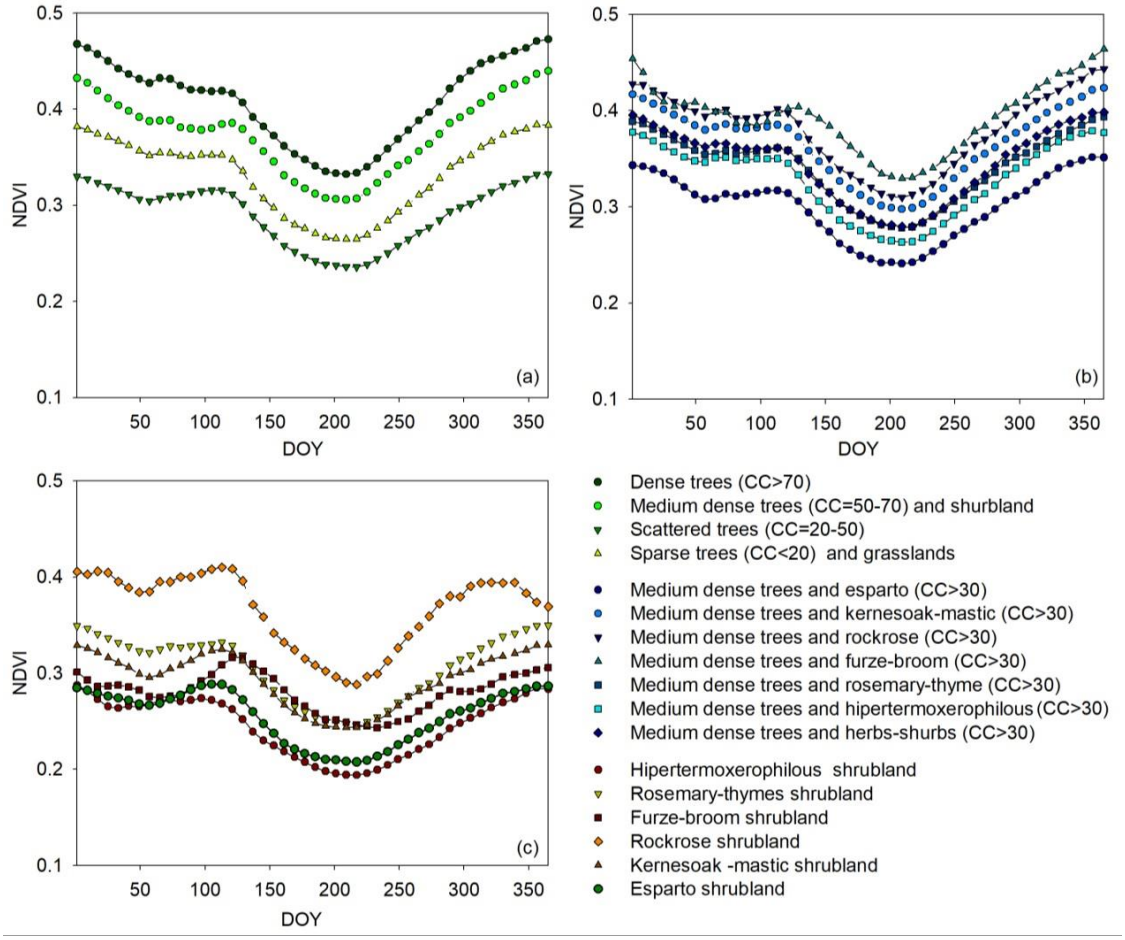


Figure 3.3 NDVI annual mean profile of the 8-day MODIS composite of the 2015–2017 time series per main vegetation types and percentage of canopy cover (CC) based on Spanish forest map (SFM) information. DOY: day of the year.

3.3.2.2. Model Fitting, Spatial Prediction, and Uncertainty Assessment

The previous VSURF selection of the initial set of 1060 covariates aimed to fit the predictive model from ALS benchmark map values decreased the dimension of the covariate space to 35 (Table 3.4), after taking into account a low VIF criterion in order to minimize the statistical redundancy of the predictors. The dynamic predictive factors were associated to dates in April, early May, late August, September, and early October. NDVI was the main vegetation index at different time frames.

Table 3.4 Selected covariates in the fitted aboveground biomass predictive modeling of the airborne laser scanning (ALS) benchmark map.

Description	Selected	Covariates[(1)]
8-day composite time series (2015–2017)	14/04/2016; 5&29/09/2016	NDVI
	15/10/2016	MSR
8-day composite mean annual profile (2015–2017)	9/05; 21/08	NDVI
16-day composite mean annual profile (2015–2017)	9/05; 13/08	EVI, NDW7
Mean value and standard deviation (2001–2016)	2 indices	NDVImean, NDVIsd
Maximum value (2015–2017)	6 indices	NDVI, MSR, DVI, RDVI, SAVI, MSAVI
Topographic variables	17 parameters	DEM, and 16 derived parameters included in Terrain Analyst functions (SAGA GIS software)

(1) Normalized difference vegetation index (NDVI), modified simple ratio (MSR), enhanced vegetation index (EVI), normalized difference index 7 (NDI7), difference vegetation index (DVI), renormalized difference vegetation index (RDVI), soil adjusted vegetation index (SAVI), modified soil adjusted vegetation index (MSAVI).

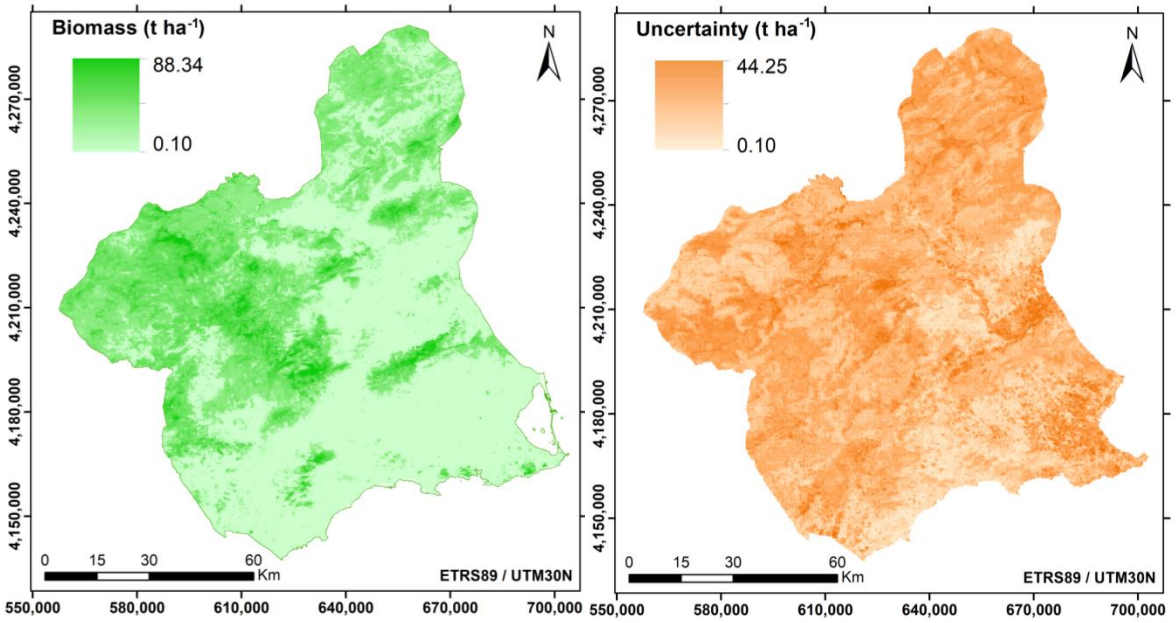
Then, two models were obtained through QFR which corresponded to conditional median for forest AGB (MODIS map) and conditional standard deviation, the latter as a surrogate of uncertainty. Based on the determination coefficient (R^2) and the root-mean-square error (RMSE) as model information criteria, a substantial accuracy increase of the biomass estimates was observed in QRF models compared to GLM techniques previously generated to calculate the VIF. The latter showed an $R^2 = 0.52$ and RMSE = 12.9 t ha^{-1} versus $R^2 = 0.71$ and RMSE = 9.97 t ha^{-1} in the QRF model. The spatial interpolation of the two QRF models resulted into a 250 m spatial resolution map with two bands related to forest AGB estimate (MODIS map) and its associated uncertainty in each pixel (Figure 3.4).

3.3.2.3. Final Predicted MODIS Map and Validation

Figure 3.4 depicts the spatial variability of the forest AGB MODIS map in the study area and its associated uncertainty. The biomass values ranged from 0.1 to 88.34 t ha^{-1} and presented a very similar spatial distribution to the 25 m ALS map (see Figure 3.2). The highest AGB values also corresponded to tree cover in the central and northwest part of the study area, and

the minimum ones were associated to shrubland covers in the southeast. The uncertainty values ranged from 0.1 to 44.25 t ha^{-1}, 11.34 t ha^{-1}, and 4.44 t ha^{-1} as mean and sd values, respectively. The relationship between the forest AGB of MODIS map predicted values and their relative uncertainty associated the highest values of uncertainty with the lowest values (dots within the red box in Figure 3.5) and vice versa.

The resulting MODIS map also showed a predicted mean value very similar to the 25 m ALS map and slightly higher than the ALS benchmark map at 250 m: 24.91 t ha^{-1} (17.16 t ha^{-1} sd). Considering the associated uncertainty of MODIS-based predicted values, the results revealed a mean value of 18.77 t ha^{-1} (13.31 t ha^{-1} sd). The external, independent validation test attained a satisfactory fitting of the predictive model, since it presented a reasonable R^2 and similar RMSE (0.703 and 10.25 t ha^{-1}, respectively).



(**a**) (**b**)

Figure 3.4 Maps of (a) MODIS-based predicted values of aboveground forest biomass (t ha-1) in the Region of Murcia (Spain) and (b) associated uncertainty estimated through the quantile regression forest (QRF) method. Data display was stretched by the cumulative pixel count cut method (default range 2–98%).

According to the AGB estimator for the global two-stage upscaling approach, the 35 selected covariates for AGB MODIS map modeling were reduced to 4 covariates based on the variable importance measures of the QRF model. These covariates were NDVI sd, 16-day NDVI composite of the 04/2016, 16-day EVI composite of the 08/2015 and DEM. The selected area for this analysis (Figure 3.2) covers an extent of 2773 km^2, and includes 72 field plots and a sample of 6474 pixels from the ALS benchmark map.

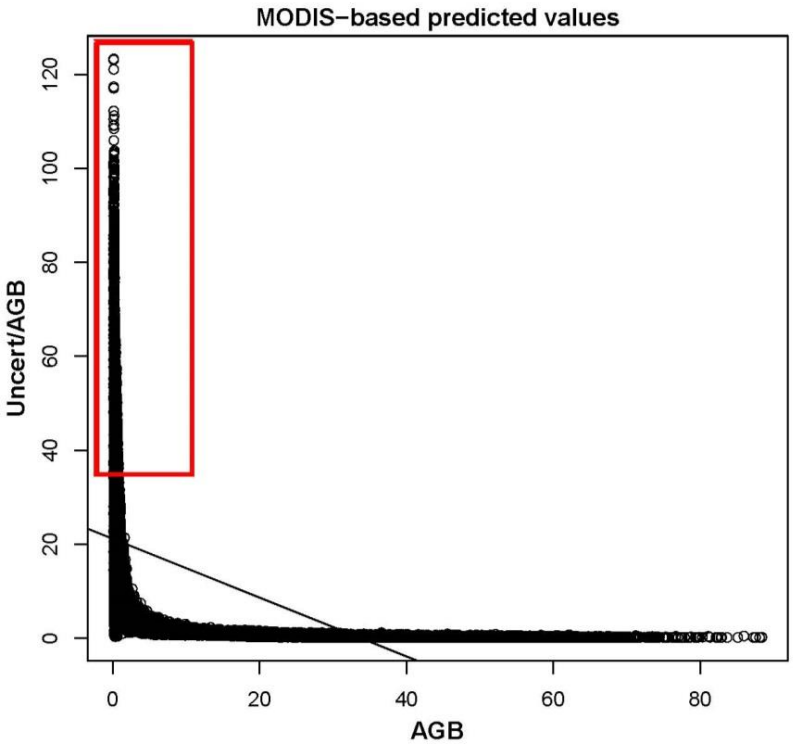


Figure 3.5 Contrast between the predicted values of the MODIS-based aboveground biomass (AGB) and their associated relative uncertainty (uncertainty/AGB).

The ghmb model attained values of 41.63 t ha^{-1} for the estimated AGB mean and 13.22 for the estimated variance with a 95% confidence interval of 34.51–48.76. Likewise, the AGB mean values were 34.72 t ha^{-1} and 31.01 t ha^{-1} for ALS and MODIS maps, respectively.

3.4. DISCUSSION

3.4.1. Forest AGB Estimate from Local to National Scale

The assessment of forest aboveground biomass at the national extent and, subsequently, carbon stock is a future key challenge still unsolved in Spain with large implications related to climate change. The novelty of this study is to apply a two-stage upscaling process for AGB assessment over large extents using open data. Even though there is a time frequency inconsistence in ALS data in Spain, the provided methodology combines a sample of ALS data to bridge the scale gap between a sparse sample of field plots and satellite images.

The methodological framework presented herein provides a robust approach to map detailed, spatially continuous, and easy to update vegetation biomass from local to large scale. We demonstrated that forest attributes such as biomass can be modeled with relevant accuracy in large areas through optical remote sensing data as long as there is free availability of ALS data and NFI plots, such as in the case of Spain. This methodology has increasingly been employed (Fassnacht et al., 2018) to estimate AGB across heterogeneous forested landscapes around the world (Meinshausen, 2017). Nevertheless, some authors such as that of (Hughes et al., 2018) advised that LiDAR-based AGB assessment

should be used with great care for further upscaling to satellite imagery, since they found lower uncertainties in NFI-calibrated biomass maps. The studies that considered the technique used in this study for feasible prediction of AGB refer to a reasonable density of training data to calibrate LiDAR data, short time lag data acquisition between NFI plots and LiDAR data, and a representative plot sample area related to LiDAR pixel resolution (Matasci et al., 2018). Once we addressed these considerations, the discrepancy between ALS biomass estimate and plot-level data was significantly reduced (Table 3.1) and, therefore, the error propagation, too. We obtained a satisfactory performance of the fitted model in terms of prediction of independent test data (explained percentage variance: 69.09; median cross-validation RMSE%: 27.18). As a general comparison, we contrasted our results with the review conducted by the authors of (Urbazaev et al., 2018) over other forestry studies, which focused on estimating growing stock volume/biomass/carbon. All the considered studies in the review used field data, mostly from local inventories of NFIs, and contrasting sources of remotely sensed data. To facilitate the comparison of the results accuracy assessment, the RMSE% (defined as RMSE normalized by the mean observed values) was the measure most commonly used and attained a median RMSE% of 37%, with a standard deviation of ± 31.6% across these types of attributes. The studies that used only LiDAR-based feature variables (81 studies) revealed a median RMSE% of 31.3%, with a median reference set size of 124 plots (our study accounted for 242 plots). These studies confirmed that to obtain low RMSE%, large reference sets are necessary depending on the complexity and diversity of the forest. Therefore, our results of the AGB estimated based on the ALS model can be deemed as a promising result relative to these and national benchmarks (Chirici et al., 2016).

Since the pixel values of the ALS benchmark map can be considered as forest AGB synthetic data, we integrated ALS pixel plots in order to increase the distribution and sample size of local data to model AGB at moderate resolution (250 m). As the authors of (Wulder et al., 2012) claimed, this integration mitigates the cost of ground plot installation and offers spatially extensive and representative sampling of calibration and validation data to support the modeling of forest attributes, such as biomass canopy cover (in our study, more than 30,000 ALS plots and 3000 test data, respectively). However, the costs, limited revisited time and spatial continuity, logistics and data volumes involved in PNOA LiDAR data have fostered the combination of field NFI plot data and synthetic ALS dataset with MODIS imagery. As preliminary results, the extrapolation of AGB assessment from ALS to 250 m pixel resolution based on VIs and topographic variables (MODIS map) offered reasonably robust, accurate, and scalable results (R^2 0.71 and RMSE 9.97 t ha^{-1}), with a satisfactory model fitting for the external, independent validation (similar R^2 and RMSE of 0.703 and 10.25 t ha^{-1}, respectively). Our MODIS model relative RMSE (48.47%) was consistent with other studies such as that of Domingo et al. (2018), which revealed an RMSE% of 69% for aboveground tree biomass estimated by NFI ground plots and 250 m MODIS data. However, other studies, which a large amount of Landsat data, provided lower RMSE% at regional scale (Beaudoin et al., 2014).

The main limitation of measure in our model was due to a discrepancy between the ALS estimation capacity and the surface reflectance signal captured by MODIS in grasslands. ALS presents measurement constraints related to its operational mode with discrete returns. It only returns a few (up to 4) points per laser shot and, therefore, represents a potential

limitation for the measurement of short understory structure (Nguyen et al., 2018). In this sense, canopy attribute measurement was correlated with the MODIS VIs reflectance signal, which also captures variability in understory structure related to composition and configuration of herbaceous species (Hancock et al., 2017). Our analysis of the NDVI annual profile characterizing the main types of vegetation in the Region of Murcia (Figure 3.3) showed slight differences in the shapes of seasonal patterns between stands of pure tree and mixed with grass or shrubland cover Figure 3.3a–c), in which the pattern of the tree canopy cover was mainly predominant. This may be partially due to an understory scattered cover or an intermediate to dense pine canopy cover, where understory reflectance can be neglected.

In addition, since SNFI plots mainly encompass forest tree areas, biomass in pure shrub areas was estimated without plot-level validation. Therefore, and according to studies carried out by the authors of (Li et al., 2017; Ozdemir, 2014), high biomass uncertainties may be found in modeling herbs/shrub areas in the MODIS map generated in this study due to low ground cover or low vegetation height. Even though shrubland biomass estimation was less accurate, the model performance analysis revealed the convenience of including it, thus showing higher final model accuracy. This can be supported by the "edge effect" of the input data aggregation on upscaling (Glenn et al., 2016) created in fragmented forest landscapes such as in the Region of Murcia, where tree canopy cover presented narrower and discontinuous spatial patches in comparison with the enlarged surface of shrubland. This is particularly relevant in forest biomass modeling in Mediterranean areas where fragmentation of forests is highly frequent due to strong human influence and management (He et al., 2002). Moreover, in Mediterranean fire-prone landscapes, there is a current knowledge gap related to shrubland fuel load at the landscape scale and, therefore, lots of assumptions need to be taken in order to assess fire behavior at large scale (Doblas-Miranda et al., 2015). In this sense, despite the acknowledged limitations, the shrubland AGB estimation carried out in this study represents a novel tool with relevant implications regarding fire regime modeling (and subsequently fire management, e.g., prescribed burning).

3.4.2. Environmental Drivers of Forest AGB Spatial Variability

Apart from the influence of dynamic type variables derived by optical remote sensing on AGB modeling, we also found a strong terrain variable influence. These factors derived from the DEM improved our forest AGB spatial estimate, though biomass changes cannot be detected from them, since topography is a static feature. The combination of both types of ecological drivers (static and dynamic) has overcome this limitation and allows us to monitor and easily update the biomass forest environmental changes.

Regarding the dynamic type variables, the previous covariates selection before modeling forest AGB at moderate spatial resolution showed that NDVI, combined with MSR and EVI, exerted a strong influence as environmental drivers among the estimated MODIS VIs. Because of the widely acknowledged close relationship between vegetation characteristics described by satellite data and biomass, remote sensing data are often used to calculate biomass estimates, showing a high correlation mainly with NDVI and describing similar insights as ours (Duane et al., 2016). Moreover, our results revealed the importance of

other covariates regarding different time dynamics which span from short to large time-scale, including seasonal mean dynamics.

The implementation of QRF to a large data set allows capturing complex hidden patterns between covariates and the response variable which are difficult to detect, or even difficult to interpret ecologically. In this sense, the results of our analysis showed quite consistent ecological interpretations. In the VIs time series, we found that the strongest correlation between forest AGB and VIs variables were associated with dates in April, September, and October 2016. This short-time response was related to the data acquisition during the 2016 ALS flight (from August to November), hence highlighting the importance of a short time lag in data acquisition between NFI plots and LiDAR data to minimize the outcome uncertainty (Matasci et al., 2018). In annual mean VIs profiles, we detected strong correlations with May and August dates coinciding with the end of the spring growing season and the start of the autumn growing season, respectively. While some covariates of short-time and annual mean scales were linked to different moments of the high photosynthetic activity in spring (beginning in March–April and reaching a maximum in May), the remaining covariates were related to the summer drought recovery, which is typical in the European Mediterranean region. Finally, long-term ecosystem functioning factors were represented by the mean annual and standard deviation of the NDVI time series as annual primary production and vegetation seasonality indicators (Mura et al., 2018).

In spite of the satisfactory model accuracy reported in this research, the fitted AGB model should be tested including other ground cover specific variables, such as forest canopy cover, and accounting for more species-specific equations to calculate AGB for Mediterranean shrubland. The latter may improve the assessment of forest AGB and its associated uncertainty, relevant in the two-stage upscaling methods due to the error propagation at all scales. In this sense, this can be checked when comparing the AGB mean assessed in each stage with the ghmb model results. The AGB mean of the ALS map for the specified testing area was within the 95% confidence interval of the ghmb model, whilst the MODIS map AGB mean was slightly out of the lower limit. Therefore, as (Saarela et al., 2016) claimed, ignoring the uncertainty in the AGB benchmark map may lead to under-estimation of estimator variance in a hierarchical process. However, both methodologies—ghmb and random forest—are not totally comparable since they are based on different modeling approaches (linear regression vs. decision trees, respectively) and the number of covariates also differs. Despite this, an independent, external ground-based validation of the final MODIS map would be desirable, as it would allow the analysis of the accuracy and reduce the high uncertainty estimates which limit the understanding of the spatial variability of AGB.

3.5. CONCLUSIONS

The methodology proposed here has been shown to provide a cost-effective and easy to update model aimed at generating spatially extensive maps at management scales concordant with AGB complexity. This confirms our hypothesis regarding the correlation between ALS-based forest AGB and MODIS–derived VIs. In this study, we specifically leveraged the benefits of two complementary remote sensing technologies to map forest AGB over large areas and integrated ALS pixel plots as a means to increase the distribution

and sample size of target data for modeling AGB as a function of VIs moderate resolution composites (MODIS).

The information necessary to perform the methodology tested in this study is available for the entire Spanish national territory and free to access. Even though there is a date inconsistency in ALS data in Spain with several years of time lag between two different areas owing to the time frequency of ALS flights each 6–7 years, enough networks of airborne ALS transect can be used to bridge the scale gap between field samples and satellite images. Due to the high temporal availability and moderate spatial resolution, MODIS imagery was shown to be feasible in forest AGB estimation for regional, even national, land use planning and management. The two-stage upscaling method for AGB assessment tested in this study can be applied over large areas, such as the whole country, to gain insights into national forest aboveground biomass, and subsequently into carbon stock, which remains a key challenge still unsolved in Spain.

Capítulo 4. Testing Modeling Strategies for Soil Organic Carbon at Different Spatial Scales in a Mediterranean Area (Southern Iberian Peninsula)

ABSTRACT

Quantification and monitoring of soil organic carbon (SOC) stocks across local-to-global scales is needed to assess soil management practices, adapt policies, and evaluate environmental impacts. Here we evaluated different digital soil mapping (DSM) techniques to develop a local SOC map and test it against available estimates derived from regional-to-global databases. The study was focused within the Region of Murcia, a topographic and climatic complex area in southern Iberian Peninsula. Furthermore, we analyzed SOC estimates derived from a local database to predict SOC concentration (SOCc) and SOC stocks (SOCs) using different spatial interpolation methods (linear model, quantile regression forest; random forest and support vector machine) at three spatial resolutions (100m, 250m, 1000m). A Quantile Regression Forest (QRF) approach parameterized with local data of SOCc had the best data-model agreement at 100 m spatial resolution and the best balance for accuracy, external validation, and interpretability of results. This predictive model showed 12.18 g/kg with an overall uncertainty of 10.54 g/kg an accuracy percentage of 79% for SOCc, and a 27,572 GgC/ha and an uncertainty of 0.016 GgC/h for SOCs. Using local environmental covariates and local legacy soil information to predict SOC within the region resulted in a relative improvement between ~40% (for SOCc) and ~65% (for SOCs) when compared with SOC products derived from national and global databases. The results showed the large discrepancy between national and global estimates for reporting SOC at a local scale. Consequently, local-to-regional efforts are needed to better describe SOC spatial variability to reduce uncertainty.

Keywords: Digital soil mapping, SOC concentration, SOC stock, Quantile Regression Forest (QRF)

4.1. INTRODUCTION

Global environmental change disrupts the biodiversity, structure, and function of terrestrial ecosystems (Pecl et al., 2017). Sustainable land use management and, specifically, soil carbon management is crucial for global change adaptation and climate regulation (Jobbágy et al., 2000; Wiesmeier et al., 2019). Quantification and monitoring of soil organic carbon (SOC) stocks across scales is needed to inform soil management, adapt local policies and assess potential impacts (Vargas-Rojas et al., 2019). However, there is a need to bridge the gap between management and the suitable spatial scale (Cash and Moser, 2000).

A current research challenge is to predict accurately SOC stocks at high spatial-resolution across the whole soil profile. However, due to high spatial variability of SOC, soil surveys are still a challenging and expensive task with operational complexity, and difficult to update (Smith et al., 2020; Vargas et al., 2017). To upscale soil surveys, soil mapping has traditionally included a framework considering soil forming processes assessed from soil-landscape and vegetation associations. Arguably, new technological modeling and computational advances, such as digital soil mapping (DSM), challenge this traditional approach, enabling large-scale implementation while providing local information on soil properties (Brus et al., 2011; McBratney et al., 2003; Savin et al., 2019; Searle et al., 2021). New developments in DSM combine data-driven models relying on direct measurements (i.e., plot scale) provide local (Filippi et al., 2021), national (Vitharana et al., 2019), continental (Guevara et al., 2018) or global (Hengl et al., 2017) estimates of different soil properties. Moreover, the integration of single and/or multitemporal remotely sensed satellite indices improved the prediction of soil carbon spatial distribution even when there exist limited soil samples (Fathololoumi et al., 2020; Liang et al., 2020; Schillaci et al., 2017a).

Even though there has been considerable progress in the accuracy of DSM based products, there are aspects in constant development and improvement. In particular, as described in more detail below, database information, carbon parameter estimation, spatial predictive models, and spatial scale (Arrouays et al., 2017; Poeplau et al., 2017).

Due to the aforementioned limitations of soil data surveys, an external validation of SOC maps is often missing, with data being used almost entirely in parameterization (Lamichhane et al., 2019). Despite regional-to-global efforts, such as Profile Analytical Database for Europe (SPADE), Harmonized World Soil Database (HWSD), World Soil Information Service (WoSIS) or the International Soil Carbon Network, differences in data density or in structured information, e.g. “map units” or “point” information, contribute to biases (Kibblewhite et al., 2008; Smith et al., 2020; Trnka et al., 2011; Willaarts et al., 2016). Furthermore, an unsolved issue across these databases is the lack of concomitant soil properties. This is critical because estimation of SOC stocks (SOCs) is dependent on information of SOC concentration (SOCc), bulk density (BD), and coarse fragment content (CRF) of the target soil depth. While SOCc is usually measured with precision in elemental analyzers, BD and CRF are often missing, which results in added uncertainty in the predictions (Durante et al., 2020; Poeplau et al., 2017).

Soil predictive methods are constantly improving to address the challenge of limited in situ data. For example, soil spatial inference is common to generate continuous maps from point data and to estimate SOCs (Wang et al., 2018b). For instance, in the SCORPAN spatial inference approach SOC can be predicted as a function of the soil forming environmental factors, such as climate, topography, vegetation, or land use (McBratney et al., 2003). To correlate SOC with this forming factors numerous strategies in statistical prediction models have been developed (Kravchenko and Bullock, 1999; Omran, 2012; Robinson and Metternicht, 2006). Linear regression approaches are popular because of its computational simplicity and interpretability (Thompson et al., 2006), but the relationships between soil properties and environmental variables are usually complex and nonlinear (Manning et al., 2015; Moni et al., 2010; Wiesmeier et al., 2019) . Recent studies have proposed alternative techniques adapted from data mining, machine learning, or multi-model ensemble methods to account for these non-linear relationships and improve DSM predictive capacity (Shangguan et al., 2017; Wang et al., 2018b).

To generate robust models, the spatial resolution of SOC variable must be consistent with spatial scales of both input covariates and land management (Hartemink, 2006). Arguably, to ensure this consistency for national and local management strategies, scales from 1: 50,000 to 1: 500,000 are recommended (Montanarella, 2015; Pásztor et al., 2019) The corresponding pixel size from these scales ranges from 25 m to 250 m, respectively (Tobler, 1988). The resolution of national SOC maps provided by global soil projects derived from HWSD is about 1 km, such as the Global SOC Map (GSOC) of the Global Soil Partnership (Yigini et al., 2018). A more detailed global initiative derived by the WoSIS-standardized data (i.e., Soil-Grids250m) has been generated (Hengl et al., 2017). Finally, a third global project is the GlobalSoilMap, a consortium conducted by the International Union of Soil Sciences (IUSS) (Arrouays et al., 2014) with the goal to create global digital maps for key soil properties prediction at finer spatial resolution (about 100 m). The diversity of available SOC maps requires comparative approaches and locally-derived benchmark information for applicability and interpretation of global maps.

Our overarching goal is to evaluate different methodological criteria relying on DSM techniques and remote sensing data for local SOC mapping to improve currently available estimates derived from regional-to-global databases. We focus on a local Mediterranean area (11,313 km2) with complex climatology and topography in the southeastern Iberian Peninsula. At least six different SOC maps products derived from global to local approaches are available in this area. Due to the spatial heterogeneity of SOC in the study area (Conant et al., 2011; Minasny et al., 2017; Xiong et al., 2016), we hypothesized that data-driven models fed by a local set of observations and with appropriate environmental data could better capture SOC spatial variation than extrapolated global estimates.

Our specific objectives were to (1) compare SOC from six available products for the study area and validate them with an independent dataset; and (2) produce a local SOC map testing different data, statistical models and spatial scales (100 m, 250 m and 1 km).

4.2. MATERIALS AND METHODS

4.2.1. Study Area

The study area is focused on a Mediterranean semiarid area of the southeastern Iberian Peninsula (i.e., Region of Murcia). This region is about 11,313 km^2 and presents a complex topography including mountains (reaching 2,000 m altitude), high plateaus (500-1000m), and advanced degradation zones or badlands (> 14% of the territory). This topographic diversity results in contrasting climatic zones. For example, the southeastern area is influenced by hot dry winds of Sahara Desert, which causes a NW to SE line of aridity. Overall, the study area has a mean annual temperature of 18º C, annual rainfall of 300-350 mm/year distributed in torrential events, and mean annual evapotranspiration of about 900 mm.

Most of the Region of Murcia (70%) is influenced by human activity occupied by cultivated areas; approximately 20% is covered by shrubland and 10% by pine forest. The lithology is represented by underdeveloped soils with scarce and hardly differentiable horizons. According to the World Referenced Based (WRB, 2014), the dominant soils are: Calcisols (43%), Leptosols (23%); Regosols (17%) and Fluvisols (9%) followed by Gipsisols, Solonchaks and Kastanozems (Alias and Ortiz, 1986).

4.2.2. Available SOC Products Across the Study Area

We considered six available SOC products derived from DSM frameworks as benchmarks, four of which correspond to Global and European, one national map, and the last is a local SOC content map (Table 4.1). These products provide two different organic carbon variables: soil organic carbon concentration (SOCc) and soil organic carbon stock (SOCs). The description of the products is reported as follows.

- At Global and European level:
    - **GOSC:** Global Soil Organic Carbon (Hiederer and Köchy, 2012). The organic carbon density for the topsoil (0 - 30cm) and the subsoil layer (30 - 100cm) were obtained from the amended HWSD with estimates derived from other global data sets for these depth layers. SOC distribution was calculated through generalized linear models, where a raster format is used for the spatial extent of the soil mapping units.
    - **SG**: SoilGrids250m system of WoSIS (Hengl et al., 2017). The digital soil mapping techniques were based on a global compilation of soil profile data (WoSIS) and environmental layers. To regionalize the SOC, global data in the SG product used machine learning techniques (random forest-kriging) implemented in the GSTAT package in R.
    - **OCTOP**: Organic Carbon Content in Topsoils in Europe (Jones et al., 2004). The OCTOP estimates were calculated from European Soil Database by combining refined pedo-transfer rules and spatially continuous data layers to generate a digital soil mapping by regression kriging approach. As a first attempt to calculate SOC contents at the European level, the authors stressed the lack of comprehensive geo-referenced, harmonized SOC data to test the reliability of their map, in sampling and analysis methodologies.

- **ocCont-LUCAS**: this Top Soil Organic Carbon map in EU-25 is based on Land Use/Cover Area frame statistical Survey 2009 (LUCAS) (de Brogniez et al., 2015). The map was produced by fitting a generalized additive model between organic carbon measurements from the LUCAS survey (dependent variable) and selected environmental covariates. Urban areas, large water bodies, and areas above 1000 m altitude were masked out. Due to improved dataset, it offers the opportunity to replace the OCTOP dataset (Sanchez et al., 2009).

– At national and local level:

- **SCSS**: Soil Organic Carbon Stock in Spain (Rodríguez Martín et al., 2016). It assessed SOCs analyzing topsoil samples from 4401 locations. SOCc, soil bulk density and stoniness were measured to estimate SOCs, using geostatistical technique (ordinary kriging) for spatial interpolation.
- **OCMRM:** Soil Organic Carbon Map in Region of Murcia (Blanco, 2015). The SOCc (g/kg) was modeled in the Region of Murcia using machine learning techniques (random forest and support vector machine) and bootstrapping methods for validation. More than 1000 topsoil samples were used for the analysis from LUCDEME (Fight against desertification in the Mediterranean, by its initials in Spanish) database.

Table 4.1 Description of available SOC products for the study area.

Name	Product	Publisher	Publication date	Soil Database	Spatial resolution	Data	Units
GLOBAL PRODUCTS							
GSOC	Global soil organic carbon	JRC	2012	HWSDB	1 km	2 horizons: 0-30 cm 30-100m	tC/ha
SG	SoilGrids250m: 7 horizons and Individuals soil layers	ISRIC	2017	WoSIS	250 m	7 horizons: 0-200 cm Top soil: 0-30 cm	g/kg tC/ha
EUROPEAN PRODUCTS							
OCTOP	Organic Carbon Content in Topsoils	JRC	2004	ESDB	1 km	Top soil: 0-30 cm	%
ocCont (LUCAS)	Topsoil Soil Organic Carbon Content	JRC	2014	LUCAS (Land Use/Cover)	500 m	Top soil: 0-20 cm	g/kg
NATIONAL PRODUCTS (Spain)							
SCSS	Soil Organic Carbon and soil organic carbon stock in Spain	INIA	2015	Spanish topsoil database	100m	Top soil (0_30 cm)	% tC/ha
LOCAL PRODUCTS (Murcia)							
OCMRM	Soil Organic Carbon Map in Region of Murcia	University of Murcia	2014	LUCDEME	25 m	top soil: 0-30 cm	g/kg

4.2.2.1. Comparison of Available SOC Products

To generate comparable information among DSM products, we extracted SOCc and/or SOCs at 30 cm depth; except for ocCont product with data only in the uppermost 20 cm. For SOCc estimates in SG, we weighted the average of the predictions within the depth interval (i.e., 0-30 cm) using the trapezoidal rule for numerical integration (Hengl et al., 2017). For OCTOP, the SOCs values were supplied from a table format per country.

As a preliminary approach, we performed a qualitative description and a quantitative comparison of the SOC estimates of the six available products at 30 cm and two different spatial domains: national (Spain) and local (Murcia). We computed descriptive statistical parameters of SOCs and SOCc for comparisons: sum (in GgC) for SOCs and mean (in g/kg) for SOCc, and standard deviation for both. Moreover, to analyze the statistical significance of the mean values of the SOC products, we performed the Tukey plot interval using the 'graphics' package in R.

We validated the products estimates with an independent local database of 255 soil profiles (see section 4.2.3) using general statistical comparisons (using the 'graphics' and 'stats' packages in R) and the determination criteria (R^2) and root mean squared error (RMSE) as information criteria.

4.2.3. Local Spatial Prediction of SOC and its Uncertainty

4.2.3.1. Local Soil Database and SOC Calculation

The independent local database was derived from the LUCDEME Project generated between years 1986-2004 by the "Ministerio de Medio Ambiente de España" and the support of "Dirección General de Medio Ambiente de la Región de Murcia" (Alias and Ortiz, 1986). This legacy database consists of a sampling of 255 soil profiles representative of soil taxonomic units over a range of 0-1700 m altitude. For each profile, there are morphological and analytical data for each horizon (903 horizons). In addition, 1100 topsoils (20-30 cm soil depth) were sampled on a regular 3 x 3 km grid, except the southeast quadrant, distributed in an altitudinal range of 0-1950 m (Figure 4.1).

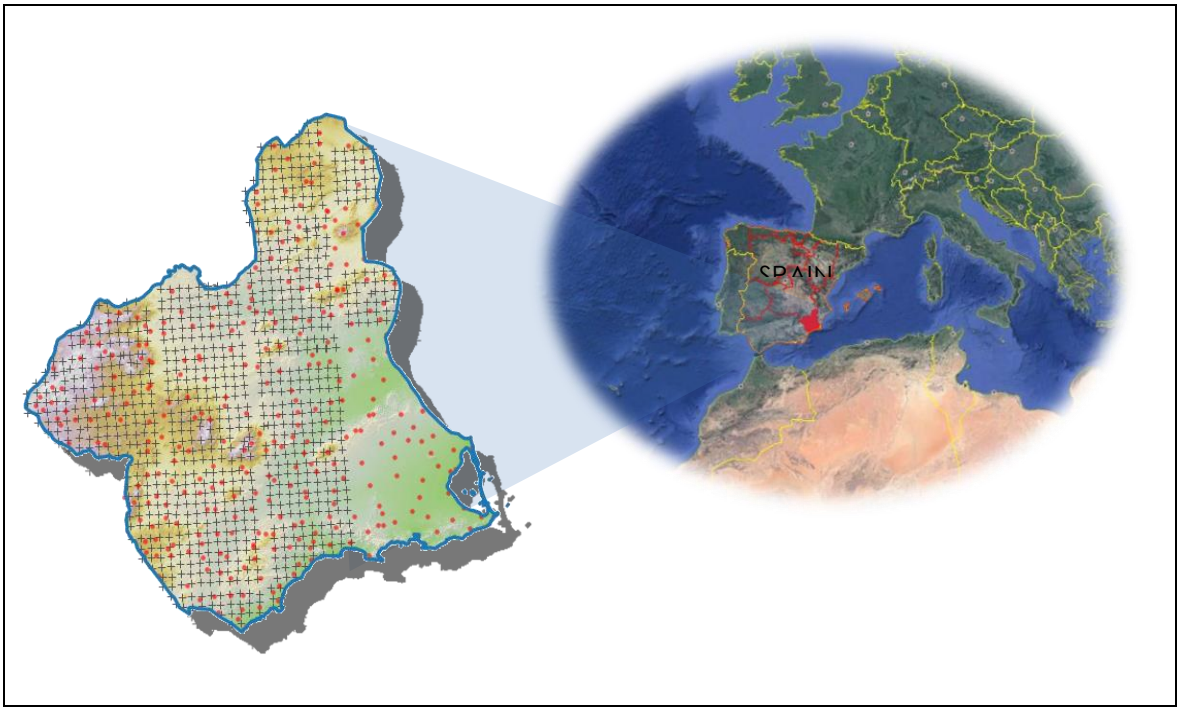


Figure 4.1. Spatial distribution of database samples derived from LUCDEME Project in the Region of Murcia (Spain). Dots depicted 255 soil profiles and cross symbol 1100 topsoils (20-30 cm depth approx.).

We harmonized and screened the soil legacy database following published guidelines (Dobos et al., 2010). To validate the aforementioned benchmark products, we generated synthetic horizons of 0–30 cm depth from soil profile of local legacy database for SOCc and SOCs data. To generate the synthetic horizons, the aggregation of horizons depth was carried out using the equal-area spline technique through mass preserving spline ('mpspline') function. This technique is based on fitting continuous depth functions for modeling the variability of carbon soil (Bishop et al., 1999). The estimation of SOCs in each soil profile were calculated from equation (1):

$$SOCs\ (Kg \cdot m^2) = SOC\ (g/Kg) \cdot BD(Kg \cdot m^3) \cdot \left[1 - \left(\frac{CRFVOL}{100}\right)\right] \cdot HSIZE(cm) \quad (1)$$

were BD is bulk density, CRFVOL is percentage of coarse fragments (above 2 mm in diameter), and HSIZE is thickness of the horizons. Due to data gaps, BD was estimated by means of a pedotransfer function adapted from a regional study (Barahona and Santos, 1981). We used the R package GSTAT for the stock estimates and mpspline, where the propagated error (attached as an attribute) was estimated by the Taylor Series Method (Hengl and Mendes de Jesus, 2016; Heuvelink, 1998; Malone et al., 2009).

In order to evaluate the SOC spatial distribution, we log-transformed the SOC original values to generate a normal distribution (Yigini et al., 2018). We tested the correlation between log SOC values with the prediction factors and compared to SOC original ones. Moreover, we provided the basic statistical data of the semivariograms of SOCc and SOCs values at 30 cm depth, as well as the SOC logarithm.

4.2.3.2. Local Spatial Covariates

Our modeling approach was based on the SCORPAN conceptual model using the soil forming environmental factors as soil spatial prediction functions (McBratney et al., 2003). We generated a covariate stack based on 34 environmental factors as SOC prediction variables (Table 4.2).

Table 4.2. Description of prediction factors used in statistical modeling of soil organic carbon (SOC)

Variables[(1)]	Source	Spatial resolution	Description
DEM	IGN (Spain)	25 m	Terrain altitude variability, basis of topographic variables.
PP	ACDPI University of Barcelona	200 m	Mean annual precipitation (mm), period 1951-1999.
TP	ACDPI University of Barcelona	200 m	Mean, minimum and maximum annual temperature (°C), period 1951-1999.
NDVI and CV_NDVI	MODIS-Terra (MOD13Q1)	230 m	Mean annual and coefficient of variation of Normalized Difference Vegetation Index (NDVI), period 2001-2016.
EVI and CV_EVI	MODIS-Terra (MOD13Q1)	230 m	Mean annual and coefficient of variation of Enhanced Vegetation Index (EVI), period 2001–2016.
Lithology	IGME (Spain)	1:200000	Lithological units and their associations.
Soil types	SEIS.net Project (MIMAM-CSIC)	1:100000	Digitized soil map from the National Atlas of Spain (1: 2,000,000), IGN 1992).
Land cover	IGN-Corine Land Cover (Spain)	1:100000	Inventory of land covers.
MFE	MAPAMA (Spain)	1:25000	Forest structural types and cover canopy area.
LiDAR (Tree biomass)	IGN (Spain)	5 m (0.5 pt/m^2)	High-precision of vegetation cover altitude from Light detection and ranging (LiDAR) data.

(1) DEM: Digital elevation model. PP: Precipitation. TP: Temperature. NDVI: Normalized difference vegetation index. CV_NDVI: Coefficient of variation of NDVI. EVI: Enhanced vegetation index. CV_EVI: Coefficient of variation of EVI. MFE: Map of forests in Spain. LiDAR: Light detection and ranging.

We used dynamic and static variables as predictors for SOC. The static variables were 16 topographic parameters derived from a local digital elevation model (DEM) using the Terrain Analyst functions included in SAGA GIS software (Conrad et al., 2015). DEM was available from Geographic Information National Centre (Spain), resulting from interpolation of LiDAR national images, with a 25 m pixel resolution. To perform the local model at different spatial resolution we re-sampled the DEM into 100, 250, and 1000 m pixel size and calculated the basic terrain parameters at each scale.

The dynamic variables included climatic variables (precipitation and temperature) (Ninyerola et al., 2005); land cover (IGN, 2012) reclassified into 13 classes; forest structural variables, aboveground biomass of forest trees cover from LiDAR data (Durante et al., 2019); and vegetation indexes (VIs). The calculated VIs were The Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI), which are link to ecosystem functional attributes related to seasonal dynamics of net primary productivity. These indices were derived from mean annual time series images (2001-2016) of MODIS-Terra images satellite using Google Earth Engine as describe in (Arenas-Castro et al., 2019).

These covariates were layer-stacked to build three different harmonized covariate stacks (at 100 m, 250, m and 1000 m pixel resolution) with the same projection, extent, and pixel size). The covariates were re-scale, re-projected or, in the case of categorical covariates, rasterized when appropriate. All the statistical and geo-information analyses in this section were performed using packages of R software (raster, rgeos, rgdal, GISTools).

4.2.3.3. Model Fitting, Spatial Prediction, and Uncertainty Assessment

To analyze the influence of different methodological criteria in estimation of SOC spatial variability, we tested different statistical models replicated at three different pixel resolutions: 100 m, 250 m, and 1000 m.

Prior to model building, a regression matrix was performed including the best correlated environmental factors with SOC local data as covariates. To select them, we used a balance among higher Pearson coefficient of multiple linear regression, lower error (RMSE), and lower variance inflation factor (VIF) to identify statistical redundancy (Heiberger et al., 2005). We used the Akaike information criterion (AIC) to determine the best compromise between model accuracy and model parsimony (Rossel and Behrens, 2010).

We tested the following models to predict SOCc and SOCs:

- Linear models (**LM**): the fitted model was specified using SOC as response variable and the regression matrix of covariates as predictors. We used the 'stat' R package to perform LM.
- Support vector machine (**SVM**): This algorithm creates a line or a hyperplane, which separates the data into classes. Prior to performing this model, the qualitative variables were transformed into factors. We used the SVM with the linear kernel method (svmLinear) Kernel since a non-linear decision surface can be transformed to a linear equation in a higher number of dimension spaces. This method was implemented in the train function of the 'caret' R package in R software.
- Random forest (**RF**): is an enable learning method (bagging) of decision trees. Decision trees learn how to best split the dataset into smaller subsets based in different conditions (or nodes) to predict the target value. The RF algorithm operates by constructing a multitude of decision trees at training time and outputting the mean of prediction of the individual trees. The number of variables available for splitting at each tree node (mtry) was set 1/3 of the total variables used in the model, and the total number of trees to grow (ntree) was 500. We implemented this method in 'randomForest' package in R software.

Quantile regression forest (**qrf**): since qrf estimates an approximation of the full conditional distribution of the response variable, the inferred conditional quantiles to build prediction intervals were estimated as surrogate of the value of uncertainty associated with the response variable (Meinshausen, 2006). We used qrf algorithm implemented in R software environment for statistical computing in two different packages: 'quantregForest' (**QRF**) (Meinshausen, 2006) and' 'GSIF' (**QRF_G**) (Hengl and MacMillan, 2019). QRF validation was calculated from out-of-bag error, and QRF_G model validation was calculated from n-fold cross-validation. In addition, the latter combines predictions by qrf regression and interpolation of residuals (kriging) via the Regression-Kriging (RK) techniques.

The information criteria to assess the fit of the different models were RMSE and R^2 (equations (2) and (3)).

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

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

The observed values of LUCDEME topsoil database were graphically compared with the estimate of available SOC products of the study area using Taylor diagrams (Carslaw and Ropkins, 2012). In these diagrams, the similarity between two patterns is quantified in terms of their correlation, their centered root-mean-square difference, and the amplitude of their variations (represented by their standard deviations).

4.2.3.4. Validation and Local Map Selection

An independent, external database was used for a robust model validation. The validation data were based on the 1100 topsoil legacy local dataset describe in 4.2.3.1section (Figure 4.1). The agreement between predicted and observed data was measured by the accuracy percentage within the interval of SOC prediction (i.e., the interval corresponding to predicted values and their associated uncertainty). The balance between the predictive model performance and the validation determined the model selection for the final maps of the SOCc and SOCs.

In order to analyze the relationship between estimated SOC and their associated uncertainty in the selected maps, we plotted them in a scatter diagram depicting the regression line to facilitate data visualization.

We calculated the relative improvement (RI, equation (4) of prediction accuracy of generated SOCc and SOCs maps rely on OCMRM and SG stock maps (i.e., the maps with the most accurate balance of model validation criteria in the external validation).

$$RI = \frac{RMSE_{AP} - RMSE_{QRF}}{RMSE_{AP}} \quad (4)$$

where $RMSE_{AP}$ and $RMSE_{QRF}$ are the root mean square errors of a given available product and our maps (SOCc and SOCs), respectively.

4.3. RESULTS

4.3.1. Available SOC Products, Comparison and Validation

For the comparison of the maps, we considered their original data characteristics, predictive models, and spatial scale. Regarding data availability, we compared the estimated values of both SOCc and SOCs at national (Spain) and local (Murcia) level, when applicable. We clarify that the OCTOP just had SOCs values at national level and the ocCont (LUCAS) masked out urban areas, large water bodies, and areas above 1000 m altitude.

The quantitative comparison of all products showed lower SOCs values in the Region of Murcia corresponding to 1.4-2.1 % of the national data (Table 4.3). The qualitative comparison of their spatial patterns depicted some agreement in the distribution of higher SOCc values across the northeast of the region (Figure 4.2). The differences between the spatial patterns of the SOCs maps were more notable than the SOCc maps (Figure 4.3).

Table 4.3. Comparison of available soil organic carbon concentration (SOCc) and soil organic carbon stock (SOCs) products for the topsoil) at national and local scale (Región de Murcia).

AVAILABLE PRODUCTS [(1)]			**GSOC**	**SG**	**OTCOP**	**ocCont (LUCAS)**	**SCSS**	**OCMRM**
Spatial domain			(Global)	(Global)	(Europe)	(Europe)	(Spain)	(Murcia)
National (Spain)	SOCs (GgC)	Sum	$3.19 \cdot 10^6$	$5.07 \cdot 10^6$	$3.50 \cdot 10^6$	$1.89 \cdot 10^6$	$2.82 \cdot 10^6$	-
		Mean±SD	0.06±0.02	0.10±0.03	0.07	0.04±0.03	0.06±0.04	-
	SOCc (g/kg)	Mean	-	26.62	22.71	24.38	25.89	-
		SD	-	17.94	23.69	16.62	20.47	-
Local (Murcia)	SOCs (GgC)	Sum	52,689.62	73,364.37	-	39,983.88	44,421.16	
		Mean±SD	0.05±0.01	0.07±0.01	-	0.04±0.02	0.04±0.01	
	SOCc (g/kg)	Mean	-	13.78	11.6	18.88	19.51	14.07
		SD	-	5.87	7.15	7.85	6.06	14.88

(1) The available SOC products in the area refer to Global soil organic carbon (GSOC), SoilGrids250m (SG), Organic Carbon Content in Topsoils (OTCOP), Topsoil Soil Organic Carbon Content (ocCont), Soil Organic Carbon Stock in Spain (SCSS), Soil Organic Carbon Map in Region of Murcia (OCMRM).

At the national scale, the different available DSM products showed a large diversity among stock values (coefficient of variation CV = 0.35), with a range from 1,892 Tg C (by LUCAS map) to 5,068 Tg C (by SG map), representing a difference of 63%. At the local scale, the

results showed slightly smaller differences between extreme values (CV =0.28), ranging from 39,984 Gg C to 73,364 Gg C (46% of difference), according to LUCAS and SG maps, respectively. We found the opposite pattern in SOCc, where the values remained much more stable at national than at local scale (CV= 0.07 and CV= 0.22 respectively). The OCTOP map showed the lowest concentration value (22.71 g/kg at national scale and 11.6 g/kg at local scale) versus the highest values of SG map (26.62 g/kg) at national scale (14%), and SCSS map (19.51 g/kg) local scale (40%). Therefore, despite the higher data dispersion at local scale, the CV values are smaller in concentration than in stock data.

The standard deviations (SD) also presented large spatial heterogeneity, being much lower at local spatial domain than at national level for both variables. The high standard deviation value of the local product OCMRM (SD= 14.88) is remarkable with respect to the SD of the other compared products.

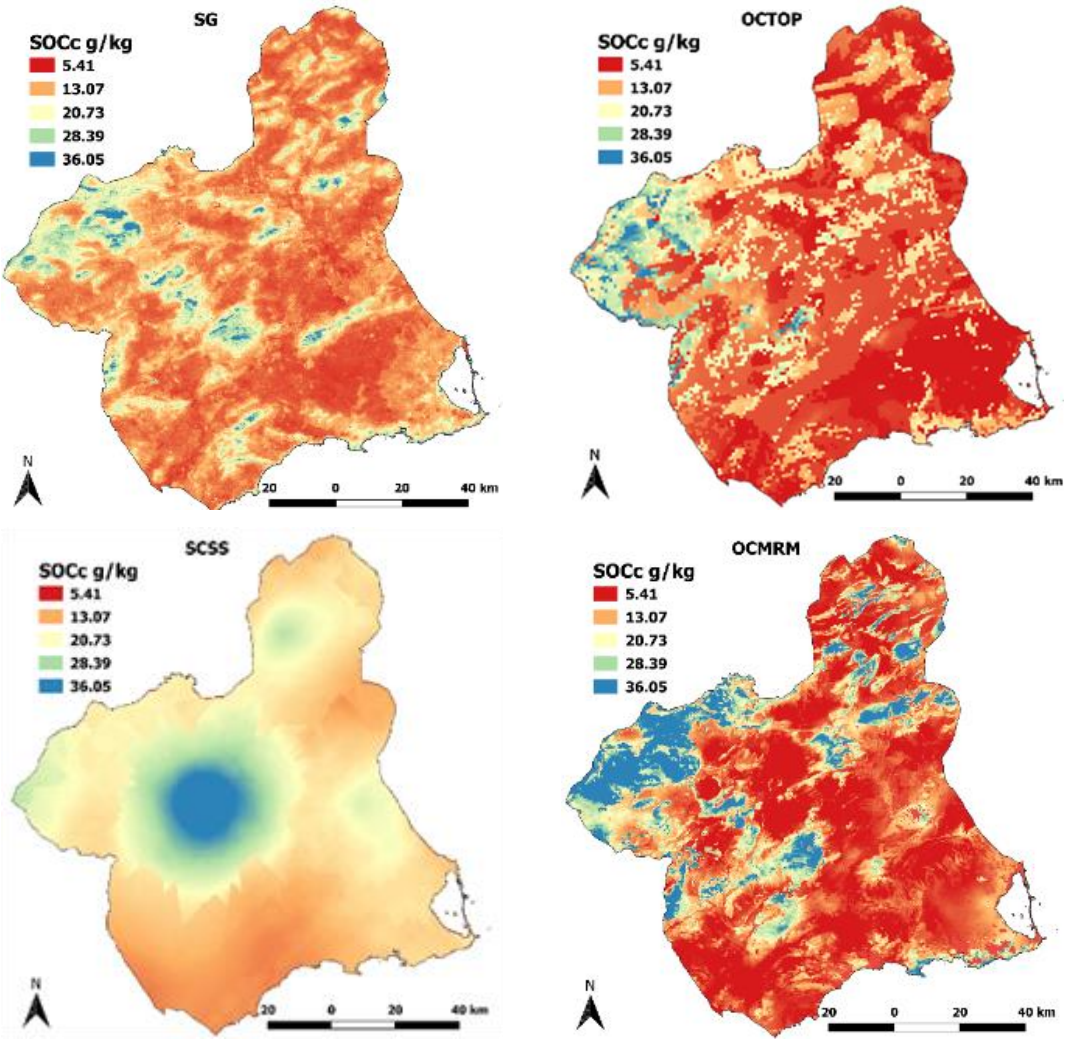


Figure 4.2. Soil organic carbon concentration (SOCc) maps prediction of available products. Data display was stretched by the cumulative pixel count cut method (default range 2%-98%). SoilGrids250m (SG), Organic Carbon Content in Topsoils (OTCOP), Soil Organic Carbon Stock in Spain (SCSS), Soil Organic Carbon Map in Region of Murcia (OCMRM)

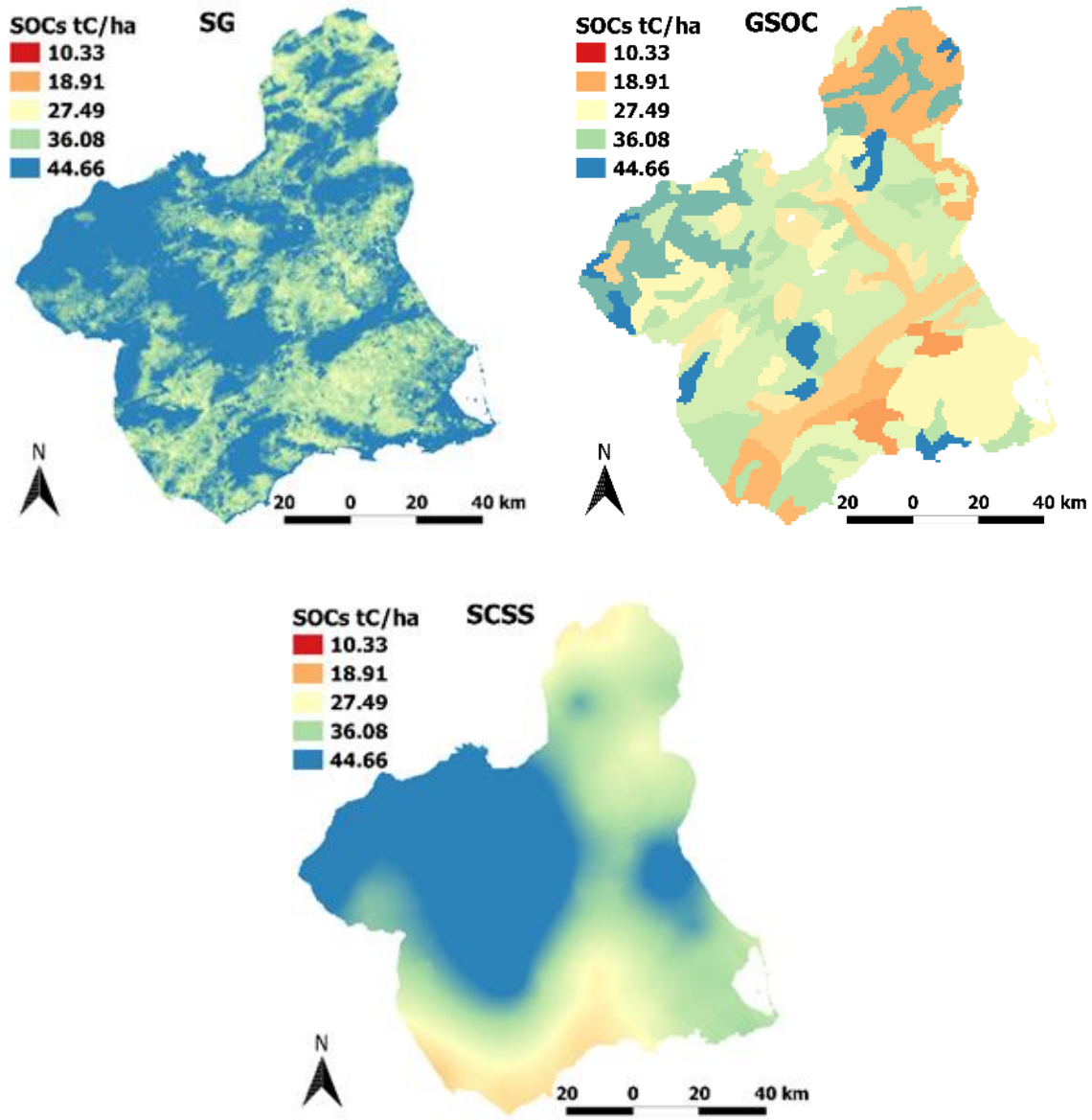


Figure 4.3. Soil organic carbon stock (SOCs) maps prediction of available products. Data display was stretched by the cumulative pixel count cut method (default range 2%-98%). SoilGrids250m (SG), Global soil organic carbon (GSOC), Soil Organic Carbon Stock in Spain (SCSS).

The external validation using the 255 soil profiles from the LUCDEME database (Table 4.4) revealed a poor fit for SOCs than for SOCc. The best fit in SOCs corresponded to SG map with R^2= 0.06 and RMSE= 25.73, and the best fit in SOCc values corresponded to the local OCMRM map with R^2= 0.53 and RMSE=14.74. A pairwise comparison with the Tukey's test, using a studentized range distribution at 95% of confident interval, indicated that the differences between the means of stock available products and the profile samples were significantly different from zero. Nevertheless, we found no statistical differences (using the 95% CI) for the SOCc values in the SG and OCMRM maps (Figure S 1).

Table 4.4. External validation of the available SOC products (SOCs and SOCc) using synthetic profiles values (255 soil samples) calculated from LUCDEME local database (MurDB). The available SOC products in the area refer to Global soil organic carbon (GSOC), SoilGrids250m (SG), Organic Carbon Content in Topsoils (OTCOP), Organic Carbon Stock in Spain (SCSS), Soil Organic Carbon Map in Region of Murcia (OCMRM).

AVAILABLE PRODUCTS[1]	SOCs		SOCc	
	R^2	RMSE	R^2	RMSE
MurDB vs. SG (Global)	0.055	25.73	0.211	11.316
MurDB vs. GSOC (Global)	$0.004 \cdot 10^{-1}$	20.88	-	-
MurDB vs. OCTOP (Europe)	-	-	0.067	12.74
MurDB vs. SCSS (Spain)	0.033	21.86	0.017	14.90
MurDB vs. OCMRM (Mur)	-	-	0.525	14.740

4.3.2. Local SOC map: Model Fitting, Spatial Prediction, and Uncertainty Assessment

4.3.2.1. Local SOC Values and Spatial Covariates

The statistical description of the SOC average profile (Figure S 2) from the local legacy database (Alias and Ortiz, 1986) showed most of the SOCc in upper horizons (20-30 cm) and decreases with depth. Profile mean value of the total depth was 8.22 g/kg (10.49 SD), and 12.22 g/kg (12.52 SD) at 30 cm depth. SOCs showed 26.71 kg/m2 (16.96 SD) as mean value in upper horizons (above 30 cm). Referring to the probability distribution, both SOCc and SOCs revealed a log-normal distribution with a right-skew (Figure S 3). However, their spatial autocorrelation behavior differed from SOCc and SOCs at 30 cm depth. We reported a nugget to sill ratio (NSR) <25% for SOCc (Figure S 4 and Table S 1), which indicates a strong spatial structure (Cambardella et al., 1994). Conversely, SOCs presented no significant spatial autocorrelation structure.

The selected environmental drivers to describe the spatial variability of SOC at 30 cm depth and at the three spatial resolutions (100m, 250m, and 1000m) were mainly related to topographic and vegetation variables (Table S 2). Specifically, the most consistent variables for predicting SOCs and SOCc were slope, plane curvature, and vegetation index (EVI) at 100 m pixel resolution. For the middle resolution (250 m) the vertical distance to channel network (related to vertical valley depth) and vegetation index (NDVI) were the most consistent. The coarsest pixel resolution (1000 m) showed that relative slope position, vegetation indexes (NDVI and EVI), and canopy cover (shrubland and total vegetation) were the most consistent variables. Moreover, we found that precipitation, vegetation indexes, tree biomass and different topographic variables as the most correlated covariates common to the three spatial resolutions in SOCs. Vertical distance to channel network, topographic variables, and vegetation indexes were for SOCc.

The log-transform of the SOC data did not significantly increase the simple correlation of SOC and its prediction factors.

4.3.2.2. Evaluation of SOC Model Fitting

The comparison among the generated local models of SOC spatial prediction based on the R^2 and RMSE as model information criteria revealed large differences in accuracy (Figure 4.4). To facilitate data interpretation, we included the tabular values in Table S 2. The QRF_G model showed much higher values than the other models for both SOCc and SOCs. In SOCc, the R^2 ranged from 0.90-0.93 and RMSE 3.38-3.93 g/kg. In SOCs the values varied form 0.91-0.94 and 4.08-5.16 GgC for R^2 and RMSE, respectively. When this model was excluded, we observed a prediction accuracy much lower in stock than concentration values at the three spatial resolutions (100 m, 250 m, and 1000 m). In this case, both information criteria (R^2 and RMSE) varied from 0.26-0.41 and 9.65-11.76 g/kg for SOCc, and 0.03-0.17 and 15.45-17.66 GgC for SOCs. The LM and QRF showed the best model performances. Referring to the spatial resolution, the errors of the models showed small differences, with the lower values at 1000 m pixel resolution.

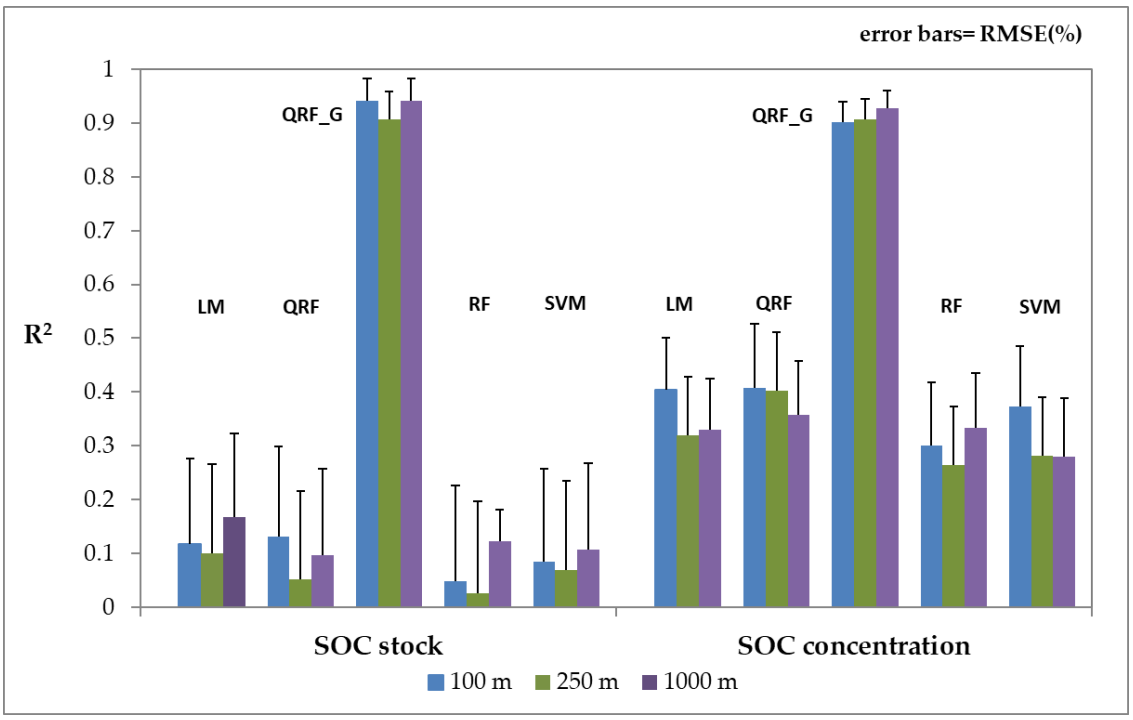


Figure 4.4. Bar diagram of R-squared and residual error –bar error- of the statistical local modeling of SOCs and SOCc (g/kg) at 30 cm depth and different spatial scales (100 m, 250 m and 1000 m). Predictive models: linear model (LM); quantile regression forest (QRF= QuantregForest R package; QRF_G= Gstat R package); random forest (RF= Caret R package); support vector machine (SVM= Caret R package).

According to the Taylor Diagrams, we compared the previous available SOCc products derived from global-to-local domain, with two generated local approaches. Among the local approaches, we selected the SOCc qrf approaches (QRF_G and QRF) not only because of the best balance between R^2 and RMSE values, but also for the advantages over uncertainty prediction. In Taylor Diagrams (Figure 4.5), the 25 m pixel resolution map (OCMRM product) had the best model-data agreement with the observed values

(LUCDEME topsoils database) and the lowest RMSE, but a greater standard deviation. Next in order were the QRF models, and more specifically those at 100 m pixel resolution. The QRF_G models showed lower accuracy of the amplitude of the variations (i.e., the standard deviation), lower correlation and higher RMSE than QRF models (Figure 4.4). In terms of model performance, SCSSc had the lowest correlation and the highest RMSE of the models.

Estimates of the covariate importance in the QRF models revealed the slope as the highest value followed by maximum temperature and plane curvature for the 100 m pixel resolution model. The most important covariates at 250 m pixel resolution were also linked to topography (vertical distance to channel network, plane curvature, and Ls-factor, listed in decreasing order) followed by climatic variables (maximum temperature) (Figure S 5).

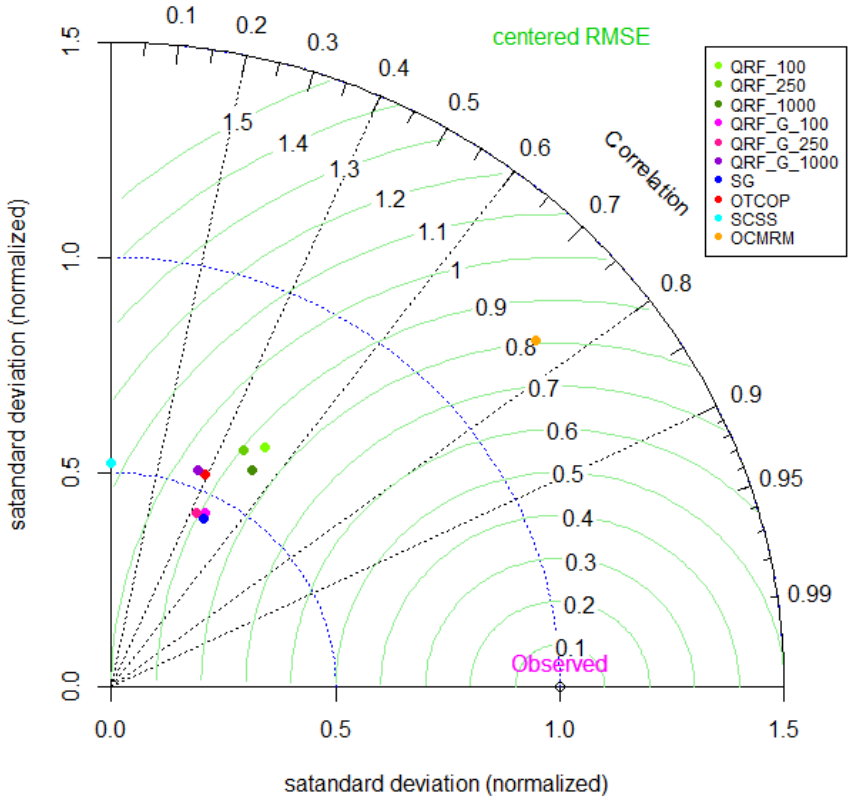


Figure 4.5. Taylor diagram of the performance of the generated local models and available products evaluated for soil organic carbon (SOC) concentration compared with the local topsoil samples (LUDEME database) as observed values. SoilGrids250m (SG), Organic Carbon Content in Topsoils (OTCOP), Topsoil Soil Organic Carbon Content (ocCont), Soil Organic Carbon Stock in Spain (SCSS), Soil Organic Carbon Map in Region of Murcia (OCMRM), quantile regression forest (QRF= QuantregForest R package; QRF_G= Gstat R package.

4.3.2.3. Local SOC Spatial Prediction and Uncertainty Assessment.

The summary of data prediction and uncertainty associated of SOCc and SOCs qrf method showed slightly different values through spatial scales in both algorithmic implementations from the R packages quantregForest (QRF) and GSIF (QRF_G) (Table 4.5). The QRF prediction had lower data than QRF_G for SOCs predicted values and their uncertainty at the three spatial resolution. The overall mean for predicted and uncertainty values for QRF were 27,763 ± 0.024 Gg C and 0.016±0.004 Gg C, respectively. The values for QRF_G were 28,761 ± 0.026

Gg C and 0.027 ± 0.017 Gg C for mean and uncertainty, respectively. However, the values for the SOCc map were higher in QRF prediction than QRF_G across the different scales, whose overall mean results were 11.58 ± 7.76 g/kg and 9.80 ± 5.32 g/kg for predicted and uncertainty values in QRF model. Their corresponding predicted and uncertainty values in QRF_G model were 10.90 ± 6.20 g/kg and 9.02 ± 13.86 g/kg where the SD values of uncertainty were more than double than QRF model.

Comparing these estimates with the previous available carbon products in the corresponding scales, the latter overestimated both SOCc and SOCs predicted values. In particular, the SCSS national map about 1.5 times in carbon concentration (19.50 vs. 12.18/11.20 g/kg C in QRF/QRF_G, respectively) and SG global map about 2.6 times in carbon stocks: 73,364 vs. 27,646/28,435 Gg C in QRF/QRF_G, respectively.

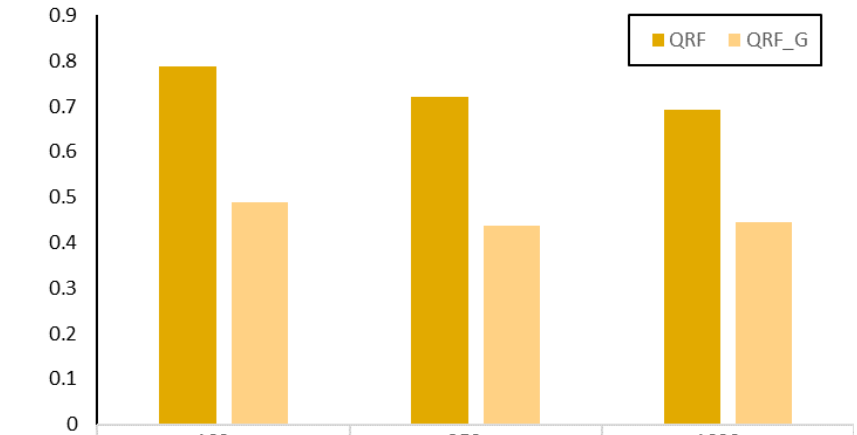
The analysis of the model residuals confirmed the absence of spatial autocorrelation structure in both SOCc and SOCs (Figure S 6).

Table 4.5. Predicted values and associated uncertainty with Soil Organic Carbon (SOC) stock and the SOC concentration estimated by two R implementations of quantile regression forest (QRF= QuantregForest R package; QRF_G= Gstat R package) at different spatial resolution for the Region of Murcia.

			Resolution spatial		
Variable			100 m	250 m	1000 m
QRF (QuantregForest)					
SOC stock (GgC)	Predicted values	Sum	27,572	27,646	28,071
		SD	0.024	0.024	0.024
	Uncertainty	Mean	0.016	0.016	0.015
		SD	0.004	0.004	0.004
SOC concentration (g/kg)	Predicted values	Mean	12.178	12.888	10.275
		SD	7.756	7.765	6.539
	Uncertainty	Mean	10.544	9.985	8.865
		SD	6.071	5.083	4.815
QRF_G (GSIF)					
SOC stock (GgC)	Predicted values	Sum	28,911	28,435	28,937
		SD	0.026	0.025	0.026
	Uncertainty	Mean	0.027	0.026	0.027
		SD	0.017	0.017	0.017
SOC concentration (g/kg)	Predicted values	Mean	11.197	10.682	10.808
		SD	6.278	5.771	6.542
	Uncertainty	Mean	9.463	8.813	8.789
		SD	14.441	12.837	14.313

4.3.2.4. Validation and Final Selection

The external validation of SOCc predicted values with the independent database of 1100 LUCDEME topsoil at 30 cm depth presented better model performance for QRF than QRF_G and, in turn, for the finer spatial scale (Table S 3). The difference between RMSE and mean absolute error (MAE) depicted larger prediction errors at coarser spatial scale (12,19 g/kg and 7.85 g/kg; and 12,21 g/kg and 8.01 g/kg for QRF and QRF_G at 1000 m spatial resolution, respectively). Additionally, we tested a final validation of the qrf maps (QRF and QRF_G) including the uncertainty of the estimated values (prediction intervals). Due to the local validation database (LUCDEME topsoils) presenting large gaps in bulk density and coarse fragments data, only the SOC concentration maps were validated (Figure 4.6). The accuracy percentage was significantly higher in QRF than in QRF_G. Regarding spatial scales, the coarser resolution the lower the accuracy percentage value was. Therefore, the results revealed higher accuracy in 100 m spatial scale, especially in QRF models (79% in QRF versus 49% in QRF_G).



	100 m	250 m	1000 m
QRF	0.787	0.722	0.692
QRF_G	0.49	0.438	0.445

Figure 4.6. Accuracy percentage (expressed as a decimal) of quantile regression forest models estimates (QRF= QuantregForest R package; QRF_G= Gstat R package). Local values of Soil Organic Carbon concentration (SOCc) of LUCDEME topsoil database were compared with the interval corresponding to predicted values and their associated uncertainty.

The QRF model at 100 m pixel resolution achieved the better balance among the predictive model performance, the agreement with the pattern of reference soil sample (i.e., topsoil local database), and the independent, external validation. In general, the values of these QRF maps ranged from 5.1 to 21.8 g/kg and 4.1 to 31.1 tC /ha for concentration and stock, respectively (Figure 4.7). Both maps were consistent in the areas of low SOC value, which were located at low-altitude areas (0 to 300 m), gentle slopes (<2%), and cultivated land. This area correspond to the driest part of the Region with less mean annual rainfall (< 200 mm) and high mean annual temperature (>15 °C). The highest values were found in coniferous forest located in the steeper and humid areas. The national-to-global available SOC products showed higher ranges with a considerable overestimation of values, especially for SOCs (Figure 4.7).

The relative uncertainty analysis revealed an inversely proportional relationship with SOC predicted data, emphasizing at the extreme values. This pattern was more pronounced for SOCc, where the lower values presented proportionally very high uncertainty (0).

The relative improvement of the prediction accuracy from SOCc and the SOCs in comparison to OCMRM and SG maps were 40.8% and 63.8%, respectively. We clarify that the OCMRM and SG maps were those with the best balance for our model validation criteria using external validation.

The final selected QRF model at 100 m pixel resolution for SOCs and SOCc, and their associated uncertainties maps will be available on a public repository at https://drive.google.com/drive/folders/1SykCBAbpHnUPHcCbguE34jiBfBIvOlnM?usp=sharing.

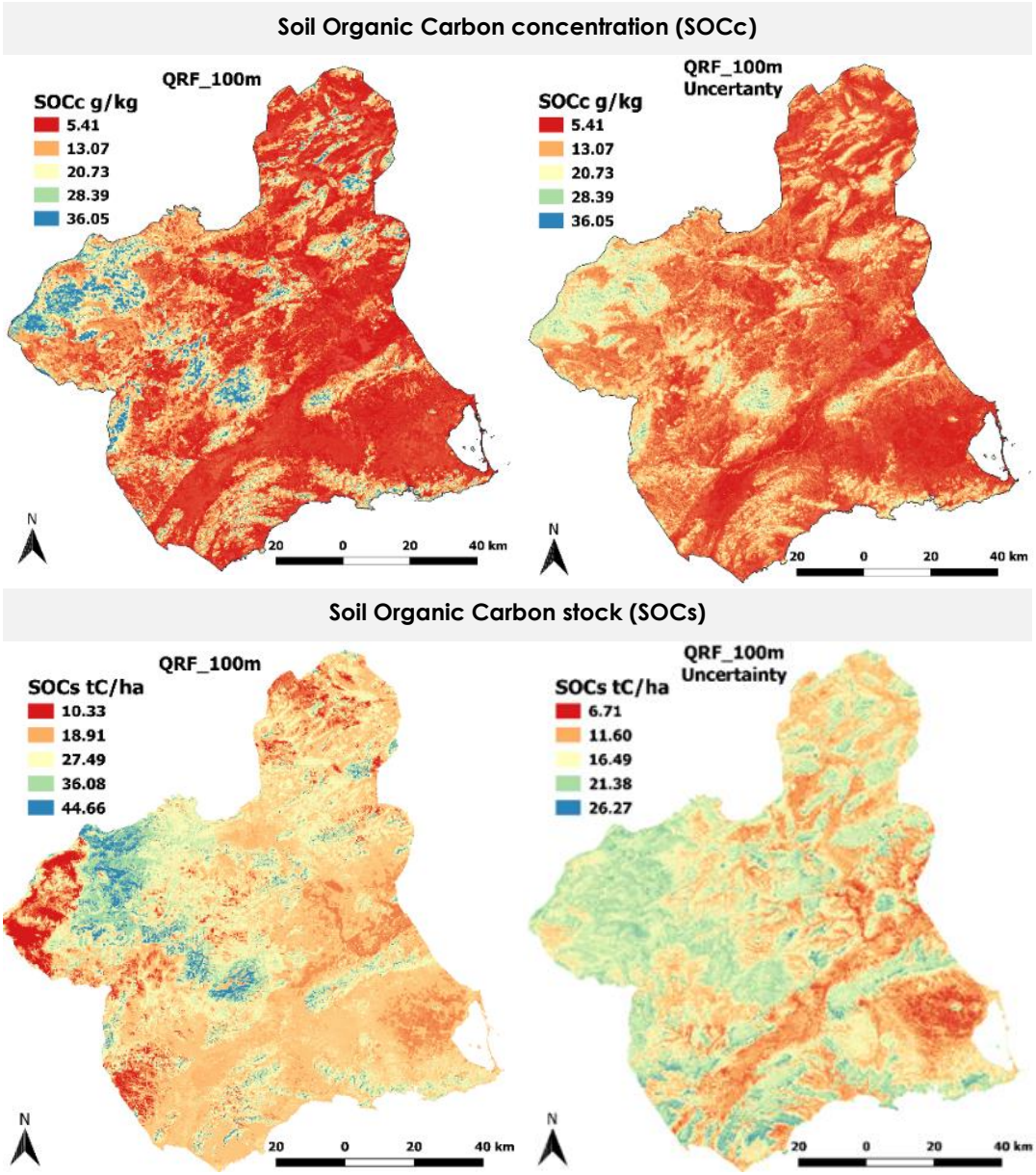


Figure 4.7. Generated prediction maps of soil organic carbon concentration (SOCc) and SOC stocks (SOCs), and their uncertainties at 100 m.

QRF= quantile regression forest models estimates by QuantregForest R package. Data display was stretched by the cumulative pixel count cut method (default range 2%-98%).

4.4. DISCUSSION

4.4.1. Comparison of Current Available SOC Products

Different soil dataset collections, scales, and spatial predictive modeling for SOC estimates lead to different interpretations for SOCc and SOCs across the study area. These differences highlight the challenges associated to SOC prediction due to its high-spatial heterogeneity, especially in semiarid regions. Discrepancies in stocks were larger than concentration values among the products analyzed. This is probably due to a generalized absence in most soil databases of coarse soil material (> 2 mm) and bulk density, the main parameters required for estimation of SOCs. These important variables are often derived from pedotransfer or extrapolation functions (Jalabert et al., 2010) resulting in additional bias and a systematic error for calculation of SOCs (Durante et al., 2020; Poeplau et al., 2017).

In relation to pixel scale, there was no clear consistency among the compared products. Our results do not support the expectation that higher spatial resolution will have a better agreement with locally derived SOC information. We highlight the similarity in SOCc maps between the global "SoilGrids250m" (SG) and the local "Soil Organic Carbon Map in Region of Murcia 25 m" (OCMRM). Both maps presented the most similar estimates (g/kg) and the more accurate balance of model validation criteria (R^2 and RMSE) in the independent, external validation. However, the number of soil samples (i.e., training data) to perform the spatial modeling of these two maps were very different. For the SG map, the estimates were extrapolated, as none of the 43 soil samples used for peninsular Spain was located in the study area. For the OCMRM map, more than 1900 topsoil samples were collected. However, both maps were modeled using a similar machine learning technique (i.e., random forest) integrating environmental soil forming factors at 250 m and 25 m pixel resolution, respectively. While the selected covariates in the OCMRM map referred to climate, land cover, soil types and terrain morphology, the SG prediction model included environmental layers from remote sensing data. Therefore, our results suggest that the application of machine learning approaches in combination with single and/or multitemporal remotely sensed satellite indices improved the prediction of soil carbon spatial distribution even when there exist limited soil samples (Fathololoumi et al., 2020; Liang et al., 2020; Schillaci et al., 2017a).

We postulate that the SG map could be considered as the best available regionally-derived product for the study area with a moderate spatial pixel resolution, but this product has some limitations. Several studies that compared estimated SOCs values with SOG map showed comparatively wide ranges and a considerable overestimation, especially in low SOCs values (Hengl et al., 2017; Silatsa et al., 2020; Vitharana et al., 2019). This overestimation may be due to the absence of profile observations that can lead to the application of greater weightage for covariates in the spatial predictions (Lombardo et al., 2018; Vitharana et al., 2019). In our results, the SG map estimated the highest stock values in both spatial domains (national and local study area) with respect to the available products compared. These observations confirmed the need of reassessing local-scale estimates of soil SOCs to gain insights into the effect of different methodological criteria in the estimation of SOC spatial variability. Thus, we tested these observations with a locally-derived information and different DSM techniques at three different scales. The relative improvement of the

prediction accuracy of the SOCc and the SOCs were about 65% and 40% in comparison to stock and concentration maps, respectively (i.e., the maps with the most accurate balance of model validation criteria in the independent, external validation).

4.4.2. Local Prediction of SOC

According to our hypothesis, our results showed how machine learning models rely on locally-derived information and appropriate environmental data better capture short-range SOC spatial variation than extrapolated global or regional estimates.

Consistent with other studies, local sampling distribution was more relevant than density to capture the SOC variation (Zeraatpisheh et al., 2019). The two maps derived using the local database showed low difference in the prediction accuracy despite their different sampling density (i.e., 1,922 systematic sampling vs. 255 representative profiles of soil taxonomic for OCMRM and QRF_100 maps, respectively).

As resulted in the comparison of current available SOC products, the accuracy of the predictive models was higher in SOCc than SOCs. This accuracy difference can be attributed to the propagation error of SOCs extrapolated parameter (i.e. coarse fragments and bulk density). The application of these parameters derived from pedtotransfer functions could introduce a considerable bias (Poeplau et al., 2017).

Among the SOC prediction models used in this study, the selection of qrf techniques was based not only on the model performance but also on the advantages of data-driven machine learning methods in DSM (Hengl et al., 2017; Yigini et al., 2018). Moreover, the added value of qrf approach is the addition of the spatial uncertainty associated with SOC estimates, calculated as suggested by (Vaysse and Lagacherie, 2017). Specially for sparse sampling studies, the lack of uncertainties values compromises the reliability of predictions for decision making (Maia et al., 2010; Ogle et al., 2010; Schillaci et al., 2017a).

Nevertheless, both qrf implementations applied in this study (QRF and QRF_G according to 'quantregForest' and 'GSIF' R statistical computing packages, respectively) showed large discrepancies in their model performance accuracy. This is probably caused by the regression-kriging techniques (i.e., predictions plus kriging of the residuals) implemented in QRF_G. This strategy can lead to overfitting when variogram modeling is under-sampling. For a well-represented multivariate feature space is preferably 300 sample points and at least 10 observations per covariate (Hengl et al., 2017; Webster and Oliver, 2001). This could explain the discrepancies of the QRF_G models, which showed the best model performance and, however, the lowest accurate in external validation at the three spatial resolution tested (100 m, 250 m and 1000 m). This approach highlight the importance of model external validation to identify the most accurate estimation. An accurate model is not necessarily related to the best model performance, as it may represent an overparameterization (Zeraatpisheh et al., 2019).

Regarding to the spatial resolution, the errors of estimated predictive models showed small differences among 100 m, 250 m and 1000 m pixel size. This is probably related to the challenge for estimating SOC high spatial heterogeneity across scales, especially in semiarid regions with complex terrain (Hoffmann et al., 2014; Hounkpatin et al., 2018; Jobbágy et al.,

2000; Kulmatiski et al., 2004; Kunkel et al., 2011).Calvo de Anta et al. (2020) also revealed this heterogeneity in the national level study where the semiarid Region of Murcia registered the highest coefficient of variation in both the SOCc (76%) and the SOCs (61%) at 0-30 cm depth.

We addressed this challenge in our framework by including a set of covariates as a function of the soil forming environment (SCORPAN model). Moreover, the integration of remote sensing covariates likely increased the predictive capacity of our models as it has been demonstrated for other applications with limited soil samples (Fathololoumi et al., 2020; Liang et al., 2020; Schillaci et al., 2017a). Our results showed that SOCc and SOCs maps at 100 m spatial resolution had the best balance among accuracy, external validation, and interpretability of results.

In terms of spatial patterns, we observed that the larger disagreement between SOCc and SOCs maps at 100 m spatial resolution were in areas with low soil sampling density. In these areas, SOCc maps depicted a spatial distribution coherent to the land use and landscape patterns (Albaladejo et al., 2009). These areas represent the highest altitude for forested regions (>1,400 m) in the northwestern and central zone. These areas have an upper horizon rich in organic matter with natural grassland and tree forest over lithosols that, although shallow (≤ 10 cm), are covered by abundant vegetation. In addition, the low SOCc values in the eastern area also showed disagreements, where the main cover type is sclerophyllous vegetation and rainfed crops over regosols, with scarce incorporation of organic matter. The highest uncertainty of SOCc and SOCs were associated with these low values corresponding to areas with an advanced process of soil degradation. These high uncertainty values could explain the low model accuracy, since degraded soils process may alter the relationship pattern with dominant environmental covariates, different from the general prediction model (Brungard et al., 2015). To better capture the representativeness of the SOC spatial variability, a stratified sampling of homogeneous sub-areas would reduce the highest uncertainty estimates (Zeraatpisheh et al., 2019).

4.5. CONCLUSIONS

In this study, we improved global and regional available SOC estimates in a complex soil-forming environment across a Mediterranean area in southeastern Iberian Peninsula. The different available SOC products for the study area showed large divergence among their values, regardless of spatial domain or scale resolution.

We observed a lower accuracy and an overestimation values in stocks maps compared to the spatial predictions of carbon concentration maps, probably due to the often-missing information of key carbon stock parameters in databases. Unless essential, we recommend to develop SOC concentration maps to adapt proper environmental policies to soil carbon management as an alternative to overused stock data.

Our high-resolution SOC map framework was based on local legacy soil data, environmental covariates (including single and/or multitemporal remote sensing indices), DMS modeling and spatially explicit uncertainty quantification. This latter aspect, together with independent, external validation, is essential to interpret the soil carbon properties distribution, especially in SOC complex quantification areas.

Our results showed potential to improve the representation of national scale SOC estimates. We provided a better understanding of SOC storage across complex soil-forming environments with persistent soil erosion and degradation. This is important for responding to the challenges of land management and climate change adaptation/mitigation policies and strategies.

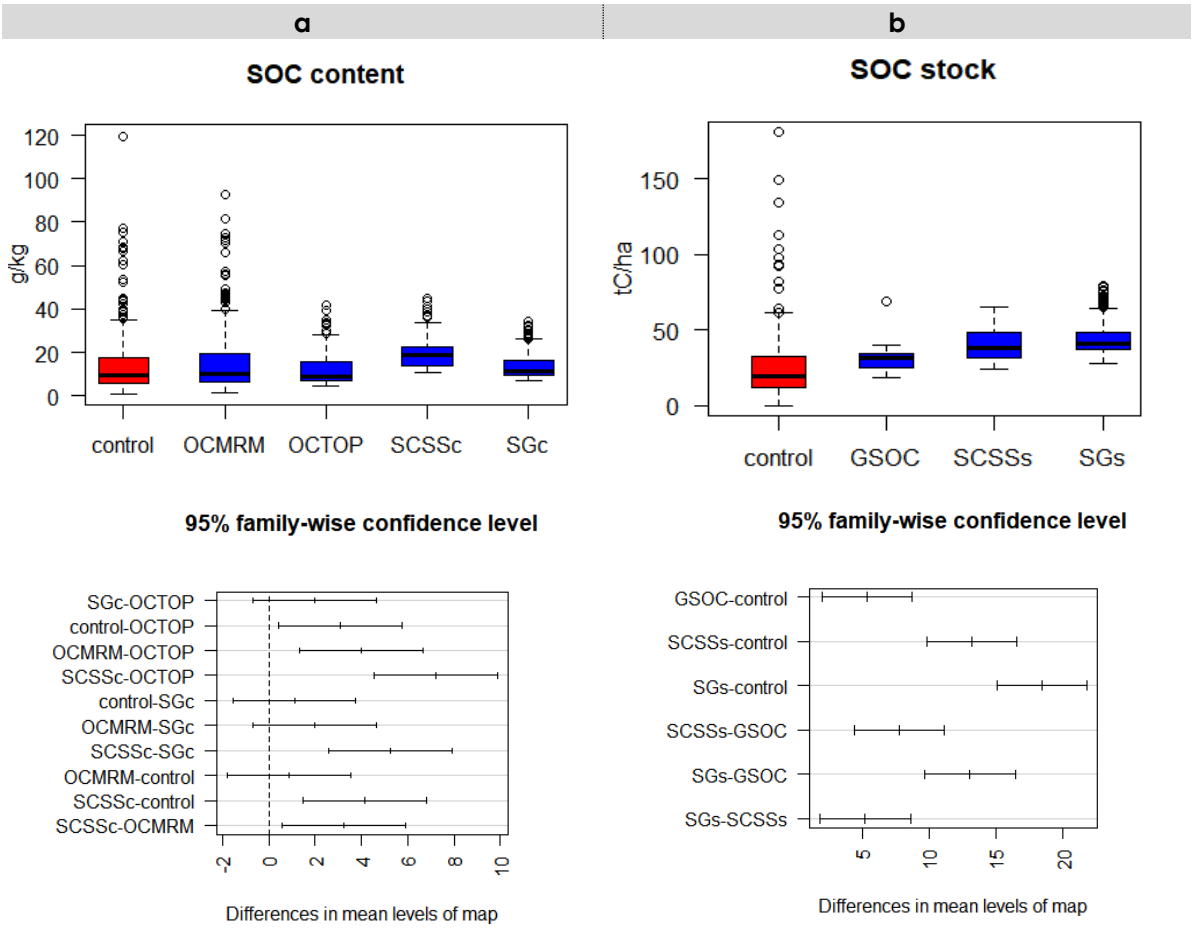
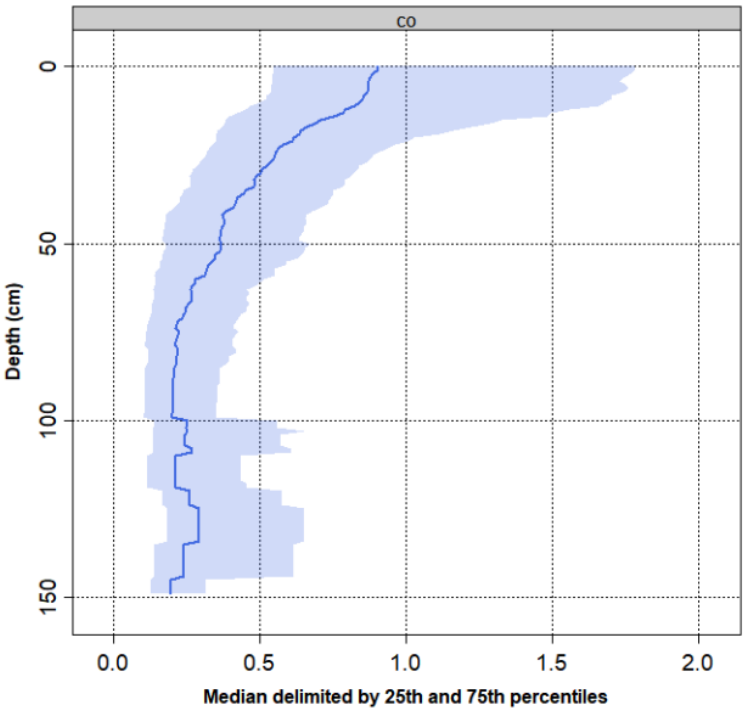


Figure S 1. Graphical representation of SOC concentration (a) and SOC stock (b) values at the profile sample locations of the available SOC products in the study area: Global soil organic carbon (GSOC), SoilGrids250m (SG), Organic Carbon Content in Topsoils (OTCOP), Topsoil Soil Organic Carbon Content (ocCont) Soil Organic Carbon Stock in Spain (SCSS), Soil Organic Carbon Map in Region of Murcia (OCMRM). Boxplot (top) and 95% confidence interval of Tukey´s test (bottom).



SOC concentration profile mean:

Min.: 0.10

1st Qu.: 2.20

Median :4.90

Mean: 8.22

3rd Qu.: 9.25

Max.: 79.40

Sd: 10.49

SOC concentration 30 cm:

Min.: 0.26

1st Qu.: 4.91

Median :7.91

Mean: 12.22

3rd Qu.: 14.59

Max.: 75.41

Sd: 12.52

SOC stock 30 cm:

Min.: 0.08

1st Qu.:13.398

Median :23.19

Mean: 26.71

3rd Qu.:35.574

Max.: 83.25

Sd: 16.96

Figure S 2. Soil organic carbon (SOC) concentration (%) average profile at different depths. On the right, summary (in g/kg) of the profile mean value of SOC concentration, SOC concentration data (g/kg) at 30 cm depth and SOC stock (kg/m^2) at 30 cm depth.

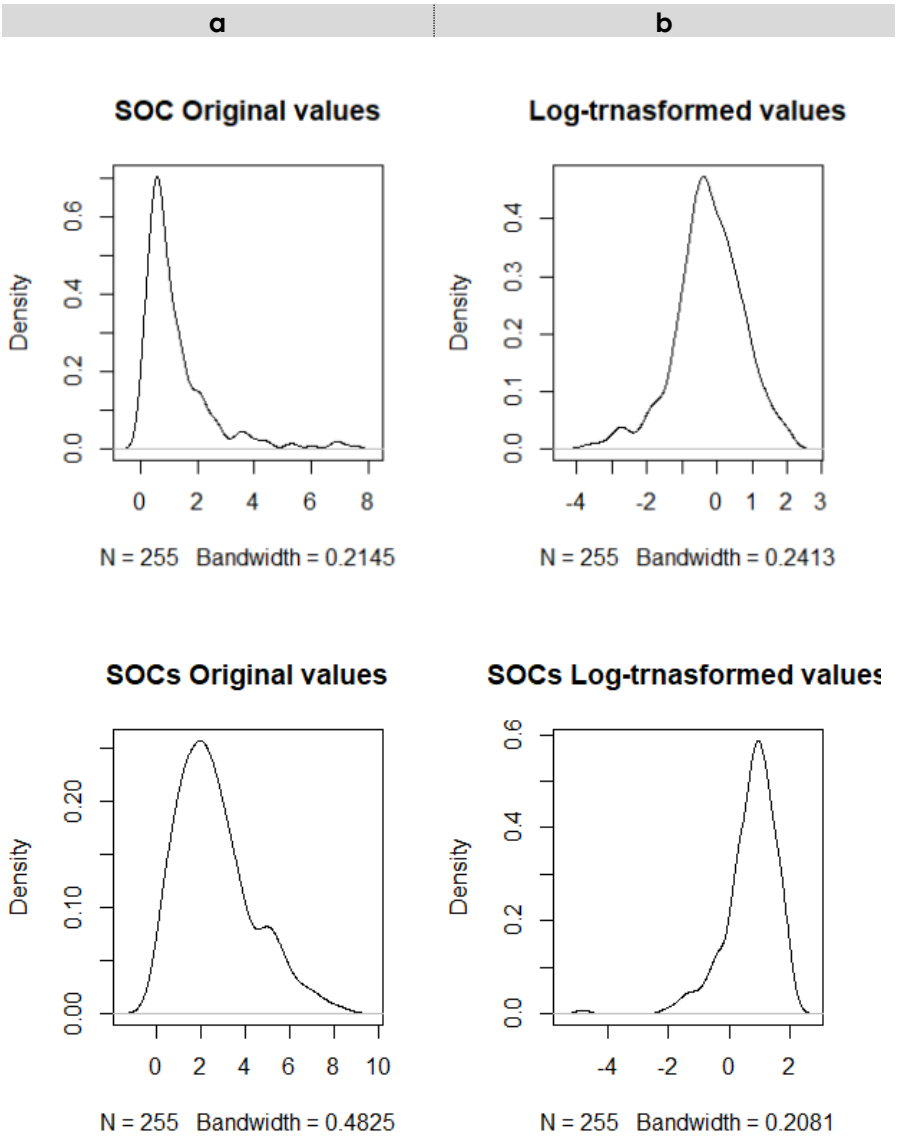


Figure S 3. Probability distribution plot of soil organic carbon concentration (SOCc) (top) and soil organic carbon stocks (SOCs) (bottom): originall values (a) and its natural logarithm (b) at 30 cm depth.

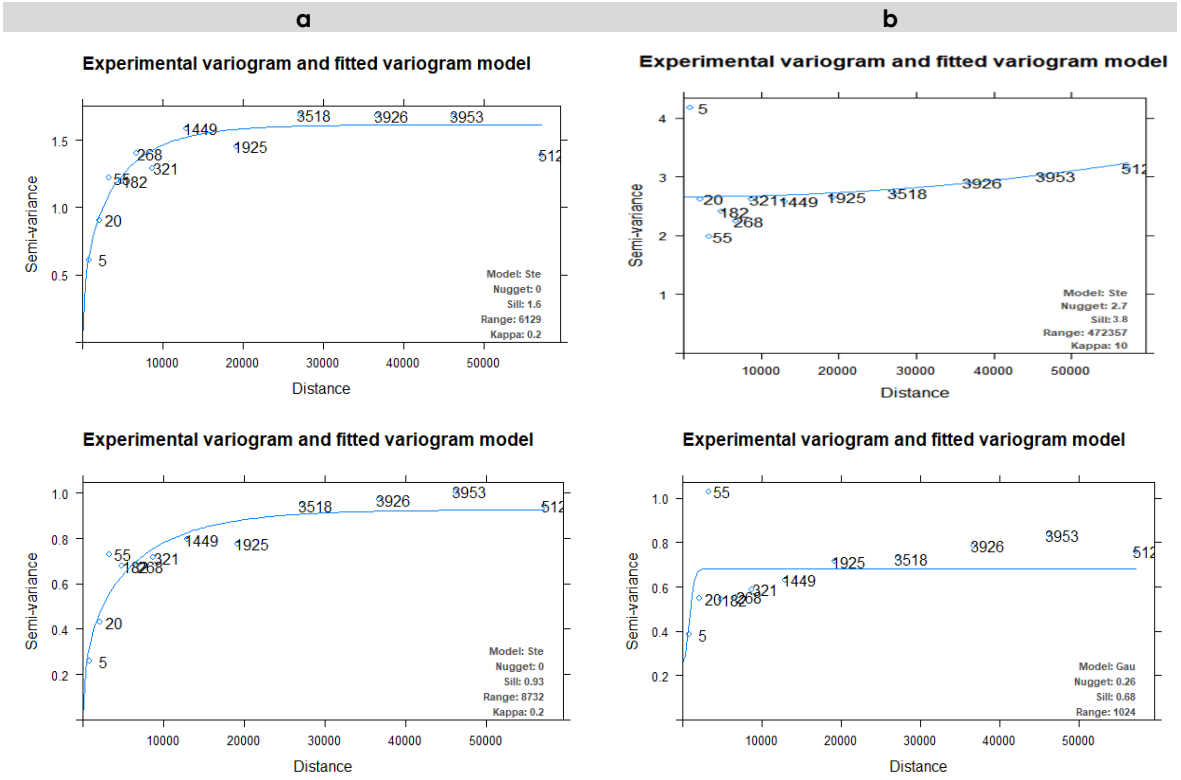


Figure S 4. Semivariogram of original values (top) of Soil Organic Carbon (SOC) concentration (a) and SOC stock (b) and their natural logarithm (bottom) of soil profiles sample.

Table S 1. Semivariogram data of Soil Organic Carbon (SOC) concentration and stock data at 30 cm, and their logarithms of soil profiles. Sph= spherical model; Ste= Matern, M. Stein's parameterization; Gau= Gaussin model.

	Model	Nugget	Sill:	Range:	NSR
SOC concentration at 30 cm depth	Ste	0	1.6	6129	0.00
Log SOC concentration at 30 cm depth	Ste	0	0.93	8732	0.00
SOC stock at 30 cm depth	Ste	2.7	3.8	472357	0.71
Log SOC stock at 30 cm depth	Gau	0.3	0.68	1024	0.44

Table S 2. R-squared and residual error of the statistical local modeling of Soil Organic Carbon (SOC) stock (GgC/ha) and SOC concentration (g/kg) at 30 cm depth and different spatial scales (100 m, 250 m and 1000 m). The predicted models were linear model (LM); quantile regression forest (QRF= QuantregForest R package; QRF_G= Gstat R package); random forest (RF= Caret R package); support vector machine (SVM= Caret R package).

MODEL(R package) / SOC VARIABLE		Spatial resolution		
		100 m	250 m	1000 m
LM				
SOC stock	R^2	0.117	0.101	0.168
	RMSE	15.99	16.05	15.45
	Covariables(1)	S, PIC, PC, VD, PP, EVI	AS, VDCN, PP, NDVI	DEM, CNBL, RSP, MRVBF, PP, EVI, CCs
SOC concentration	R^2	0.405	0.320	0.329
	RMSE	9.65	10.30	9.65
	Covariables(1)	S, PIC, VDCN, EVI, TB	AH, PLC, VDCN,Tmax, NDVI	AH, CI, VDCN, RSP, NDVI, TCC
QRF				
SOC stock	R^2	0.132	0.051	0.097
	RMSE	16.002	16.487	16.085
	Covariables(1)	PP, CNBL, VDCN, AS, EVI	PP, MRVBF, CNBL, NDVI	DEM, CNBL, RSP, MRVBF, PP, EVI, CCs
SOC concentration	R^2	0.408	0.403	0.358
	RMSE	10.856	10.849	10.010
	Covariables(1)	S, Tmax, PIC, AS, TCC, EVI, AA	VDtCN, PIC, LsF,Tmax, AS, CI, CCt	AH, CI, VDCN, RSP, NDVI, TCC
QRF_G				
SOC stock	R^2	0.941	0.907	0.942
	RMSE	4.119	5.16	4.08
	Covariables(1)	S, PIC, PC, VD, PP, EVI	AS, VDCN, PP, NDVI	DEM, CNBL, RSP, MRVBF, PP, EVI, CCs
SOC concentration	R^2	0.901	0.907	0.927
	RMSE	3.929	3.82	3.38
	Covariables(1)	S, PIC, VDCN, EVI, TB	AH, PIC, VDCN,Tmax, NDVI	AH, CI, VDCN, RSP, NDVI, TCC
RF				
SOC stock	R^2	0.049	0.026	0.122
	RMSE	17.66	17.16	15.90
	Covariables(1)	S, PIC, PC, VD, PP, EVI	AS, VDCN, PP, NDVI	DEM, CNBL, RSP, MRVBF, PP, EVI, CCs
SOC concentration	R^2	0.300	0.265	0.334
	RMSE	11.76	10.78	10.2
	Covariables(1)	S, PIC, VDCN, EVI, TB	AH, PIC, VDCN,Tmax, NDVI	AH, CI, VDCN, RSP, NDVI, TCC
SVM				
SOC stock	R^2	0.085	0.069	0.107
	RMSE	17.17	16.55	16.14
	Covariables(1)	S, PIC, PC, VD, PP, EVI	AS, VDCN, PP, NDVI	DEM, CNBL, RSP, MRVBF, PP, EVI, CCs
SOC concentration	R^2	0.373	0.282	0.279
	RMSE	11.23	10.78	10.96
	Covariables(1)	S, PIC, VDCN, EVI, TB	AH, PIC, VDCN,Tmax, NDVI	AH, CI, VDCN, RSP, NDVI, TCC

(1) Covariables: DEM: Digital Elevation Model, AH: Analytical Hillshading, S: Slope AS: Aspect, PlC: Plan Curvature, PC: Profile Curvature, LC: Longitudinal Curvature, CI: Convergence index, CD: Closed Depressions, TCA: Total Catchment Area, TWI: Topographic wetness index, LsF: Ls Factor, CNBL: Channel network base level, VDCN: Vertical Distance to Channel Network, VD: Valley depth, RSP: Relative slope position, MRVBF: Multiresolution valley bottom flatness index, MRRTF: Multiresolution ridge top flatness index, NDVI: Normalized Difference Vegetation Index (mean annual), EVI: Enhanced Vegetation Index (mean annual), PP: Precipitation (mm), Tmax: Annual Mean Maximum Temperature (°C), Tmin: Annual Mean Minimum Temperature (°C), Tmed: Annual Mean Temperature (°C), TCC: Total Canopy Cover, CCs: Shrub Canopy Cover, CCt: Tree Canopy Cover, TB: Tree Biomass.

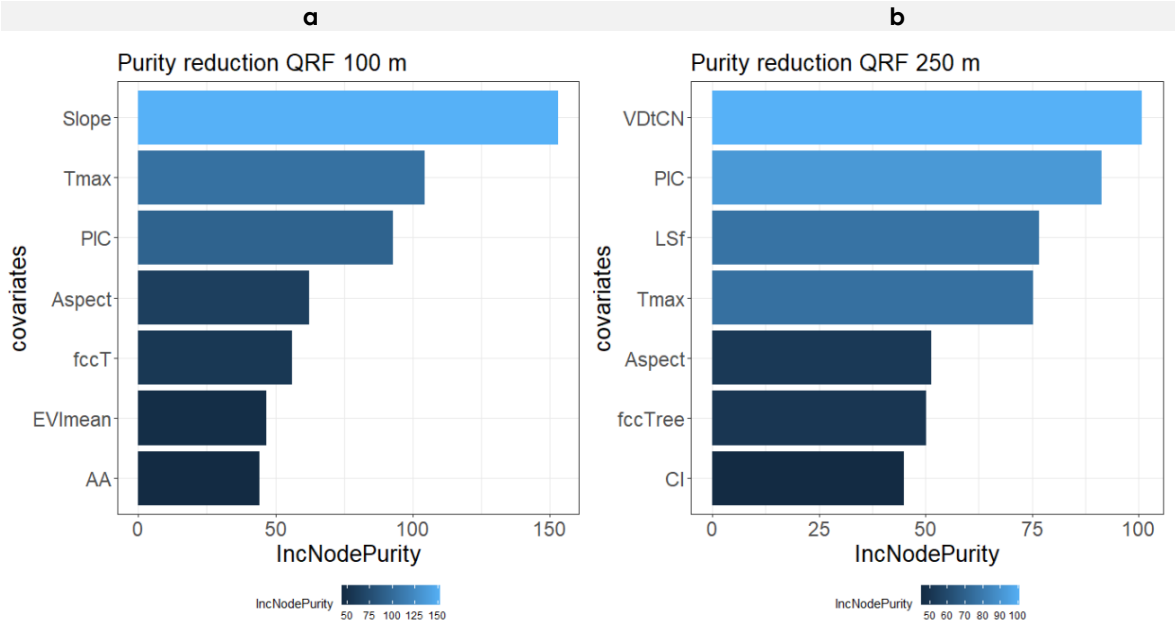


Figure S 5. Importance of predictors according to their influence on increment mean squared error and node purity for Soil Organic Carbon concnetration (SOCc) at 100 m (a) and 250 m (b) spatial resolution.

Covariables: AA: Analytical Hillshading, PIC: Plan Curvature, LsF: Ls Factor, VDtCN: Vertical Distance to Channel Network, CI: Convergence index, EVI: Enhanced Vegetation Index (mean annual), Tmax (°C), fccT: Tree biomass, fccTree: Tree Canopy Cover.

Table S 3. External validation of predicted values of soil organic carbon concentration (g/kg) with LUCDEME topsoil database at 30 cm depth. R-squared (R^2) and residual error (RMSE) and MAE were used as information criteria of the validation for both quantile regression forest methods (QRF= QuantregForest R package; QRF_G= Gstat R package).

MODEL / VARIABLE		Spatial resolution		
QRF		100 m	250 m	1000 m
Concentration (g/kg)	R^2	0,272	0,279	0,144
	RMSE	10,99	10,97	12,19
	MAE	7.49	7.75	7.85
QRF_G				
Concentration (g/kg)	R^2	0,208	0,179	0,128
	RMSE	11,41	11,69	12,21
	MAE	7.72	7.93	8.01

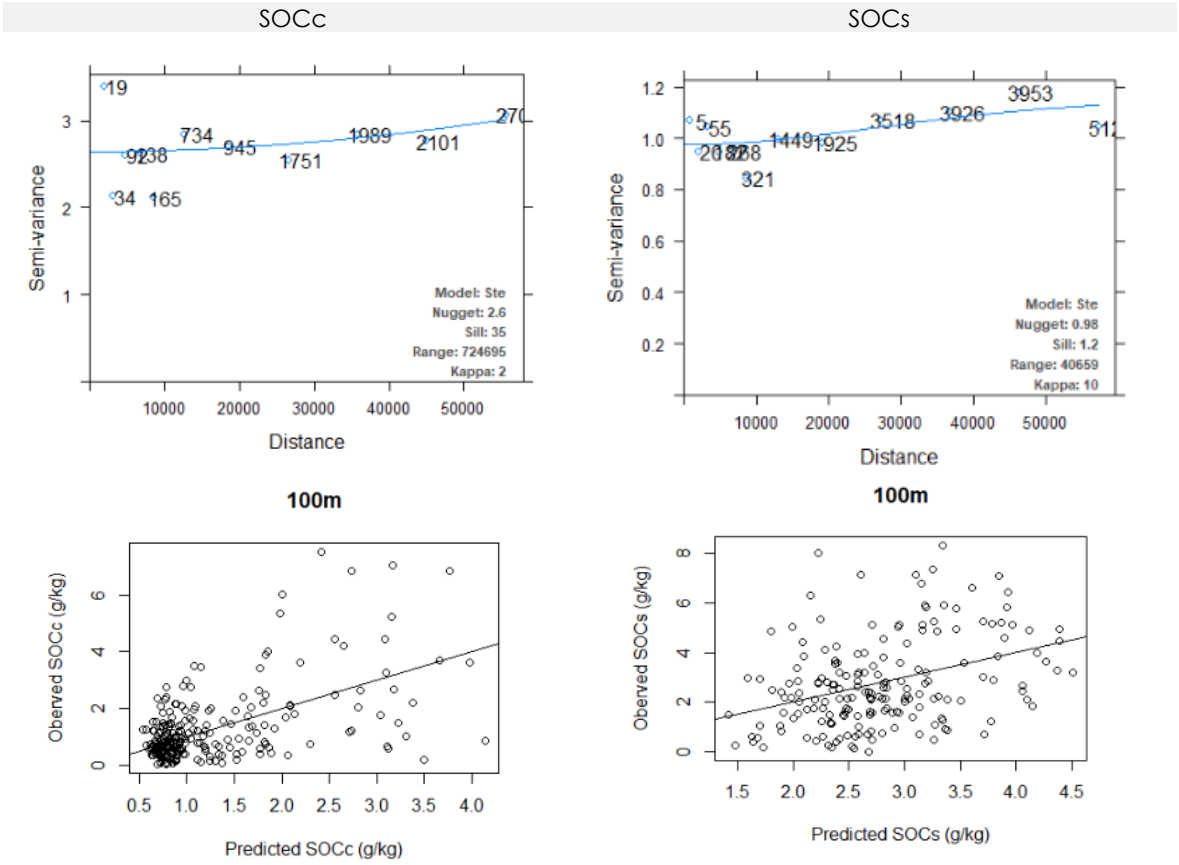
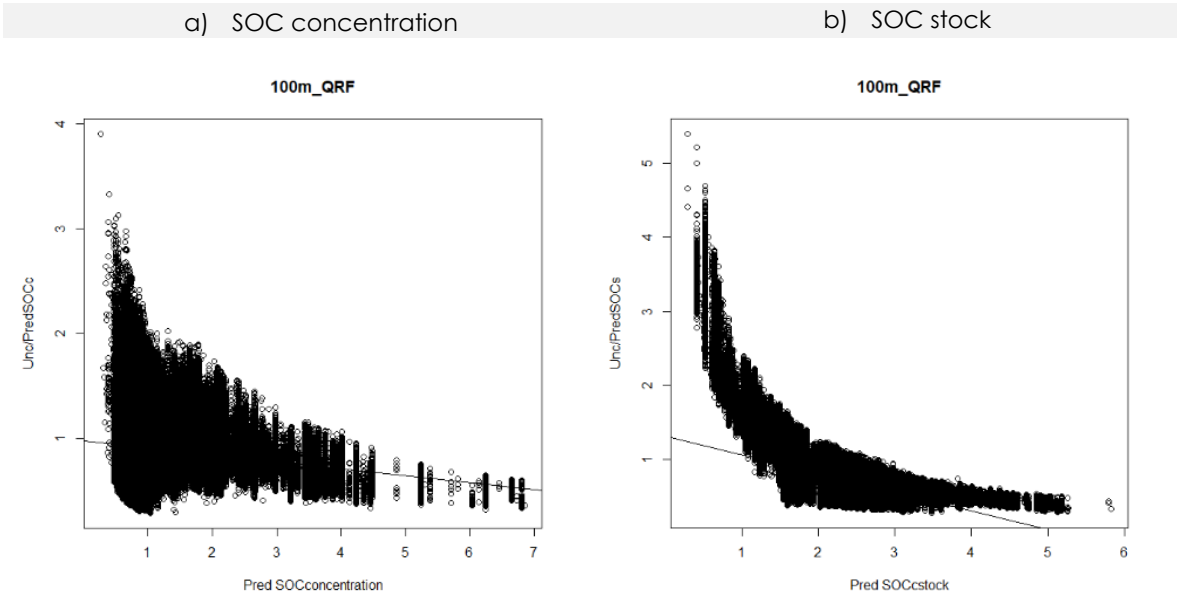


Figure S 6. Semivariogram of the residuals (top) and residuals plot (bottom) for the Soil Organic Carbon (SOC) concentration (SOCc) (left) and stock (SOCs) (right) of the local quantile regression forest method (QRF= QuantregForest R package) at 100 m spatial resolution.



Scatter diagram of predicted values of Soil Organic Carbon (SOC) concentration(g/kg, a) and stock (GgC/ha, b) versus their associated relative uncertainty of quantile regression forest method (QRF= QuantregForest R package) at 100 m..

Capítulo 5. Spatial Distribution Modeling and Quantification of Soil Organic Carbon in Peninsular Spain

ABSTRACT

Soil organic carbon (SOC) must be quantified and monitored to assess soil management practices, adapt policies, and evaluate environmental impacts. However, due to SOC spatial variability, soil surveys become a very challenging task because of the high costs of acquiring data, operational complexity, and updating. Digital soil mapping based on machine learning approaches in combination with remote sensing techniques have enabled soil carbon spatial distribution to be significantly improved, even with limited soil samples. A legacy soil database of 8,361 georeferenced profiles and a selection of environmental data-driven covariates intimately related to soil-forming factors (e.g., biota, climate, parent material) were used to generate SOC maps. Modeling of data was based on three supervised learning approaches. quantile regression forest, ensemble machine learning and auto-machine learning**.** For the final SOC spatial distribution maps, each pixel was assigned the prediction from the most accurate model, i.e., lowest uncertainty.

We applied this modeling technique to generate cost-effective, high-resolution maps (90 m pixel resolution) of SOC distribution, and its associated spatially explicit uncertainty, in peninsular Spain. These maps showed 15.7 g.kg^{-1} mean SOC concentration at 0-30 cm and 3.6 g.kg^{-1} at 30-100 cm depth. The total SOC stock at its effective depth was 3.8 Pg C, storing the 74% in the upper 30 cm (2.82 Pg C). The correlation between SOC observed and predictions final values showed R^2=0.68 for SOCc and R^2=0.54 for SOCs at the upper 30cm.

The methodology proposed in this study aims to improve benchmark SOC estimates in support of the National GHG Emissions Inventory Report.

Keywords: Effective soil depth, stock, feature selection, Bayesian analysis, partial dependence plot, ensemble learning.

5.1. INTRODUCTION

Human activity has altered the carbon cycle, increasing the amount of CO_2 in the atmosphere and modifying the natural capacity of ecosystems as sinks. In this context, the soil, able to store from about two to four times more organic carbon in the first meter than is found in vegetation (1500 Pg C), is the main reservoir of carbon in terrestrial ecosystems. Soil organic carbon (SOC) residence time is also longer than in biomass, making carbon storage in soil a key service in climate change mitigation strategies (Jobbágy et al., 2000; Saatchi et al., 2007).

Knowledge of soil properties becomes necessary for adequate decision-making on resource management and planning, as well as quantifying and monitoring changes. However, long-term change is inherent in soil, so it is critical to work on various spatial and temporal scales (Stockmann et al., 2015). Therefore, numerous national and global efforts have been made to compile and standardize empirical information on soil properties generated over decades and make them available (Harden et al., 2018; Shangguan et al., 2014).

In the case of Spanish soils, the extremely valuable information acquired over more than fifty years is currently fragmented and dispersed. Organizations and institutions have generated this information for years and compiled information on thousands of soil profiles in databases. The information comes from a wide variety of sources using different procedures, laboratory methods, standards, scales and georeferencing systems (Llorente et al., 2018). Using these legacy databases, that is information on soil compiled historically within a territory, involves an enormous search and processing effort to form a set of common, coherent and geographically well-defined data. Thus, although SOC in particular is one of the most well-studied properties in Spanish soils (Calvo de Anta et al., 2020), not all the databases generated are available. And even though SOC stock (kg C/ha) is a frequently estimated variable due to its relationship with regulation of the organic carbon service, the databases available are significantly short on parameters fundamental to its calculation, such as apparent density and coarse fragments (Poeplau et al., 2017). Therefore, the lack of accessibility to databases with consistent, complete information could lead to biased interpretation of the stock in different ecosystems, and further, to the distribution of disperse or clustered sampling points. At present, soil sampling is still challenging due to its high cost and operational complexity, which impedes updating of soil databases over time, and hinders interoperability (Smith et al., 2020; Vargas et al., 2017).

The behavior of SOC is determined by soil physical and chemical properties, and its interrelationships with environmental soil-forming factors. These relationships have been widely described, as in the soil formation equation by Jenny, (1941). However, soil properties show diverse and complex scales of variation resulting from the wide spatial-temporal range of soil genesis (Allen & Starr, 1982). As a result, information on soil has been evolving from an originally qualitative perspective toward accurate quantitative estimations and evaluation of uncertainty through empirical models, such as the CLORPT or SCORPAN (Jenny, 1941; McBratney et al., 2003).

Advances in digital soil mapping (DSM) have provided a large number of modeling techniques based on empirical models. These techniques rely on estimation of soil properties measured *in situ* (target variable), where predictive environmental factors are available at the same location (explanatory variables). These statistical models estimate soil properties quickly and reliably, minimizing the number of samples necessary for prediction (McBratney et al., 2003; Searle et al., 2021).

Nevertheless, the wide variability in methodologies used, due mainly to disparity in soil data periods, sampling density, protocols, stock calculation approaches, prediction models, or spatial resolution, have resulted in considerable variation in SOC estimates. In Spain, different SOC stock values have been described for specific agricultural systems (Albaladejo et al., 2009; Álvaro-Fuentes et al., 2008; Muñoz-Rojas et al., 2012), and forests and pasture land uses (Doblas-Miranda et al., 2013). At national scale, several authors have estimated SOC stock ranging from 2.82 Pg C (Rodríguez Martín et al., 2016) to 3.25 Pg C (Calvo de Anta et al., 2020) in the upper 30 cm for peninsular Spain.

In addition, most of the information on SOC is currently standardized to the upper 30 cm. However, the subsoil carbon pool (>30 cm) may contribute up to 50% of the total stock in Mediterranean soils (Mulder et al., 2016), and would therefore be underestimating a large part of carbon in the effective soil depth.

Thus, standardized and comparable procedures are crucial for sharing and using information, removing conceptual, technological, organizational and cultural barriers (Vargas et al., 2017). Among the major global standardized procedures for SOC spatial estimations are the Global Soil Organic Carbon Map (GSOCmap) a project of the FAO and the Global Soil Partnership (GSP), and the GlobalSoilMap.net Project, an initiative of the DSM Working Group of the International Union of Soil Sciences (IUSS). Both projects have designed, validated and verified their methodologies, developing guidelines and technical specification reports for their replication.

Based on this background, the aim of this study was to provide basic information on the storage and spatial distribution of SOC in peninsular Spain. This information could be used as a national benchmark in estimating and managing carbon sinks, as well as in the national GHG Emissions Inventory Report. To achieve this aim (a) accurate consistent and standardized information and methodologies must be applied to (b) generate cost-effective, high-resolution maps of SOC distribution and its associated spatially explicit uncertainty, ensuring interoperability of the results from national to global scales.

As final products (deliverables) four national SOC maps with a 90-m spatial resolution have been generated with their associated uncertainties. These maps show the spatial estimation of the SOC concentration (SOCc) at depths of 0-30 cm and 30-100 cm, and SOC stock (SOCs) at 0-30 cm and at its effective depth.

5.2. MATERIALS AND METHODS

The general methodology used in this study (Figure 5.1) is divided into three main stages:

a) Acquisition and analysis of the soil profile and environmental covariate database
b) Modeling of the SOC spatial estimate
c) SOC spatial prediction and mapping

The final maps were generated by DSM based on the SCORPAN (Florinsky, 2012) conceptual spatial inference model. The SOC point prediction (target variable) on a continuous surface was based on relationships with the selected soil forming factors (predictor covariates) in a given location. Modeling of data was based on three supervised learning approaches. quantile regression forest, ensemble machine learning and auto-machine learning**.** For the final SOC Spatial distribution maps, each pixel was assigned the prediction from the most accurate model, i.e., lowest uncertainty.

5.2.1. Study Area

The study area was peninsular Spain (491,258 Km^2) (Figure 5.2). The complex orography is characterized by abrupt mountain systems, high plateaus and wide watershed depressions. Most of the peninsula is dominated by the large Central Plateau, with a height of 600 to 760 m asl and slightly inclined to the west, therefore most water courses discharge into the Atlantic ocean. In the interior and surrounding the plateau are gentle hills and high, steep mountain chains in which the highest peak is 3,478 m asl (Serrano, 2000).

Spain is characterized as one of the most varied climate country, from humid to semiarid (AEMET IPMA, 2011). Mean annual precipitation varies from less than <200 mm (in the southeast) to >2200 mm (in the north and mountainous zones), the mean annual temperature ranges from <2.5ºC at high altitudes to >17.5ºC, especially in the south and southeast. The dominant climate is Mediterranean, extending over the interior plateaus (continental Mediterranean) to the coast (coastal Mediterranean). It is characterized by rainy, cold-to-mild winters and dry, hot or mild summers with variable temperatures and precipitation in autumn and spring. This contrast is smoother on the coast, becoming predominantly arid or subdesert in the southeast. In the north and northwest, the climate is predominantly oceanic, very humid and with mild temperatures.

The complex topography and wide climatic variety have led to a mosaic of land covers and uses. Although the agricultural area in Spain has been decreasing in the last 20 years, it still occupies about 33% of the total area (MAPA, 2021). It is characterized by heterogeneous agricultural systems including herbaceous crops, mainly dryland cereals, and woody crops (grapes, olives, almonds and fruit). The forested area occupies over 59% of the peninsula with predominantly natural forest and plantations, mainly in mountain systems in humid and subhumid regions, and shrubland. Grasslands and other herbaceous vegetation are extensive, especially on high mountains and in the north. Wetlands and water surfaces comprise 0.9% of Spain, and artificial surfaces are 7.1% of the total area.

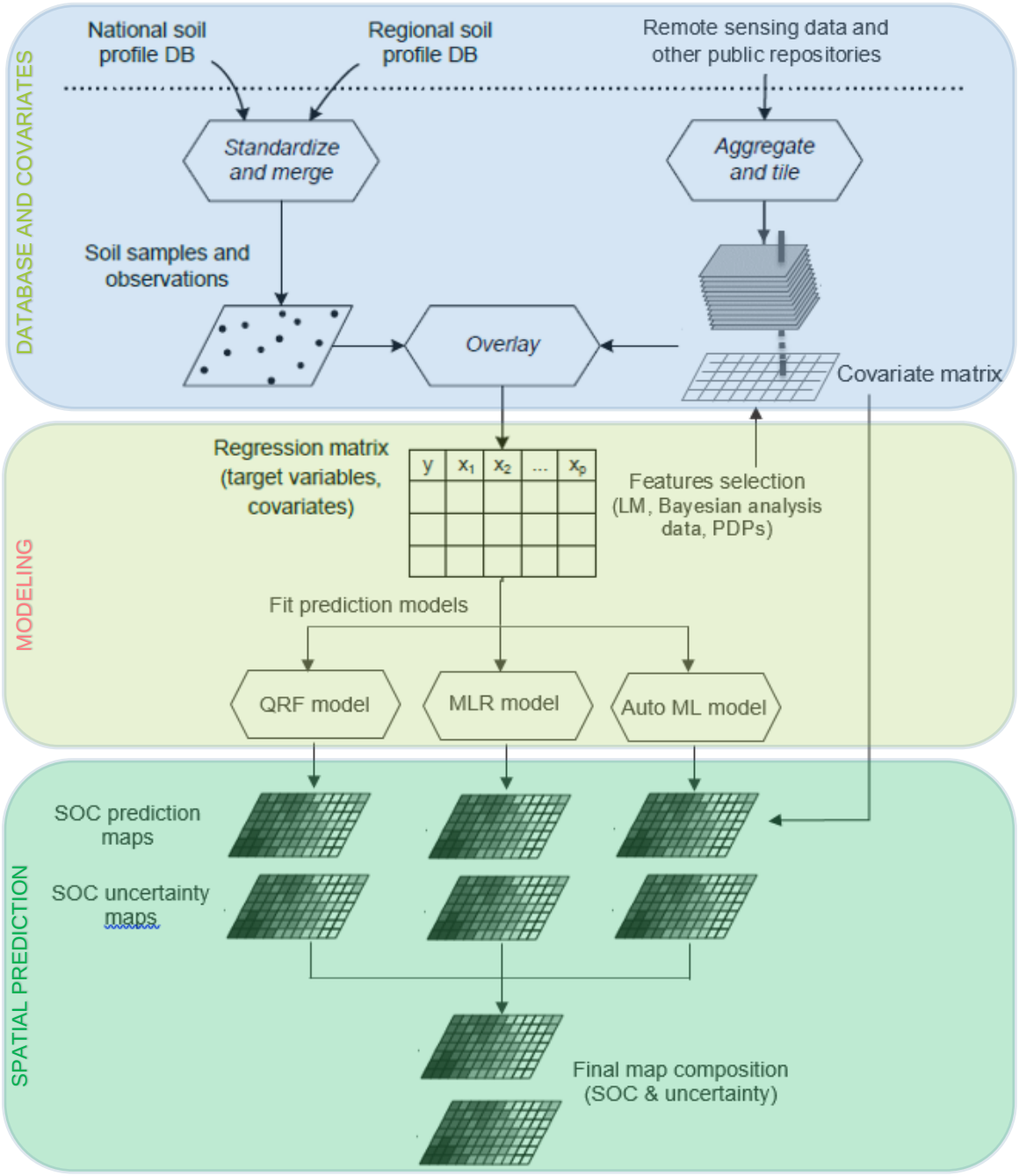


Figure 5.1 General methodological outline adapted from the World Soil Information Service (WoSIS, Hengl et al., 2017)

The Iberian massif, which occupies the western half of the peninsula, is made up mainly of siliceous metamorphic rock (mainly slates, schists and quartzites) and intrusive rocks (mostly granites). The Central Plateau is composed of two large groups of lithological material: sedimentary rock (Mezozoic), mainly calcareous or calcareous cement (limestones, marls, sandstones and conglomerates), and clayey to sandy sediment (Tertiary and Quaternary). In the mountain chains surrounding the Central Plateau, Pyrenees and Betic System, calcareous materials predominate. The wide alluvial plains (Ebro and Guadalquivir) are Tertiary and Quaternary sediments (SGE-IGME, 2004).

Soils in peninsular Spain are predominantly Calcisols. These soils formed on calcareous parent material are usually present in areas with water stress. Because of their high rate of mineralization, they are vulnerable to decrease in organic matter. In the north, they are predominantly Umbrisols, which are typical of cold humid zones with hardly any or no water deficit and rich in organic matter. The south is dominated by Regosols, i.e., little developed mineral soils frequent in eroded areas, particularly in arid and semiarid areas and mountains. Luvisols, Rendzinas or Leptosols are frequently found along with the Regosols. Fluvisols and Vertisols, predominant in fluvial areas, are characterized by a large proportion of swellable clay (De la Rosa et al., 2001).

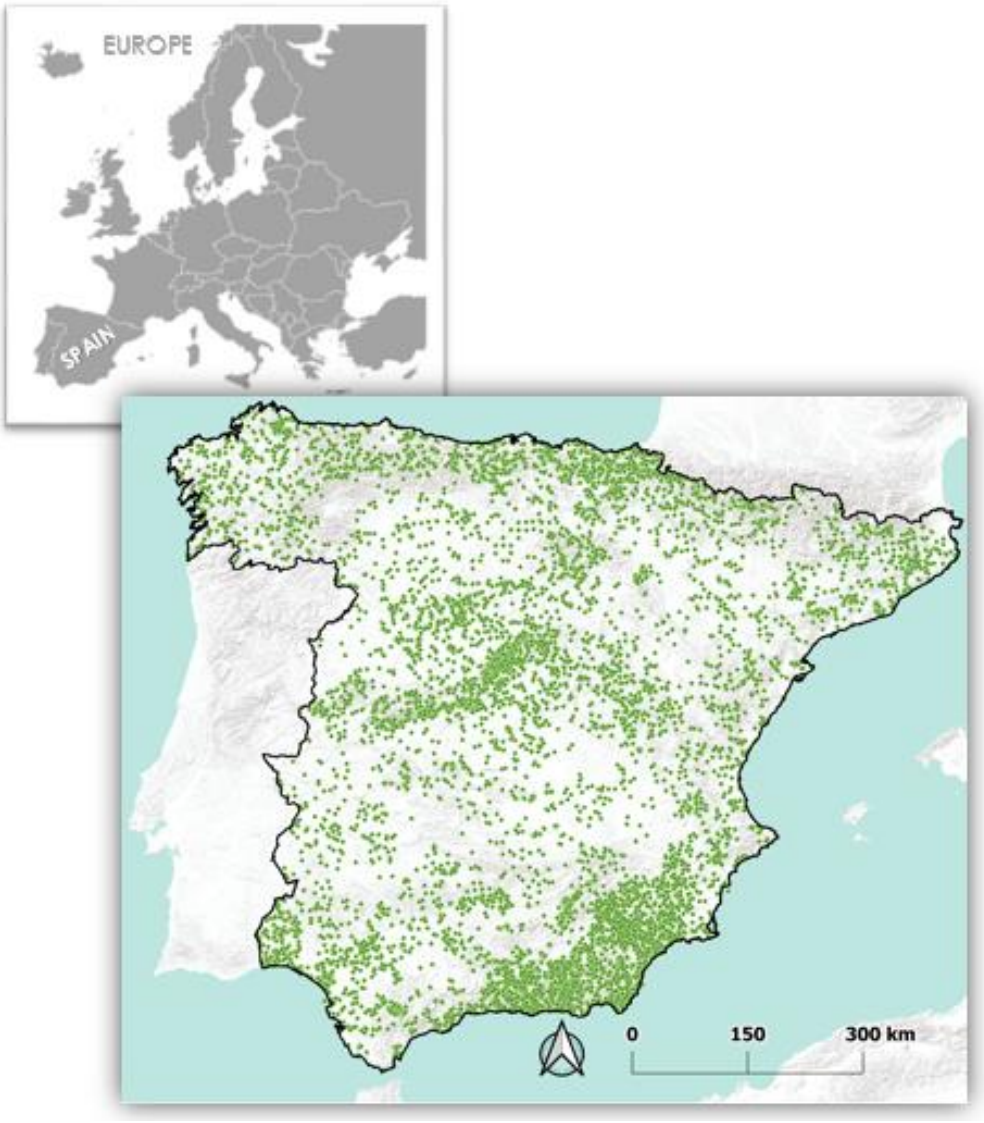


Figure 5.2 Location of the study area and soil profiles

5.2.2. Database

The soil profile database used in this study has 8,361 georeferenced profiles (Figure 5.2) with 27,931 pedogenetic soil horizons. The soils were described and sampled following

pedomorphogenetic criteria in the various versions of the FAO Guidelines for Soil Description Profile (Jahn et al., 2006).

The database was built from public domain data or facilitated by institutions responsible for the information. The main source for profiles was the Red Carbosol database (78% of the samples), a collaborative Spanish network of soil experts at various research centers and universities who have compiled information on profiles from 635 different sources (Llorente et al., 2018). The second source of information (18% of the profiles) was derived from different cartographic projects provided by the CMAyOT (Junta de Andalucía). The rest of the profiles (4% of data) were from the LUCDEME database from 'Región de Murcia' (Alias and Ortiz, 1986), Agrarian Technological Institute (Junta de Castilla y León) and from the University of Castilla La-Mancha (Bravo et al., 2019). The sampling periods range from 1954 to 2018, being mostly from 1965 to 2000.

The structure of these data was modified to generate a common database of standardized reconciliated information on soil properties and their measurement units. Profile inclusion in the final database considered basic quality criteria in accord with our aims and based on quantitative pedological routines (Beaudette et al., 2013). These criteria were correct georeferencing, duplicate information, missing data values, error in the information's agreement with the horizon, and format error. All fields in the database were checked using basic descriptive statistics, such as minimum, maximum, average, and standard deviation values to check the consistency of the data and the behavior of the soil properties (e.g., extreme values, pH range, CN ratio). The inconsistent data were reclassified as "no data". The information on profiles and horizons included in the final database are shown in Table 5.1

As SOC tends to have a log-normal distribution (Yigini et al., 2018), to determine the SOC spatial distribution, the original SOC data were log-transformed to test the increase in correlation between SOC and its prediction factors.

As the purpose of modeling was to estimate the organic carbon in mineral soils, all the profiles or horizons with over 20% organic matter were excluded from the database according to WRB criteria (IUSS Working Group WRB, 2015). Thus, Histosol profiles and horizons composed of organic materials (H, O, L) were eliminated.

Table 5.1. Soil information in the final database used in the study.

Code	Property	Units	Description
		Profile information	
Source	Source of data		Original repository of the database
Province	Province		
Location	Location		
Latitude	Latitude	degree	Latitude WGS84 projection (EPSG:4326)
Longitude	Longitude	degree	Longitude WGS84 projection (EPSG:4326)
UTM east	UTM X coordinates	m	X coordinate ETRS89 / UTM zone 30N projection (EPSG:25830)
UTM north	UTM Y coordinates	m	Y coordinate ETRS89 / UTM zone 30N projection (EPSG:25830)
Soil_WRB	Soil classification into WRB system		According the classification system used in the original bibliographic source
Soil Taxonomy_USDA	Soil classification into USDA system		According the classification system used in the original bibliographic source
		Horizon information	
Id_Horiz	Unique identification number of horizon		
Id_Profile	Unique identification number of profile		
		Morphological properties	
Horizon_Main	Horizon main reference		
Horizon	Horizon reference		
Position Horizon	Horizon relative position in the soil profile		I: initial position; M: medium position; F: final position
Depth top	Upper limit	m	
Depth bot	Lower limit	m	
Depth med	depth from the surface to the centre of the horizon	m	
Thn	effective depth	m	
Color HLS	Horizon color		Munsell Code
		Physical properties	
Density	Bulk density	g/cm3	Wide variety of methods, including application of pedotransfer functions. Block Method is the most widely used
Litho	Coarse material	%	> 2 mm; % of total volume
Sand	Sand	%	USDA System
Silt	Silt	%	USDA System
Clay	Clay	%	USDA System

Table 5.1. (cont.) Soil information in the final database used in the study.

Code	Property	Units	Description
		General chemical properties	
OM	Organic matter	%	Conversion factor from TOC: 1.724
TOC	Total Organic Carbon	g/kg	Walkley-Black Method or Dry Combustion Methods (elemental analyzer)
pH	pH in water		Soil:water ratio 1:1, 1:2.5 or 1:5
Carb	Carbonate content	%	Bernard's Calcimeter or Titrimetry
C/N	Carbon and nitrogen ratio		Direct calculation, from TOC and N data
TN	Total Nitrogen	g/kg	Kjeldahl Method or Dry Combustion Methods (elemental analyzer)
P	Phosphorus	mg/kg	Olsen or Mehlich Methods
K	Potasium	mg/kg	Extraction with ammonium acetate and quantification by ICP spectroscopy or Flame Photometry Methods
Ca	Calcium	mg/kg	Extraction with ammonium acetate and quantification by ICP spectroscopy or Atomic Absorption Spectrometry Methods
Mg	Magnesium	mg/kg	Extraction with ammonium acetate and quantification by ICP spectroscopy or Atomic Absorption Spectrometry Methods
Na	Sodium	mg/kg	Extraction with ammonium acetate and quantification by ICP spectroscopy or Flame Photometry Methods
CEC	Cation-exchange capacity	cmol/Kg	Cation Summation Method
EC	Electric Conductivity	dS/m	Soil:water ratio 1:5
Gp	Gypsum content	%	Conductometry in the water extract

5.2.2.1. Standard Soil Depths and Calculation of SOC Concentration and SOC Stock

To facilitate interoperability of the results, the variable depths of the morphological horizons were standardized and discretized. New limits were set applying the commonly applied worldwide criteria for SOC estimation (Brus et al., 2017).

The information in the database on soil organic carbon concentration (SOCc) was measured analytically, and therefore, that data was used directly in the estimation. Some data were converted into SOCc from organic matter values, using 0.67 as the conversion factor, which assumes 58% C in OM (Rosell et al., 2001). SOCc values were discretized in two standard depths of 0-30 cm and 30-100 cm. The final value assigned to each standard depth was the sum of the SOCc morphological horizons, weighted by their original depth.

The SOC stock (SOCs) in the profile, defined as the total amount of soil carbon at its effective depth per unit area, was calculated using equation (1). The effective soil depth (ESD) was

considered the *solum*, which includes surface and subsurface horizons with presence of roots and biological activity (Soil Survey Staff, 1975).

$$SOCs\ (Kgm^2) = \sum_{i=1}^{n} SOC\ (g/kg)_i \cdot BD(Kg \cdot m^3)_i \cdot \left[1 - \left(\frac{CRFVOL}{100}\right)\right]_i \cdot HSIZE(cm)_i \quad (1)$$

where 'i' is the horizon and 'n' the total number of horizons in the soil profile, BD is bulk density, CRFVOL is the percentage of coarse fragments (over 2 mm in diameter), and HSIZE is the horizon thickness.

Following the principle of interoperability abovementioned, the SOCs was also calculated for the standard depth of 0-30 cm by the Equation 0, prior to standardization to 0-30 cm.

Based on the results of the previous chapte 4r, to avoid propagating errors in the estimation of parameters absent in some profiles, stock was calculated only for those profiles with information available on apparent density and coarse fragment content.

The final number of profiles for modeling carbon as a function of depth was distributed as follows:

- For estimating SOCc (g/kg) at a standard depth of 0-30 cm, 8332 profiles.
- For estimating SOCc (g/kg) at a standard depth of 30-100 cm, 6947 profiles.
- For estimating SOCs (tC/ha) at a standard depth of 0-30 cm, 1475 profiles.
- For estimating SOCs (tC/ha) at its effective depth, 1499 profiles.

The main SOC characteristics in peninsular Spain were calculated using basic descriptive statistics of the variables related to carbon, such as density distribution, mean profiles and spatial autocorrelation.

The data processing and spatial analysis were developed using the open-source software of QGIS and R (main "aqp" "Performance Analytics", "GSIF" and "gstat" packages).

5.2.2.2. Representativeness of the Database

The spatial representativeness of the database was developed by the probability distribution based on the maximum entropy method (Phillips et al., 2006). Maximum entropy (Maxent) has been mainly used in modeling predictive species distribution maps. However, it is widely used for estimating the relationship between spatial observations (i.e., soil samples) and environmental properties (i.e., soil-forming factors) for a specific area (Villarreal et al., 2018). The "logistic" predictive model was selected to perform the probability distribution based on environmental variables to define different representative areas. These areas were defined by the most relevant environmental covariates in modeling SOC, according to the results of the covariate selection methods (see section 5.2.4.2).

The predictive capability of the resulting model was tested with the Area Under the Curve (AUC) of the training data. This value was compared with the AUC expected for a random model (0.5). Values lower than 0.5 would mean that the model prediction was worse than a random estimation (Fielding and Bell, 1997). The output resulted in a map with color gradients showing the similarity of the environmental predictor variables in each pixel and minimizing the relative entropy between them with regard to the pixels with the located soil

samples (Elith et al., 2011). This approach expressed the information on the spatial representativeness of the soil samples for different types of environmental factors in peninsular Spain.

5.2.3. Environmental Covariates

Following our methodological outline (Figure 5.1), the environmental variables to be included in the SOC model were identified to be applied into the SCORPAN conceptual spatial inference model (McBratney et al., 2003). According to this conceptual model, soil property predictions were based on their forming factors categorized in seven environmental factors: Soil properties (S), climatic variables (C), biota or organisms (O), relief (R), parent material (P), time, age (A), and spatial location (N).

5.2.3.1. Compilation of Variables

Based on the SCORPAN categorization of the forming factors, the environmental variables stacked and included in the SOC spatial prediction were:

- Static variables:
  - Relief factor. Geomorphometric variables of the terrain. the topographic relief was assessed by geomorphometry and feature extraction derived from the Geomorpho90m global dataset, at 90 m resolution under the WGS84 geodetic datum. The fully-standardized 26 geomorphometric variables in the dataset, derived from the MERIT-Digital Elevation Model (DEM), consist of layers that describe the (i) rate of change across the elevation gradient, using first and second derivatives, (ii) ruggedness, and (iii) geomorphological forms (Amatulli et al., 2020).

    The source and resolution of these products are described in Table 5.2.
  - Human factor. Land cover and uses (IGN, 2012) reclassified into 13 classes.
  - Parent Material Factor. Lithological classes from the lithological map of Spain 1M, with 22 hierarchical levels (IGME, 1995).
  - Soil properties. Digital district soil atlas (USDA, 1987). Soil map digitalized based on the 1:2.000.000 map in the Atlas Nacional de España (Soil science) published by the CSIC/IRNAS (De la Rosa et al., 2001).
- Dynamic variables:
  - Climate Factor. Precipitation and temperature climatic variables (Ninyerola et al., 2005).
  - Biota Factor. Vegetation indices and ecosystem functional attributes derived from remote sensing. The indices were related to carbon cycle, water cycle and energy or heat balance. The functional attributes were related to the amount, seasonality and phenology of each index.

    These indices were calculated from satellite products with a long enough time series to ensure their stability over time (2000-2019). The complete remote sensing dataset included 172 ecosystem functioning indices as candidate predictors. The sources and resolution of these satellite products are described in Table 5.2.

Google Earth Engine (Gorelick et al., 2017) was used to derive the interannual mean and monthly mean of the following summary metrics of their seasonal dynamics: annual mean (surrogate for annual total amount), annual maximum and minimum (indicators of the annual extremes), seasonal standard deviation (descriptor of seasonality), and sine and cosine of the dates of maximum and minimum (indicators of phenology) (Alcaraz-Segura et al., 2017).

Finally, a total of 254 environmental covariates describing the forming factors in the peninsular territory were estimated as a prior step to SOC spatial modeling (Table 5.2).

Table 5.2. Description of candidate predictors for spatial modeling of soil organic carbon (SOC).

Category	Variables	Source	Scale/ Resolution
Topographic(1)			
	(i) Slope, Aspect, Aspect cosine, Aspect sine , Eastness, Northness, Convergence, Compound topographic index, Stream power index, East-West first order partial derivative, North-South first order partial derivative (ii) Profile curvature, Tangential curvature, East-West second order partial derivative, North-South second order partial derivative, Second order partial derivative (iii) Elevation standard deviation, Terrain ruggedness index, Roughness, Vector ruggedness measure, Topographic position index, Maximum multiscale deviation (dev-magnitude), Scale of the maximum multiscale deviation, Maximum multiscale roughness, Scale of the maximum multiscale roughness (iv) Geomorphon	Geomorpho90m (Amatulli et al. 2020)	90m
Climate			
	Mean annual precipitation (mm). Mean, minimum and maximum annual temperature (ºC). Radiation (kW/m^2). Period 1951-1999.	University of Barcelona (Ninyerola et al., 2005);	200 m
Land features			
	Soil types	Proyecto SEIS.net (MIMAM- CSIC)	1:100000
	Lithology	IGME (Spain)	1:200000
	Land use/cover	IGN-Corine Land Cover	1:100000
Satellite-indices			
Carbon cycle	Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI	MOD13Q1	250 m
Water cycle	Precipitation, Normalized Difference Water Index (NDWI) Evapotranspiration (ET)	CHIRPS MCD43A4	1km 500 m
Radiative balance	Albedo	MCD43B3	500 m
Sensible heat	Land Surface Temperature (LST)	MOD11A2	1 km
Ecosystem Functional attributes (inter-annual and monthly mean):		(Alcaraz et al, 2017)	
Amount: mean, maximum, minimum **Seasonality**: *standard deviation, coefficient of variation, range, relative range* **Phenology**: *sine and cosine of the dates of maximum and minimum*			

(1) Order topographic variables: (i) First order derivative, (ii) Second order derivative, (iii) Ruggedness, (iv) Geomorphological forms.

5.2.3.2. Covariate Matrix

The environmental covariates were stacked in a covariate space, or covariate matrix, with spatially explicit information. The different formats of the coordinate reference systems (CRS) and spatial resolutions of the covariates were processed and harmonized. An alternative to reprojection and rescaling of the covariates was used to decrease geometric distortion and avoid computational limitations due to the large amount of data (Bauer-Marschallinger et al., 2014). A reference matrix was generated based on the pixel center locations (x,y) of the most detailed resolution layer, i.e., MERIT-DEM (90 m), and WGS84 (EPSG 4326) as the CRS. The covariate data were extracted for each reference matrix location using geoprocessing techniques, prior to reprojecting the reference matrix coordinates to the CRS of each covariate, when applicable. The resulting covariate matrix consisted of the value of each covariate (column) extracted for each point in the study area at a distance of 90 m (row). This matrix was organized into a tiling system for faster computing processing time in geoprocessing analyses.

The categorical variables were previously rasterized, considering only those categories with a sufficiently representative number of soil samples (i.e., over 100 data). These categories were transformed into a binomial variable (0-1) or "dummy" indicating the presence or absence of the specific category as specified in Yigini et al. (2018).

R software was used to generate these matrices, in particular "sp", "rgdal" and "raster" packages.

5.2.4. Modeling of Soil Organic Carbon

Following the stages of our methodological outline (Figure 5.1), details of the SOC spatial estimation modeling and model performance methodologies are described below.

5.2.4.1. Regression Matrix

SOC and covariates data were combined in a regression matrix. This matrix included the carbon database (i.e., SOCc, SOCs and their corresponding log transformations) as well as the covariate point data, extracted at the same location as the profiles. As the covariance matrix, the coordinates of the SOC profiles were reprojected to the CRS of each covariate, if applicable, to decrease geometric distortions. This structure facilitated analysis of the relationship between carbon and the covariates.

5.2.4.2. Feature Selection

Due to the large number of covariates generated (254), their variable importance (VI) was analyzed to gain insight into the behavior of soil carbon data and interpretability of covariates influencing soil carbon models. The selection of the most relevant covariates, also called feature selection, does not compromise model accuracy, and is useful for avoiding the risk of statistical redundancy leading to potential overfitting (Gregorutti et al., 2017). This variable importance approach, based on the relationship between soil carbon data and covariates, was three-fold:

a) **Multiple linear regression.** In this simple and easily interpretable approach, the dependent variable is modeled as a linear function of a set of regression parameters and a random error term, which represents variation in the dependent variable unexplained by the function of the dependent variables and coefficients. Thus, two major assumptions were made when applying this approach: a) the linear relationship between the response variable (SOC) and the predictors (environmental covariates), and b) multicollinearity of the predictors (Yan and Su, 2009).

The absolute value of the t-statistic was used as a measure of VI for each model parameter. The t-statistic is the coefficient associated with the predictors divided by its standard error, and is therefore considered a measure of predictor accuracy (Greenwell and Boehmke, 2020). It was performed with the 'vip' package in R software.

b) **Bayesian analysis.** The Bayesian inferential approach is based on the probability distribution of the parameters derived from observed data and further information available. The main difference from multilinear models is that in the Bayesian probability model, the parameters are considered unknown, and both the data and prior information are combined with a likelihood function in a new posterior distribution or prediction (Gelman et al., 2013). The probability model was fitted to data by specifying magnitude and different data assumptions (SOC, log SOC, covariates and scaled covariates) and prior data (mean and standard deviation).

The models were fitted for a normal distribution using the Markov Chain Monte Carlo (MCMC) methodology, which facilitates the implementation of Bayesian analysis of complex datasets. The resulting estimates are known as the posterior distribution, which is conditional on the data. This conditioning is governed by the rules of probability theory, which define a uniquely logical posterior for every prior, likelihood, and data (McElreath, 2018). Based on the best combination of the distributional assumptions, the model was built with an iterative selection of the most significant covariates. The final covariate selection depended on the model with the best information criteria based on the lowest Watanabe–Akaike information criteria (WAIC), effective number of simulated samples (>= 300), fewest covariates and lowest standard error.

We implemented the Bayesian approach in the 'rethinking' package in R.

c) **Projection pursuit regression (PPR**) **model using partial dependence plots (PDPs)** for assessing the feature effects. PPR is a statistical modeling technique that can ignore variables with low explanatory power thus is useful for determining variable importance or feature selection. The PPR method is based on linear combinations of non-parametric functions of the explanatory variables and was applied to explore nonlinear relationships in the data (Friedman and Stuetzle, 1981). The PDPs is a technique for constructing variable importance plots. It is I based on assessing the "flatness" of the PDP of each covariate to graphically display the effect of the feature space on the estimated prediction. PDPs not only visualize this relationship while accounting for the average effect of the other predictors in the model, but also ranks and scores the predictors in terms of their relative influence on the predicted outcome. Then PDP-based VI scores coefficients were computed based on the fitted PPR model. These VI scores captured the variability in the partial dependence

values for each main effect by computing the standard deviation of the y-axis values for each PDP. For the final selection, we ranked the covariate VI scores in terms of their relative influence on the SOC data.

We fitted the projection pursuit regression method and construct PDPs for each covariates using the 'stats' and 'pdp' packages in R.

The final covariate dataset was selected by the highest VI scores. The covariates common to the three feature selection techniques were prioritized to reduce the dataset dimension. This final covariate selection was supervised by expert criteria.

This procedure was performed for each SOC variable (SOCc and SOCs) at the 0-30 cm and 30-100 cm standard depths and effective depth, as appropriate. This analysis enabled a considerable reduction in the dimensions of the regression covariance matrix used in the predictive models.

5.2.4.3. Predictive Models

There is a wide range of optimal SOC predictor algorithms for digital soil mapping (Guevara et al., 2018). Furthermore, considering the complex and frequently non-linear relationship between SOC and environmental variables, multi-model ensemble methods using a machine learning (ML) algorithm to predict the SOC spatial variability were combined with its associated uncertainty (Shangguan et al., 2017; Wang et al., 2018a). Ensemble learning is a machine learning branch that combines several base ML models (whether homogeneous or heterogeneous) to achieve better predictive performance than individual ML models by reducing the noise or error between observed and predicted data (Zhang and Ma, 2012).

Ensemble methods are usually grouped into bootstrap aggregating (bagging), boosting, and stacking methods. All three categories attempt to tune their predictions to the observations by decreasing model variance, bias, or both simultaneously. The main difference is that bagging and boosting usually work with homogeneous models, whereas stacking excels in combining heterogeneous models (Nguyen et al., 2021). We used the following three modeling approaches to predict SOC:

– ***Quantile regression forest (QRF***): In view of the results of chapter 4, we considered the QRF technique based purely on regression, i.e., no kriging interpolation of residuals, a good performance model for mapping SOC with sparse legacy data, as is the case in this study. The bagged ensemble of 'QRF' decision tree is an extension of random forest (RF) used for accurate estimation of the full conditional distribution of the response variable for each pixel value.

The QRF algorithm was used without kriging interpolation of residuals as implemented in the quantregForest (QRF) package in R software (Meinshausen, 2006). The QRF validation statistics were calculated from out-of-bag error.

– ***Ensemble Machine Learning (MLR)***: For this predictive ensemble approach, linear model regression (lm) predictors were combined with non-parametric models based on a bagging and boosting algorithm. A stacked learner was fitted using predictions from five individual learners: lm, random forest (RF), deep learning (DL), cubist (Cb) and weighted k-

nearest neighbor classifier (kknn). A super learner was generated using the base learner predictions as features by the compress method, i.e., training a neural network to compress the model from the collection of base learners. An independent assessment of all individual learners was performed by ten-fold cross validation. Spatial partition in cross-validation by k-means clustering was used to partition the dataset into equally sized, spatially disjointed subsets by two classification layers as a spatial component: xy locations of soil data and the Köppen climate classification (Kottek et al., 2006).

The ensemble was implemented with the MLR (Machine Learning in R) package.

– ***Auto-machine learning (AutoML):*** AutoML modeling ensembles of a large number of models by efficiently automating learning algorithms and featurization, and optimizing computing power and time. Given the huge and growing variety of ML algorithms, their selection and hyperparameter tuning have become challenging tasks. We used the H2O package for R, which trains a variety of algorithms, such as generalized linear models (GLMs), distributed RF, deep neural networks, XGBoost or gradient boosting machines (GBMs), among others. The H2O AutoML algorithm relies on efficient training of high-quality algorithms to produce a large number of models in a short amount of time, yielding diversity across candidate models, which can be exploited by stacked ensembles to produce a powerful final model (LeDell and Poirier, 2020). This process was highly scalable and fully-automated. The resulting AutoML object included a "leaderboard" of models that were trained in the process, including 10-fold cross-validated model performance. The models were ranked by root-mean-square error (RMSE).

The information criteria for assessing the fit of the three models were the determination coefficient (R^2), the root-mean-square error (RMSE) and mean absolute error (MAE). Each model was evaluated by comparing the SOC values performed with 75% of the dataset, with a validation dataset (remaining 25%). We repeated this process three times for each model approach to obtain the mean of the model evaluation statistics (i.e., concordance correlation coefficient (CC, Lin, 1989), normalized RMSE).

For the log SOC estimates, the model diagnostics and prediction values were back-transformed into their original units.

5.2.5. Spatial Prediction and Final Map Compilation

The last stage of the study methodological outline (Figure 5.1) refers to the spatial prediction of SOC and generation of the final maps.

5.2.5.1. Spatial Inference and Uncertainty of SOC

For the carbon variable (SOCc and SOCs), the model predictions for each carbon data profile were extrapolated to the full extent of the study area by the covariance matrix, generating three prediction maps, one for each model approach (*QRF, MLR, AutoML*).

The three different prediction maps were generated from the model approach by randomly partitioning the dataset (75% training and 25% validation) from three simulations of the model. The standard deviation of the resulting maps was calculated for each pixel to estimate model precision. This standard deviation map was interpreted as a surrogate of the spatial uncertainty associated with the SOC estimates for each pixel value. This procedure was replicated for each model approach (*QRF, MLR, AutoML*).

The SOC spatial prediction maps contributed to the final 90-m spatial resolution map for peninsular Spain. This final map was generated by assigning each pixel the prediction value of the most accurate map, i.e., the one with the lowest uncertainty value for the given pixel (Figure 5.3).

This workflow was repeated for SOCc at the 0-30 cm and 30-100 cm depths and for SOCs at 0-30 cm and the effective depth. The result was eight maps corresponding to the mean SOC spatial prediction at the two standard depths, and their respective associated uncertainty maps as surrogates for the most probable interval of carbon values.

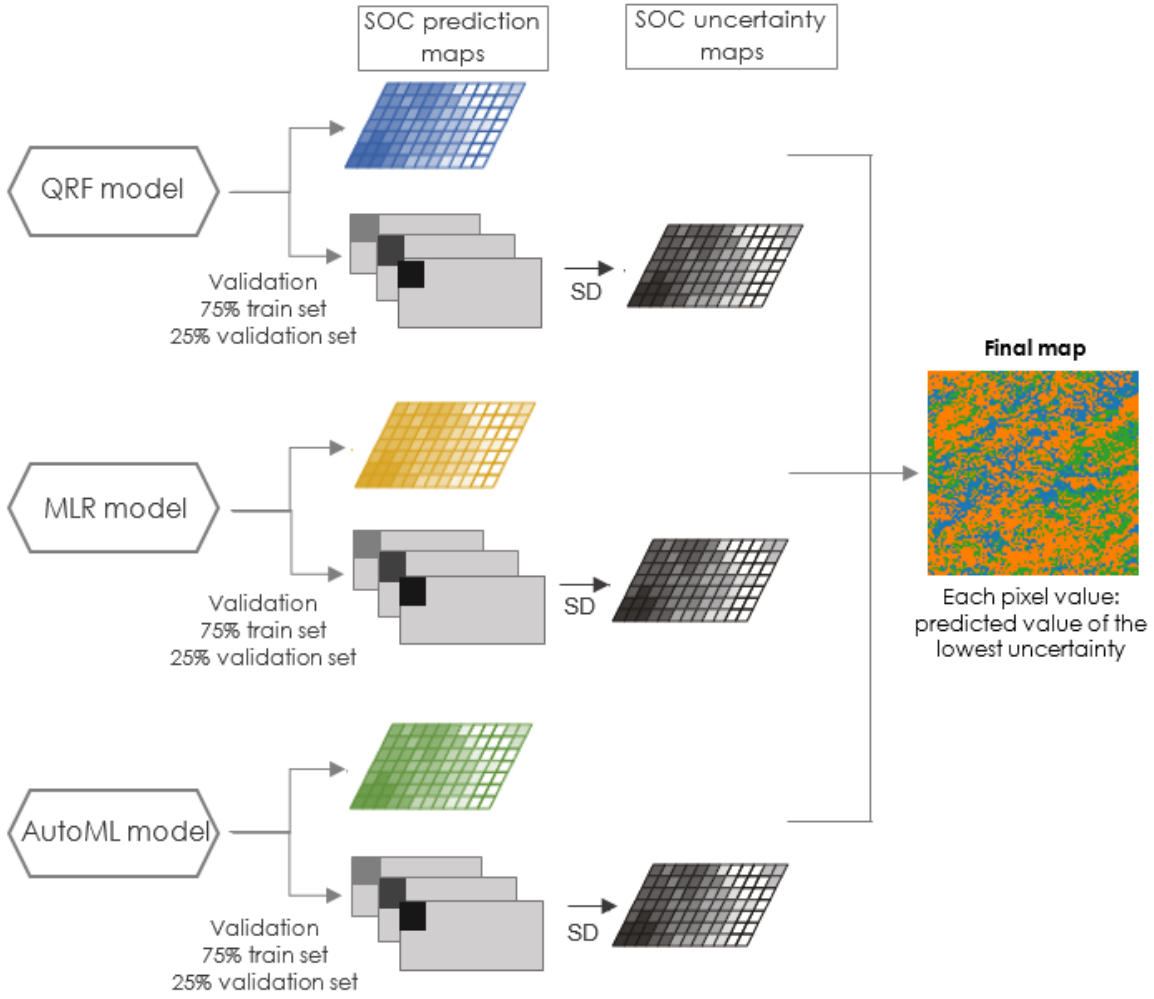


Figure 5.3 Soil organic carbon (SOC) spatial prediction mapping design. SD= standard deviation

5.2.5.2. Model Accuracy Assessment

As mentioned above, there is no perfect model on digital SOC mapping, especially for extensive areas and using disperse soil profiles from a legacy database, collected over long periods of time. Thus, different approaches were used for spatial prediction of SOC and different model evaluation metrics were considered to evaluate and compare the

accuracy of the resulting models. The metrics are based on comparison of observed and predicted data for the validation set in each model (QRF, MLR, AutoML) as Gray et al. (2015) proposed. Specifically, the four statistical indices used for this were: (a) the coefficient of determination (R^2), which measures the percentage of variation explained by each model; (b) the mean absolute error (MAE), which indicates how close the prediction is to the observation; (c) the root mean square error (RMSE), which measures the overall accuracy of the prediction; and (d) the CC, which measures concordance level between predicted and observed values.

The conditional quantiles were plotted to verify the predictive to verify the effectiveness of the models across the full distribution of SOC observed values (Wilks, 2019). The conditional quantile plot splits the predicted value evenly and identifies the corresponding values of the observations, and the median, 25/75th and 10/90 percentile (quantile) calculated. The data are plotted to show how well predictions agree with observations across the full distribution of values.

The 'openair' package in R was used to perform the conditional quantiles plots.

5.3. RESULTS

5.3.1. Database

5.3.1.1. Input Soil Database Descriptive Statistics

After harmonization data and screening for errors, the database included 8,332 soil profiles distributed in 25,370 morphological horizons distributed throughout peninsular Spain (Figure 5.2).

Only profiles with complete information were included in the stock calculation to avoid error propagation. The remaining number of profiles was 1,499, slightly more than the 1,475 selected for stock estimation at the standard depth of 0-30 cm.

SOCc content in the morphological horizons was highly variable, from 0.02 g/kg to 296.9 g/kg, and a mean of 16.53 g/kg. The mean SOCc in the standard 0-30 cm depth interval was 20.7 g/Kg, and in 30-100 cm, it was 5.8 g/kg. The latter represents 35% of the profile mean. Similarly, SOCs calculated in the profile varied from 0.006 kg/m^2 to 87.1478 kg/m^2 with a mean of 2.965 kg/m^2. In this case, the carbon stock in the upper 30 cm was 65% of the profile mean (Table 5.3).

Table 5.3. Statistical summary of soil organic carbon concentration (SOCc) and and Soil Organic Carbon stock (SOCs) at different standard depths.

Variable[(1)]	Depth	Number of profile	Minimum	1st Quantile	Median	Mean	3rd Quantile	Maximum
SOCc	0-30	8,332	0.017	7.148	14.008	20.691	27.098	257.95
(g/kg)	30-100	6,947	0.017	1.700	3.371	5.833	6.814	185.743
SOCs	0-30	1,475	0.119	2.066	3.738	5.31	6.95	39.967
(kg/m^2)	ESD[(2)]	1,499	0.119	3.000	5.300	8.198	10.260	93.892

[(1)] SOC=Soil Organic Carbon concentration; c=concentration, s=stock. [(2)] ESD= Effective soil depth.

The statistical distribution of SOCc and SOCc showed a normal distribution with a right-skew once their original values had been transformed to a natural log (Figure 5.4).

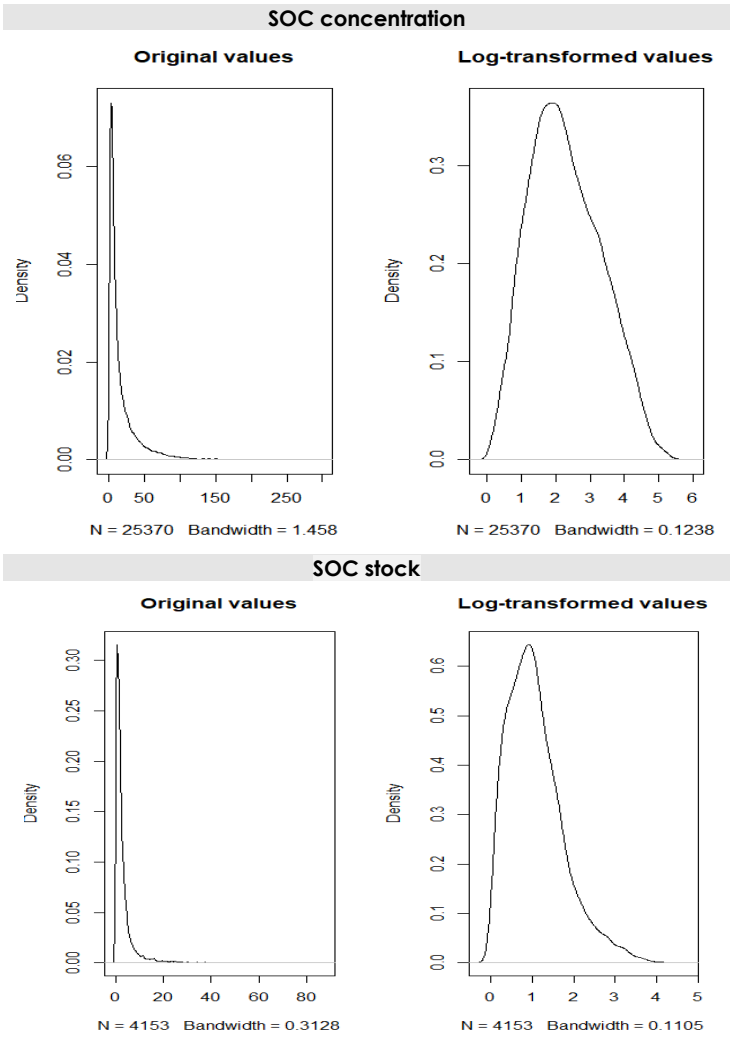


Figure 5.4 Statistical distribution of original soil organic carbon (SOC) data versus their log-transforms.

The statistical description of the SOCc average profile (Figure 5.5) showed most of the SOC in upper 20-30 cm layers, decreasing rapidly with depth. The mean profile values ranged from 23 g/kg in the upper horizon (0-5 cm) to 3 g/kg in the deepest horizon (>200 m).

This analysis demonstrated the shallow depth of peninsular soils, where only 35% of the horizons were over 100 cm, decreasing to 3% at >150 cm.

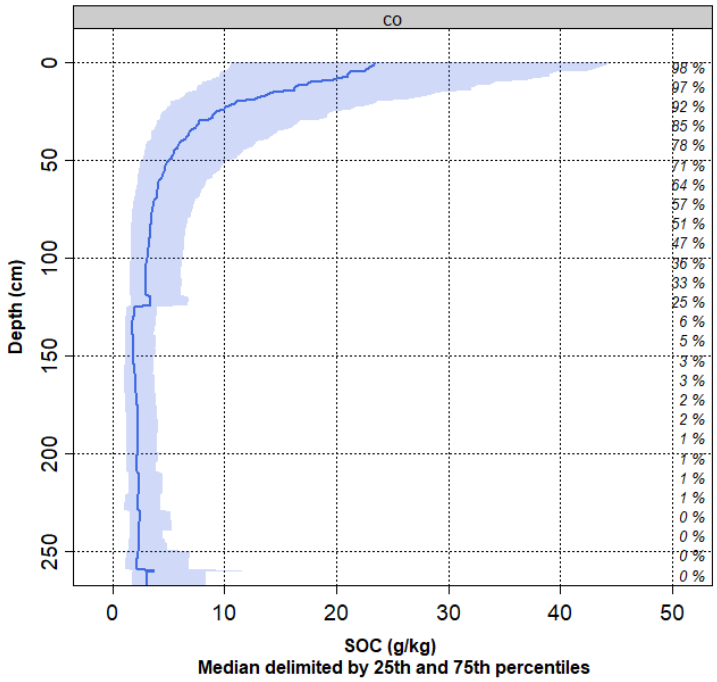


Figure 5.5 Soil organic carbon (SOC) concentration (g/kg) average profile at different depths.

SOC spatial autocorrelation was evaluated using the ratio of the nugget effect to the total sill called the nugget-to-sill ratio (NSR). The analysis of the SOCc log variable, which presented a normal distribution, showed NSR >75%. This value indicated a weak spatial structure (Cambardella et al., 1994). In addition, the large nugget effect (nugget = 1.3) could mean that a large proportion of fine-scale SOCc variation was not captured by the data. Despite the lack of evidence of autocorrelation, the SOCc kriging prediction to the new location based on the soil type map showed similarities with the general distribution expected and a relatively low standard error (Figure 5.6).

To analyze a more detail autocorrelation dependence with soil depth, this analysis was generated at six different depths (down to 5, 10, 15, 30, 60, 100 and >100 cm). There was a general moderate dependence on the best spatial correlation (NSR = 35%) at the first horizon (0-5 cm) and a gradual decrease with depth up to NSR = 72% at 2 m (Table 5.4).

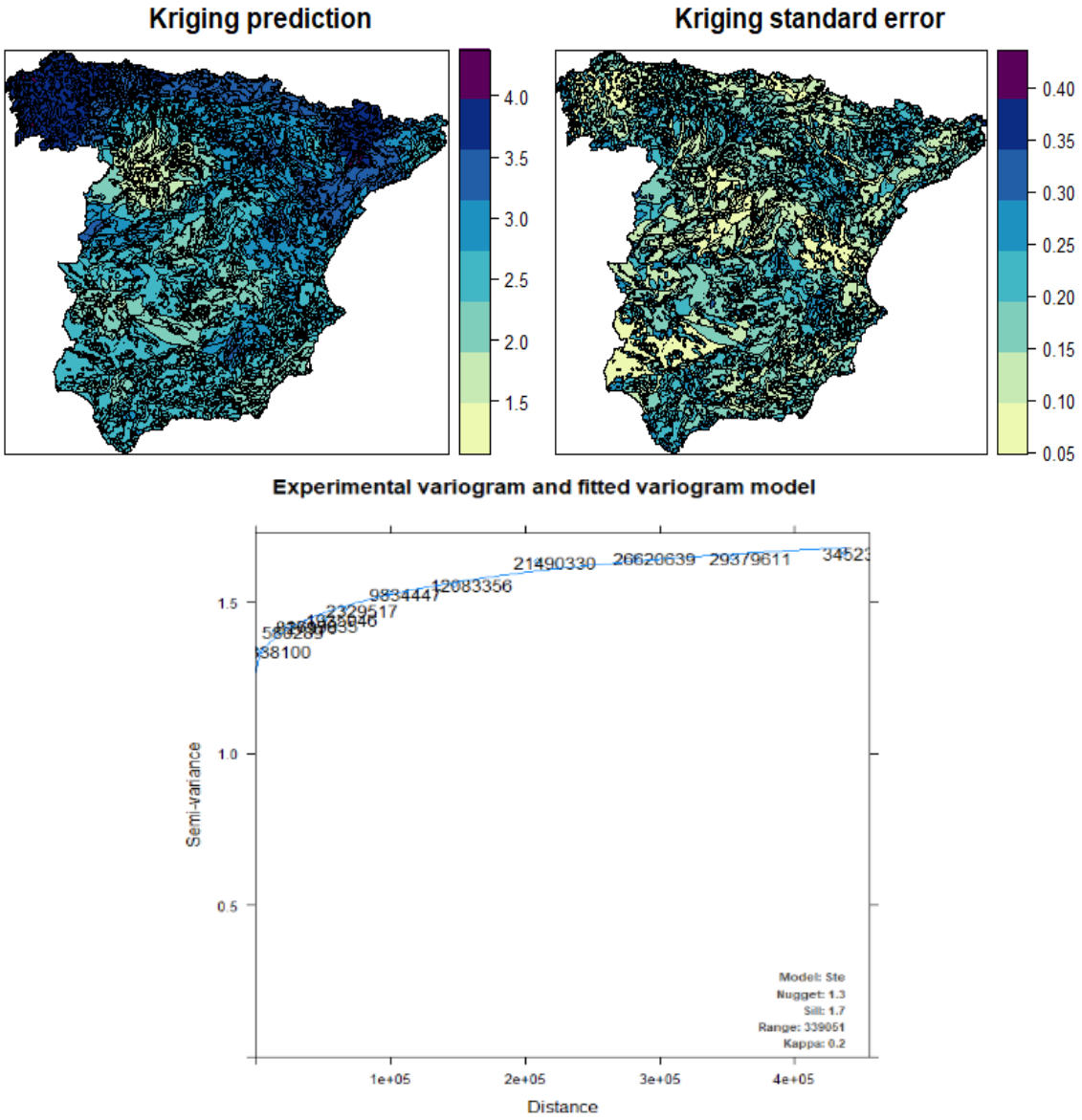


Figure 5.6 Spatial autocorrelation prediction on the soil type map (IGN, 1995) and semivariogram plot of the natural logarithm of SOC concentrations in soil profile samples.

Table 5.4 Data autocorrelation for the logarithm of soil organic carbon concentration (SOCc) at six different depths.

Depth (cm)	Model	Nugget	Sill	Range	Kappa	Ratio (NSR)	Spatial autocorrelation
0-5	Ste	0.22	0.67	28657	0.20	0.34	moderate
0-15	Ste	0.47	0.95	163474	0.20	0.50	moderate
0-30	Ste	0.65	1.05	133406	0.40	0.62	moderate
0-60	Ste	0.85	1.33	221661	0.20	0.64	moderate
0-100	Ste	1.14	1.53	210823	0.30	0.74	moderate
0-200	Ste	1.25	1.75	352425	0.20	0.72	moderate

5.3.1.2. Profile spatial distribution: database representativeness

Spatial representativeness of the database used for modeling SOC was analyzed with the maximum entropy distribution technique (Maxent). The predictive capacity of the model was tested with the mean of the area under the curve (AUC) of the training data. The AUC calculated by the model (AUC=0.611) was higher than the AUC of a random prediction (0.5), which confirms the applicability of maximum entropy analysis for evaluating the representativeness of the database.

Figure 5.7 depicts the probability distribution showing that the ecosystems were well represented by the profile database used in this study (0 low representativeness, 1 high representativeness). This map shows that the ecosystems defined by the environmental variables in the SOC modeling are not uniformly represented by the profile data. The best represented areas were the mountain systems, both in the interior (Central System, Sierra de la Demanda and Sierra Morena) and in the north and southeast (Cantabrian Range, Pyrenees and Betic System). In the Betic System, in particular, representativeness of these abrupt zones were interspersed with gentler reliefs, including coastal zones. The worst represented areas were in the Central Plateau (both in the northern and southern areas), the Ebro Depression and the Tajo and Guadiana Basins.

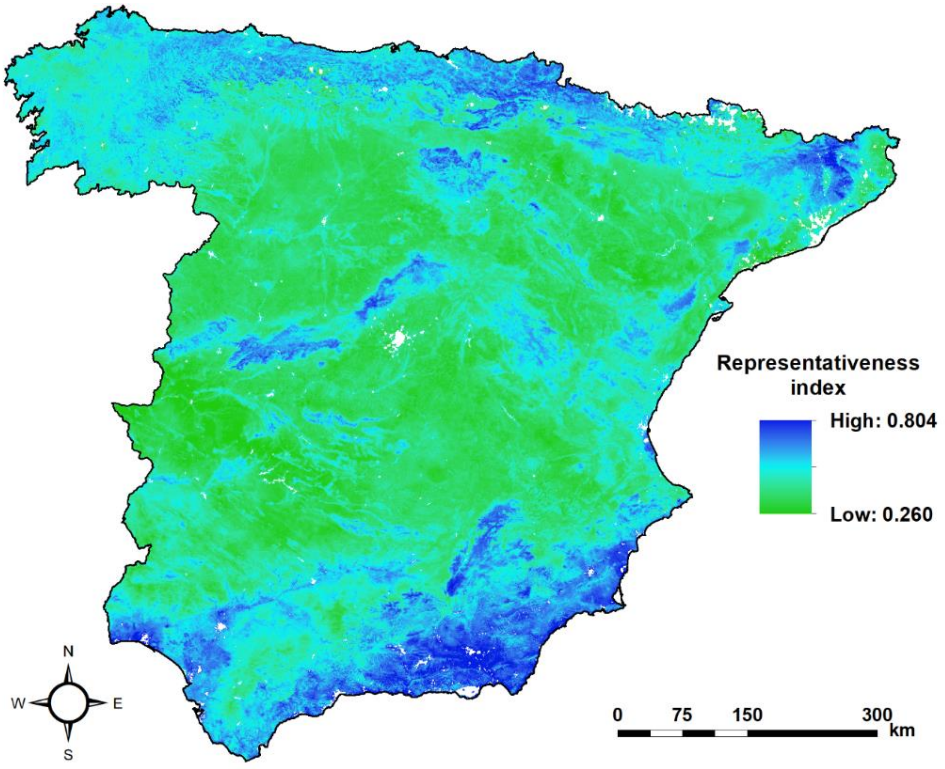


Figure 5.7 Maximum entropy model of soil data distribution spatial representativeness (MaxEnt program).

5.3.2. Modeling and prediction

5.3.2.1. Feature selection

A total of 254 covariates were generated for modeling SOC in peninsular Spain. These covariates were categorized into four main groups: topography, climate, remote sensing and soil. However, the high number of covariates and combinations would have led to an enormous number of submodels for predictor selection, causing problems for their computation and synthesis. Therefore, selection of the covariates was based on those categories, which ensured representation of the various soil-forming factors. In each category, combinations were made to evaluate the importance of the variables in the different models: SOCc, SOCs and their normal logarithms with covariate original values and their rescaled values. The most accurate combination in the selection models was the log SOC and covariate original values. Finally, the model diagnostic criteria and prediction maps were back-transformed into their original units.

According to the criteria for evaluating the importance of the variable in each selection method (multiple linear regression, Bayesian analysis and partial dependence plots), the covariate space was considerably reduced in all SOC estimations. Table 5.2 shows the covariates selected for each SOC modeling at the different depths. For modeling SOCc at the 0-30 cm standard depth, 17 covariates were selected, and for 30-100 cm 19 covariates were selected. Due to the slight difference between the number of profile samples used in stock spatial modeling at the 0-30 cm depth and the effective depth (1475 and 1499, respectively) the selection of covariates was the same in both cases. Therefore, 13 covariates were selected for SOCs in both depths.

The indices derived from terrain roughness, specifically, the maximum (rough-magnitude), scale size related to topographic position (dev-scale), and slope, were the most relevant topographic variables. Annual precipitation and mean and minimum temperatures in Spring (May) were the most significant climate variables. Among the remote sensing covariates related to the carbon cycle, NDVI (mean and maximum) was the most important of the annual covariates, and the mean EVI in Spring (March) the most important monthly covariate. The importance of indices related to the water cycle (ET and NDWI) in both annual and monthly covariates should also be emphasized. The month that most influenced SOC was March. The covariates related to soil, i.e., lithology, soil type and land use, were included directly in all the SOC models, because they were considered essential to modeling soil properties.

Table 5.5 Selected covariates for spatial modeling of soil organic carbon concentration (SOCc) and soil organic carbon stock (SOCs) variables at different standard depths.

Variable/depth	SOCc (0-30 cm)	SOCc 30-100 cm	SOCs 0-30 cm/ESD [1]
Climate variables [2]			
Annual	- Min Temp	- Mean Pp	- Mean Pp - Mean Temp
Monthly	- Mean Pp for May - Mean Temp for May - Min Temp for May	- Max Temp for Feb - Mean Temp for May - Min Temp for March	- Max Temp for April - Min Temp for May
Satellite-indices variables [2]			
Annual	- Max Albedo - Mean NDVI - Max NDWI	- Mean ET - Max LST - Max NDVI - Max NDWI	- Max ET - Mean NDVI
Monthly	- Mean Albedo for August - Mean ET for March - Mean LST for March - Mean NDVI for June	- Mean EVI for March - Mean ET for May - Mean LST for July - Mean NDWI for July	- Mean EVI for March
Topographic variable [2]			
	- dev-magnitude - dev-scale - rough-magnitude	- dev-magnitude - dev-scale - rough-magnitude - slope	- elev-stdev - rough-magnitude - slope
Land features			
	- Lithology	- Soil types	- Land use/cover
Total number	17	19	13

(1) ESD: effective soil depth. **(2)** Max: maximum; Min: minimum; dev-magnitude: Max multiscale deviation; dev-scale: Scale of the Max multiscale; rough-magnitude: Max multiscale roughness; elev-stdev: Elevation standard deviation; Pp: precipitation; Temp: temperature; NDVI: Normalized Difference Vegetation Index; NDWI: Normalized Difference Water Index; ET: Evapotranspiration, LST: Land Surface Temperature.

5.3.2.1.1. Linear models

To evaluate variable importance (VI) using the assumption of linearity between the covariates and the SOC, the absolute value of the t-statistic was calculated for each parameter in the linear model applied. Values over 2 are considered acceptable since the higher the t-value, the higher confidence in the coefficient as a predictor (Greenwell and Boehmke, 2020). Table 5.6 shows the VI in descending order, associated with the selected covariates by category for each study case.

Table 5.6 Variable importance (VI) scores based on the linear regression model (absolute value of the t-statistic) for covariate selection in soil organic carbon (SOC) estimation at standard depths. The final selection of covariates is in bold.

SOCc [(1)] ***0-30 cm***			***SOCc*** [(1)] ***30-100 cm***			***SOCs 0-30 cm/ESD*** [(2)]		
Variable [(3)]	VI	Sign [(4)]	Variable	VI	Sign	Variable	VI	Sign
Category: Climate								
Mean Pp August	5.301	POS	**Min Temp March**	5.254	NEG	**Max Temp April**	1.616	NEG
Min Temp	4.849	NEG	**Mean Temp May**	4.992	NEG	Max Temp Feb	1.001	NEG
Mean Temp May	3.813	NEG	**Max Temp Feb**	3.003	NEG	**Min Temp**	0.957	NEG
Mean Pp May	3.430	POS	Mean Pp July	2.572	POS	**Min Temp May**	0.782	NEG
Max Temp	2.871	NEG	**Mean Pp**	1.166	POS	**Mean Pp**	0.489	POS
Min Temp May	2.346	NEG	Max Temp March	0.157	NEG			
Category: Satellite-indices								
Max Albedo	5.741	NEG	**Mean ET**	3.861	NEG	**Mean EVI March**	1.318	POS
Max ET	4.309	NEG	**Mean LST July**	3.560	POS	Mean LST April	1.185	NEG
Mean NDVI	4.303	POS	**Mean ET May**	3.778	POS	**Mean NDVI**	1.059	POS
Mean ET May	4.119	NEG	Max Alb	2.649	NEG	**Max ET**	0.757	NEG
Mean Albedo August	3.344	POS	**Max NDWI**	2.462	POS			
Max NDWI	3.253	POS	**Mean NDWI July**	2.259	POS			
Min Alb	2.726	POS	**Max LST**	2.207	POS			
Mean ET March	2.661	NEG	Mean ET July	2.106	POS			
Mean NDWI	2.508	POS	**Max NDVI**	1.842	POS			
Mean NDVI June	2.388	POS	Mean NDWI March	1.501	NEG			
Mean LST March	2.287	NEG	**Mean EVI March**	1.069	NEG			
Mean NDW December	0.006	NEG						
Category: Topography								
rough-magnitude	17.593	POS	**rough-magnitude**	4.158	POS	**rough-magnitude**	1.822	POS
dev-magnitude	11.177	POS	**dev-scale**	3.904	NEG	dev-scale	1.538	POS
geom	6.409	POS	tcurv	1.533	NEG	**elev-stdev**	1.132	NEG
dev-scale	6.249	NEG	**dev-magnitude**	1.390	POS	spi	0.307	POS
			slope	0.829	POS			

(1) VI scores refer to the logarithm of SOC concentration (SOCc) and SOC stock (SOCs). **(2)** ESD: effective soil depth. **(3)** Abbreviations: Max: maximum; Min: minimum; dev-magnitude: Max multiscale deviation; dev-scale: Scale of the Max multiscale; rough-magnitude: Max multiscale roughness; elev-stdev: Elevation standard deviation; geom: Geomorphon; tcurv: Tangential curvature; spi: Stream power index; Pp: precipitation; Temp: temperature; NDVI: Normalized Difference Vegetation Index; NDWI: Normalized Difference Water Index; ET: Evapotranspiration, LST: Land Surface Temperature. **(4)** POS: positive sign, NEG: negative sign.

The most significant climate variables were related to temperature, and were negatively correlated with SOC. The vegetation indices (NDVI, EVI), ET and LST were the most important remote-sensing indices in the linear models, especially the annual data for SOCc at 0-30 cm and the monthly data in spring (mainly March and May) for the others. The topographic variable, rough-magnitude, was especially important, and clearly the most influential in all

cases. This variable represents the magnitude of deviation from the normal vector from the surface, measured in degrees. Thus, as can be deduced from the positive data trend, the more irregular and abrupt a surface, i.e., a wider deviation, the larger the amount of SOC.

The highest VI measure was found in SOCc at a depth of 0.30 cm, and the lowest in SOCs in every category of covariates.

Regarding the evaluation metrics of the linear models, the residual standard error (RSE) was high and very similar at the standard depths, with a mean of 13.5 g/kg for SOCc and 3.5 g/m^2 for SOCs. However, the adjusted coefficient of determination (Adjusted R^2) was higher for SOCc at the 0-30 cm depth (0.36), lower for SOCc at 30-100 cm (0.18) and intermediate for SOCs (0.28).

5.3.2.1.2. Bayesian analysis

Metrics for the precis of model fit were acquired with Bayesian model analysis. Assuming a normal distribution of log SOC, the displayed output included the estimates (mean) and standard error (StdDev) with optimal confidence intervals (5.5% and 94.5% by default) and parameter correlations. The StdDev is the standard deviation of the posterior distribution, while the mean is its peak. The posterior intervals, which are the quadratic estimates derived from StdDev, showed the 89% confidence intervals.

The resulting precis values of the models were plotted to evaluate the VI for SOCc (Figure 5.8) and for SOCs (Figure 5.9). The estimated means are marked by circles and their corresponding 89% confidence intervals by solid horizontal lines. The covariates which did not include 0 at the 89% confidence intervals were considered statistically significant in modeling SOC. In addition, for a same model, a narrow confidence interval may be interpreted as a strong relationship between the covariate and the predicted value. Accordingly, the most important variables in the SOCc model were those related to the remote sensing indices, especially ET and EVI (Figure 5.8). For SOCs, they were the climate-related variables (Figure 5.9).

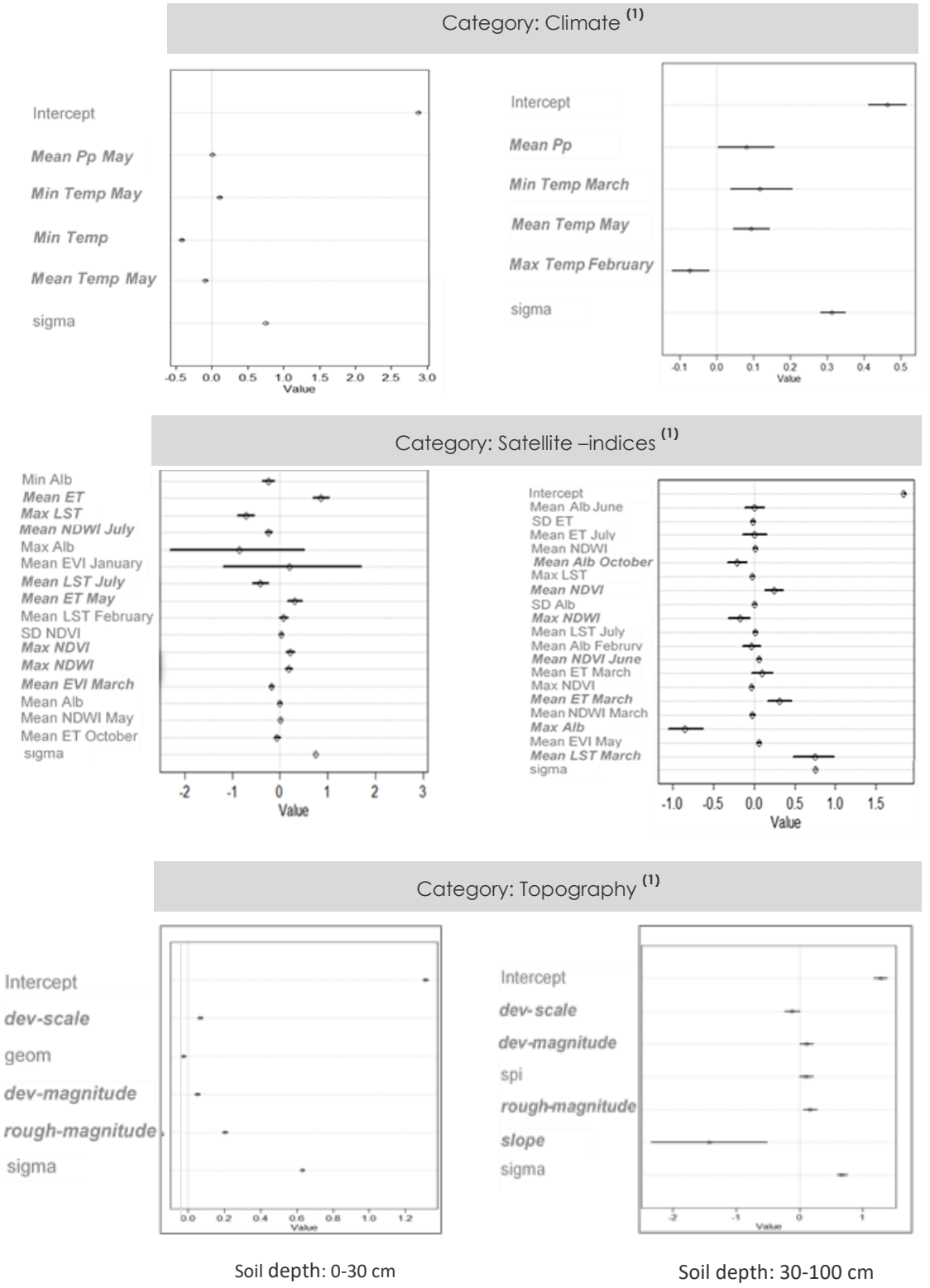


Figure 5.8 Soil organic carbon concentration (SOCc) plots of posterior means and 89% highest density intervals of parameters based on Bayesian analysis models for the covariate selection. Data refer to the log SOCc at standard depths of 0-30 cm (left) and 30-100 cm (right). The final selected covariates are in cursive and bold.

(1) Max: maximum; Min: minimum; SD: standard deviation; dev-magnitude: Max multiscale deviation; dev-scale: Scale of the Max multiscale; rough-magnitude: Max multiscale roughness; geom: Geomorphon; spi: Stream power index; Pp: precipitation; Temp: temperature; NDVI: Normalized Difference Vegetation Index; EVI: Enhanced Vegetation Index; NDWI: Normalized Difference Water Index; ET: Evapotranspiration, LST: Land Surface Temperature

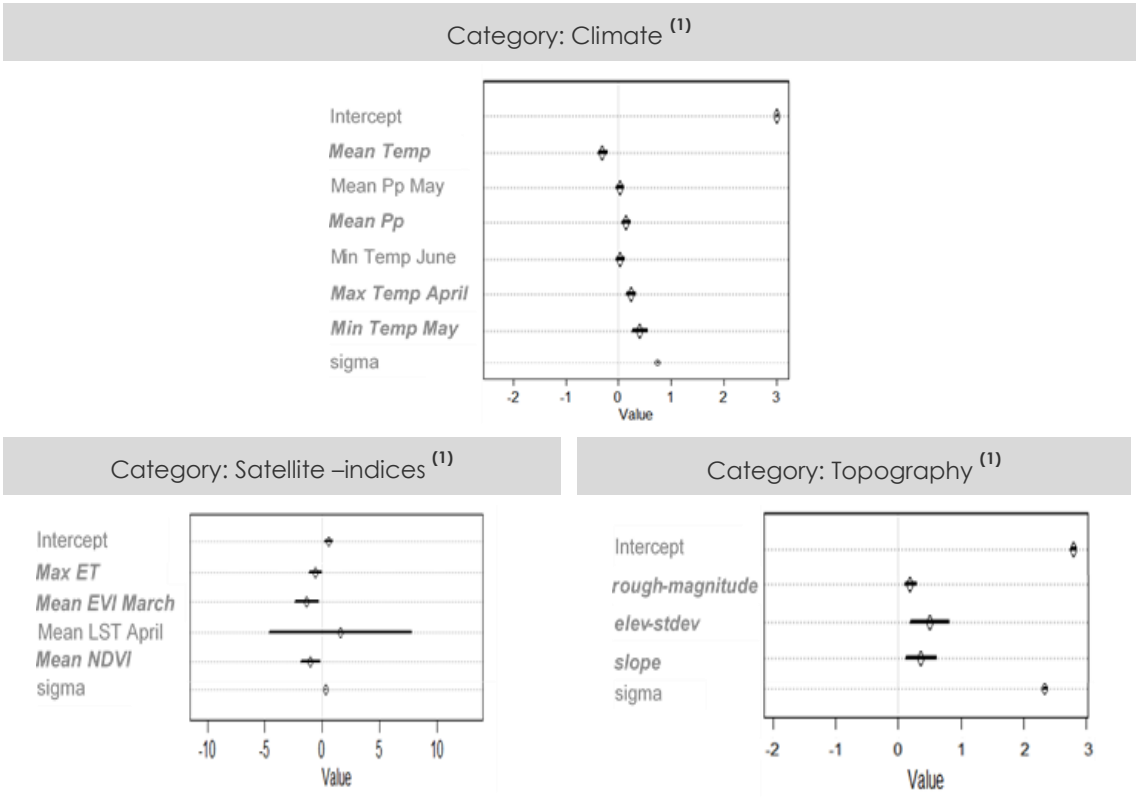


Figure 5.9 Soil organic carbon stock (SOCs) plots of posterior means and 89% highest density intervals of parameters based on Bayesian analysis models for the covariate selection. Data refer to the log SOCs. The final selected covariates are in cursive and bold.

(1) Max: maximum; Min: minimum; SD: standard deviation; dev-magnitude: Max multiscale deviation; dev-scale: Scale of the Max multiscale; rough-magnitude: Max multiscale roughness; geom: Geomorphon; spi: Stream power index; Pp: precipitation; Temp: temperature; NDVI: Normalized Difference Vegetation Index; EVI: Enhanced Vegetation Index; NDWI: Normalized Difference Water Index; ET: Evapotranspiration, LST: Land Surface Temperature.

5.3.2.1.3. Partial dependence plots

The marginal effect of the covariates on the SOC at different depths analyzed with univariate partial dependence plots (PDPs) is depicted in Figure 5.10. This figure shows the most influential covariates by category (climate, remote-sensing indices and topography) for each case of SOC. The x-axis represents the covariate, and the y-axis is the log SOC. The plots depict the direction of the association and the type of relationship (linear or nonlinear) between the covariate and SOC. Based on this visual representation, the topographic rough-magnitude covariate described a linear relationship in both cases of SOCc, stronger in the upper 30 cm. This was in agreement with the results found for the VI scores of the multiple linear regression model, and also observed in the Bayesian analysis precis plot. However, for SOCs, the most representative topographic covariate was slope, where predicted probability was higher on slope values over 30%, and even negative on slope values between 15-30%. In the multiple linear regression models, this covariate was not significant.

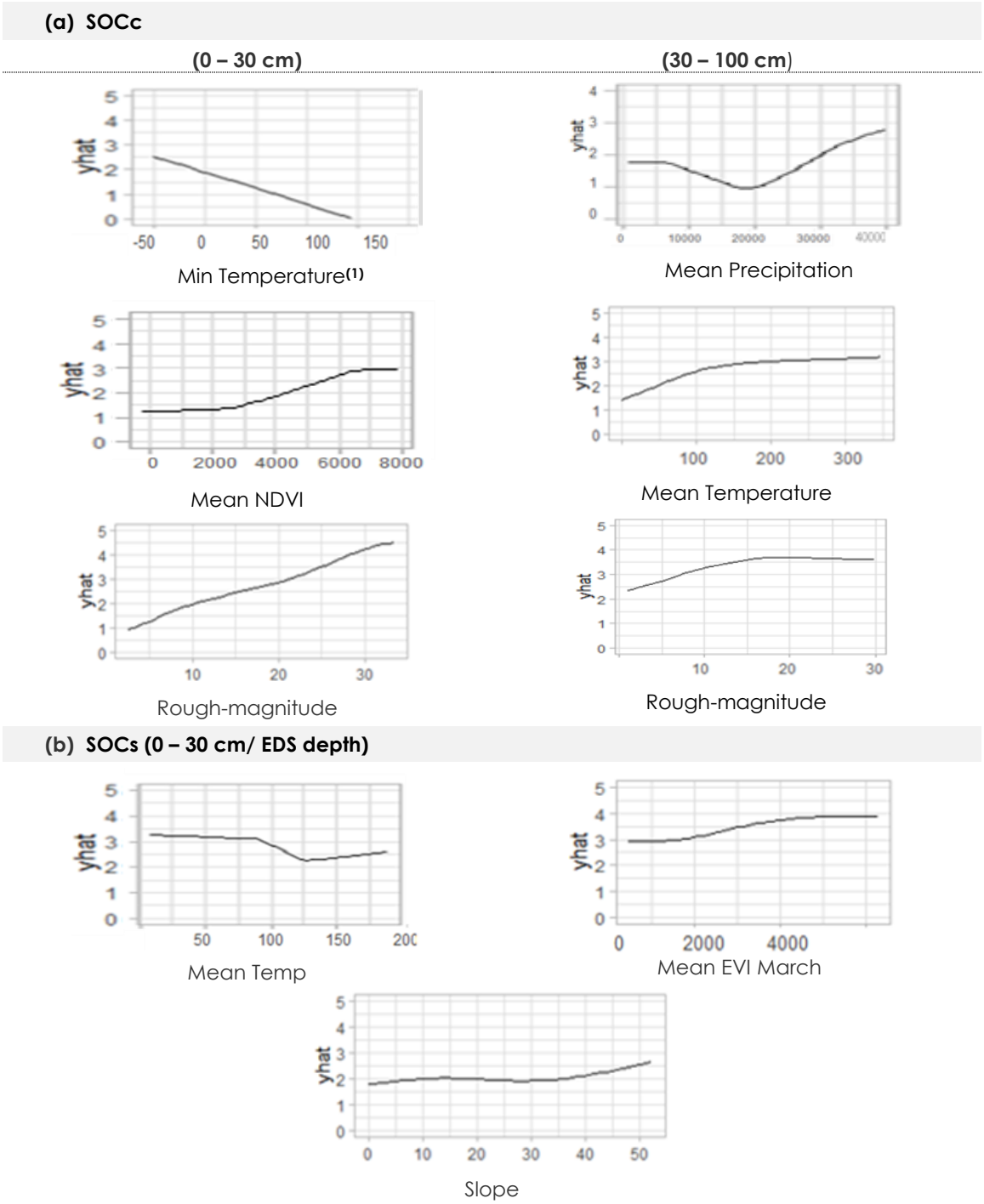


Figure 5.10 Univariate partial dependence plots based on projection pursuit regression for the covariate selection by category. Plots refer to the relatively most influential covariate on the log soil organic carbon concentration (SOCc) at standard depths of 0-30 cm and 30-100 cm and log soil organic carbon stock (SOCs) at 0-30 cm and effective soil depth (ESD). (1) The x-axis represents covariate values. The y-axis represents the log SOC. Max: maximum; Min: minimum; SD: standard deviation; rough-magnitude: Max multiscale roughness; elev-stdev: Elevation standard deviation; Pp: precipitation; Temp: temperature; NDVI: Normalized Difference Vegetation Index; EVI: Enhanced Vegetation Index; ET: Evapotranspiration, LST: Land Surface Temperature.

The correlation between SOC and temperature was negative, with a clear linear relationship with SOCc at 30 cm, as observed in both the linear and Bayesian analyses. The mean annual temperature trend for SOCs was also negative but with a more complex relationship, showing a slight rise starting at 15°C. This could explain the discrepancy in importance of this variable on SOCs in linear analyses

versus the PDP and Bayesian analyses. The remote-sensing index predictors revealed quite similar "flatness patterns", especially for SOCs, highlighting the annual means of NDVI, EET and EVI for SOCc at 30 cm, SOCc at 30-100 cm and SOCs, respectively.

5.3.2.2. Model Calibration Data

For the SOC variability prediction and its associated uncertainty, we combined three methods of multimodel ensembles using machine-learning algorithms: quantile regression forest (QRF), ensemble machine learning (MLR) and auto-machine learning (AutoML).

In view of chapter 4 results, additional kriging of residuals was no considered for the QRF model, since the density and clustered distribution of the soil data reduced the advantage of this method. The most accurate model selected in the MLR predictive ensemble approach was Cb, which was validated by spatial cross-validation with the xy locations of the soil data spatial component. To generate the final ensemble in AutoML modeling (stacked ensemble with all models), the "GLM" and "DeepLearning" algorithm families were excluded due to their strong influence on decreasing accuracy in the final spatial prediction.

The statistical summary of the models used for the SOC spatial prediction grouped by depth is described in Table 5.7. For model diagnosis, the concordance correlation coefficient (CC) was calculated for both the calibration and validation dataset (CC_{cal} and CC_{val}), and nRMSE and nMAE were calculated for the validation dataset.

The differences between CC_{cal} and CC_{val} were rather slight, except for the AutoML model, where the GBM algorithm family contributed to a higher CC_{cal}. Although the RMSE and MAE may both be used to measure the accuracy of continuous variables, the average prediction error is easier to interpret with the MAE, and furthermore, it does not penalize large errors. Based on the CC_{val} and nMAE metrics, a decrease in accuracy with depth was observed in both SOC variables, e.g., CC_{val}= 0.583 and nMAE =0.441 for the 0-30 cm horizon and CC_{val}= 0.351 and nMAE =0.668 for 30-100 cm, for SOCc.

At the same depth, i.e., in the upper 30 cm, the SOCc was more accurate than SOCs in all the models, while at 30-100 cm it was the least accurate (Table 5.7). In general, the differences between the nMAE in the three predictive ensemble approaches for each study case were rather slight, however, CC showed significant differences.

Table 5.7 Evaluation of model performance for soil organic carbon concentration (SOCc) and stock (SOCs) at 0-30 cm, 30-100 cm and effective soil depth (ESD). nRMSE and nMAE are expressed as parts per unit.

Parameter	CC_{cal}	CC_{val}	nRMSE	nMAE	CC_{cal}	CC_{val}	nRMSE	nMAE
SOC concentration								
Predictive model	**0-30 cm**				**30-100 cm**			
AutoML	0.825	0.583	0.684	0.433	0.629	0.351	1.296	0.668
QRF	0.485	0.472	0.733	0.458	0.354	0.307	0.680	0.610
MLR	0.434	0.350	0.775	0.474	0.278	0.227	0.737	0.578
SOC stock								
	0-30 cm				**Effective soil depth**			
AutoML	0.672	0.417	0.548	0.441	0.563	0.378	0.651	0.516
QRF	0.445	0.381	0.545	0.504	0.358	0.295	0.646	0.552
MLR	0.401	0.270	0.628	0.535	0.323	0.232	0.711	0.570

These predictive models contributed to obtaining the final SOC spatial distribution maps for peninsular Spain at a 90-m spatial resolution (Figure 5.11 and Figure 5.12). These final maps were generated by assigning the most accurate model prediction for each pixel, i.e., with the lowest uncertainty for the specific pixel.

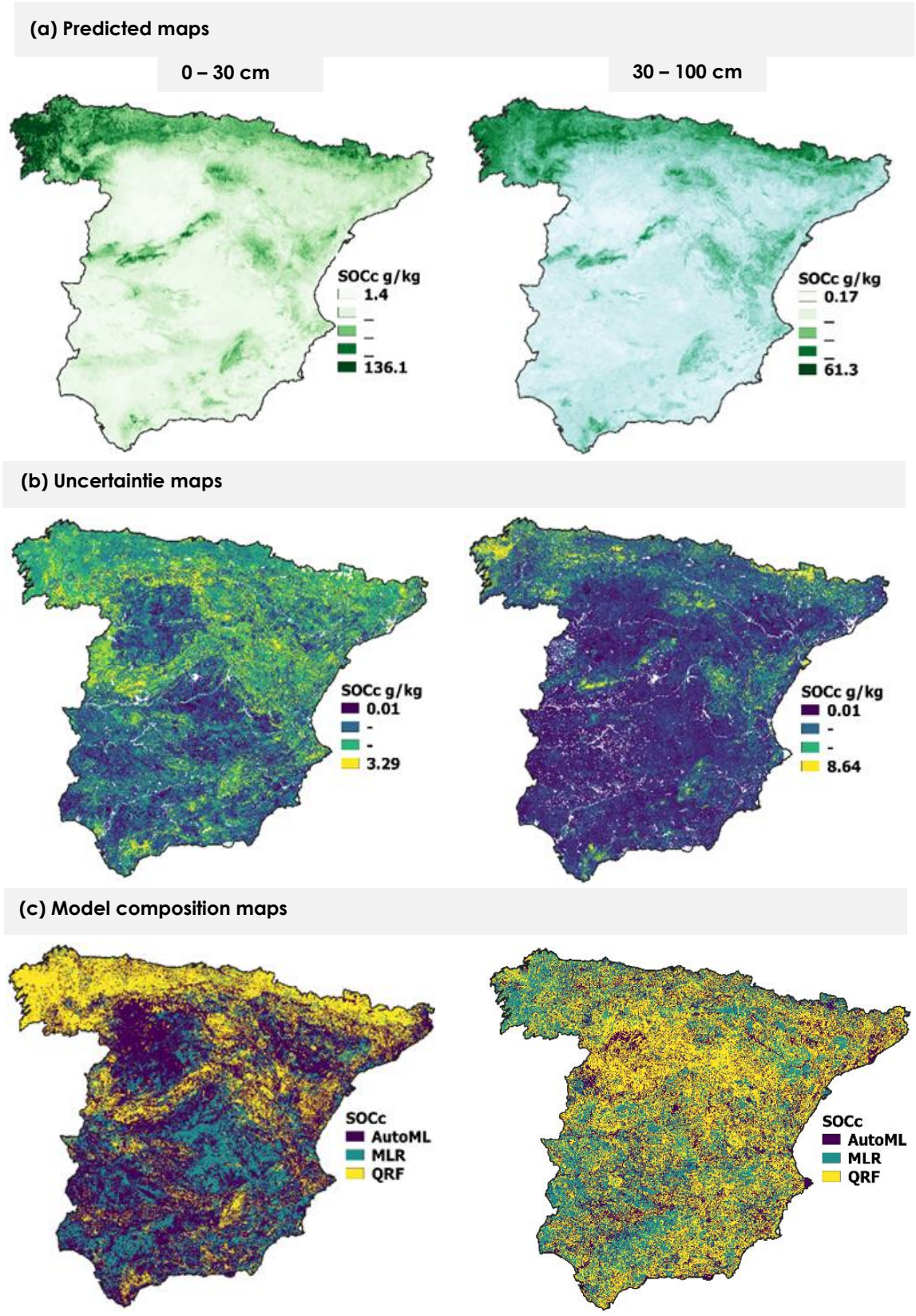


Figure 5.11 Soil organic carbon concentration (SOCc). Final spatial distribution maps at different depths for peninsular Spain (90-m pixel resolution). Data display was stretched by the cumulative pixel count cut method (default range 2%-98%).

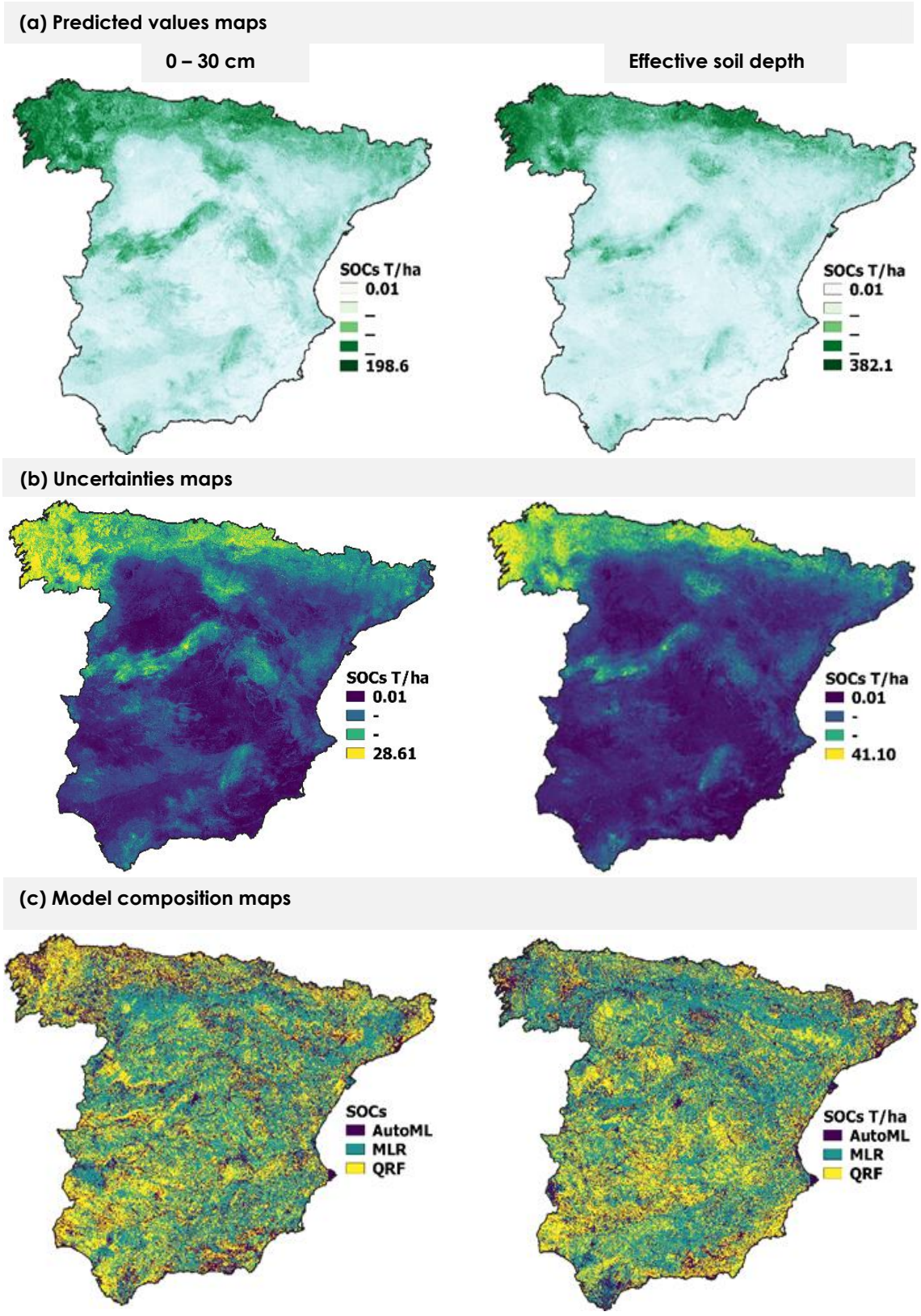


Figure 5.12 Soil organic carbon stock (SOCs). Final spatial distribution maps at different depths for peninsular Spain (90-m pixel resolution). Data display was stretched by the cumulative pixel count cut method (default range 2%-98%).

The SOC maps described a strong spatial pattern, far from any random distribution. In general, the distribution of SOCc and SOCs at the depths analyzed showed very similar geographic patterns. The highest predictions for both SOCc (Figure 5.11) and SOCs (Figure 5.12) corresponded to a both a climatic and a topographic pattern. Climate explained the gradient from maximum SOC in the northwest and north, associated with the humid area of the peninsula, to a progressive decrease to the east. The topographic pattern explained the highest values in the main mountain systems in the interior of the peninsula (Central Mountain System, Iberian System and east of the Subbetic System). The land cover factor was also represented. In general, the lowest values corresponded to the large agricultural areas of the Central Plateau (both north and south), the southwest Guadiana alluvial plain and the Guadalquivir depression, along with the subarid areas in the southeast.

The prediction uncertainty maps showed a different spatial pattern for SOCc, which is clearly divided into north, except the Central Plateau, and south of the peninsula. This spatial pattern was similar to the predictive maps, with higher variability in areas with higher SOC. The uncertainty map was calculated based on the SD of the simulations generated for each predictive model approach (Figure 5.3). Therefore, it may be deduced that the models were less accurate in the higher SOC areas than in lower ones. This could also explain the lower uncertainty found in the SOCc prediction in the 30-100 cm map.

Even though the spatial pattern of SOC uncertainty maps was similar, the values on the stock maps were lower than for concentration maps, with lower uncertainty associated with the lowest estimated values, and vice versa. Similarly for SOCc, the uncertainty in the upper horizons (0-30) in stock was higher than in the complete profile (EDS), although with slight differences.

The spatial distribution of the different models (QRF, AutoML and MLR) in the final map differed in combinations of SOC and depth. For SOCc at 0-30 cm, the spatial distribution of the three models was homogeneous, except for the higher SOC areas in the north of the peninsula and in mountain systems, where the QRF model was dominant. The prediction of SOCc at 30-100 cm depicted a much lower contribution of the AutoML ensemble, where the MLR and QRF models were dominant with different distribution areas.

Visual analysis of the maps showed that model capacity for SOCs prediction at different depths remained relatively constant. The distribution of the models assigned to each pixel did not follow a clearly defined spatial pattern, which complicated the characterization of the models associated with SOCs values.

The analysis of residuals in the final map compilation based on the linear relationship between observed and predicted values showed R^2=0.68 for SOCc and R^2=0.54 for SOCs in the upper 30cm, and low values in deeper horizons.

The mean SOCc in peninsular Spain was 15.7 g.kg^{-1} in the upper horizons (0-30), decreasing to 3.6 g.kg^{-1} in horizons bellow to 100 cm. The total stock in the profile, i.e., at the ESD, was 3.8 Pg C, storing 74% in the upper 30 cm (2.82 Pg C) (Figure 5.13). In this case, SOC stock estimates accounting only for the upper 30 cm would be underestimating over 25% of the stock stored in the subsurface horizons.

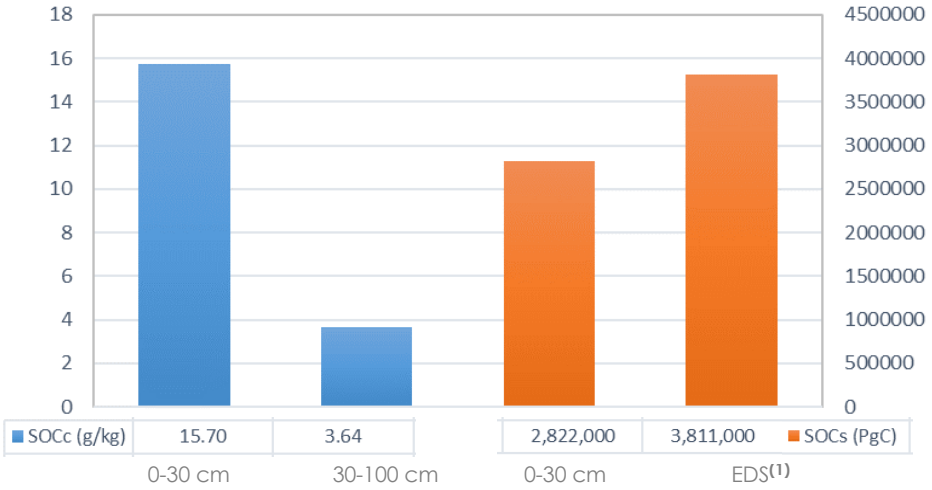


Figure 5.13 Mean of soil organic carbon concentration (SOCc, g/kg) and sum of stock (SOCs, Pg) in peninsular Spain at 90 m spatial resolution in the different soil depth layers. (1)ESD= effective soil depth.

5.3.2.3. Model Validation

The prediction capacity of the final maps was evaluated by plotting the conditional quantiles to verify the effectiveness of the models across the full distribution of SOC observed values (Figure 5.14).

The final values predicted by the models did not cover the full range of observed values, with the highest predicted for SOCc at 0-30 cm (136.1 g/kg) and for SOCs at the ESD (38.33 kg/m^2). The median values of the predictions (red line) showed more variability in high SOC. Compared to a perfect model (blue line), which would have low dispersion, the SOCc at 0-30 cm had the highest model performance based on these metrics, and the lowest at 30-100 cm.

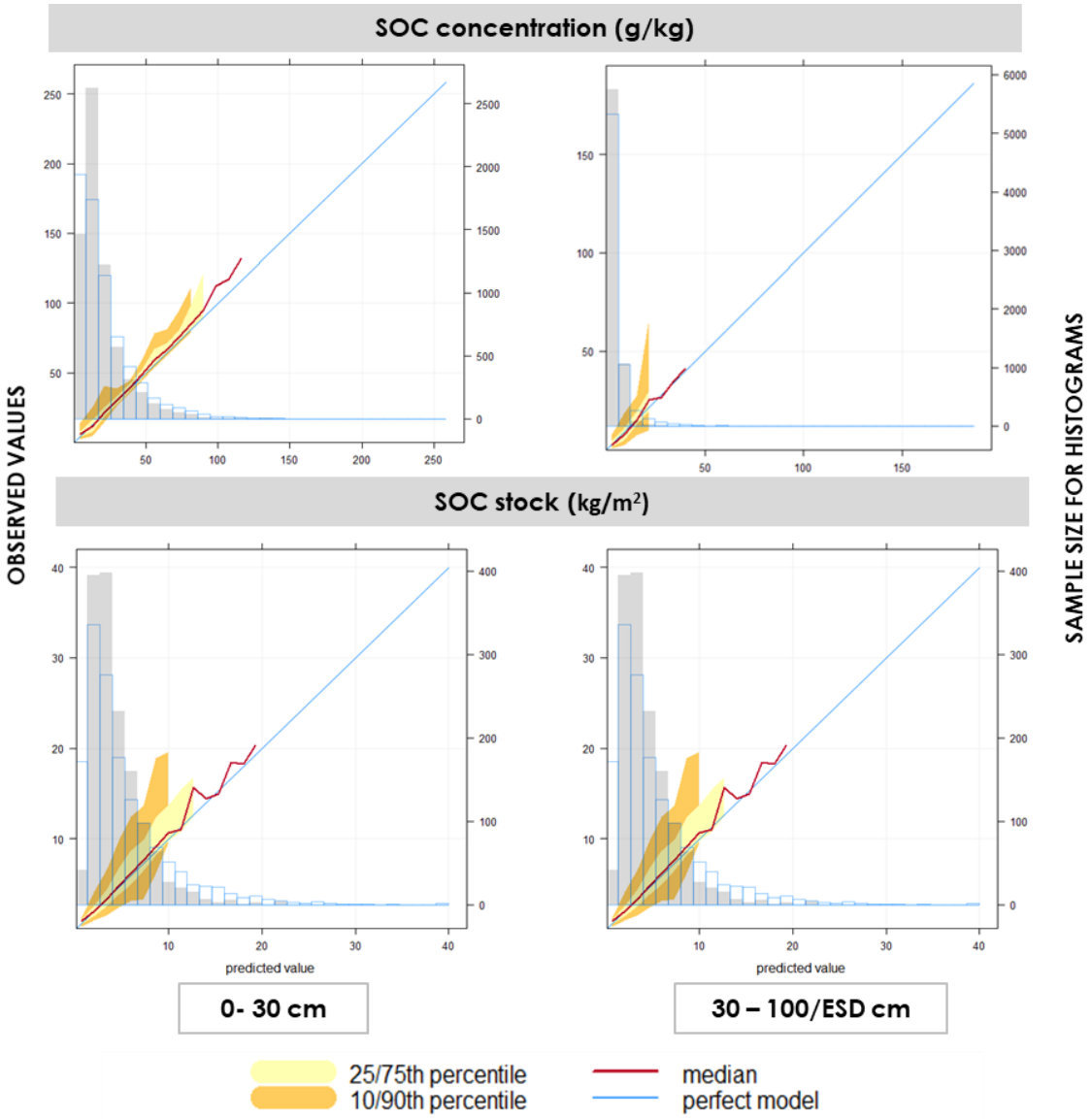


Figure 5.14 Conditional quantile plots across the full distribution of observed soil organic carbon (SOC) concentration (top) and stock (bottom) at different depths.

5.3.3. Data Availability

The final 90-m-resolution maps of SOCc and SOCs spatial estimation at different depths in peninsular Spain are available on a public repository at https://drive.google.com/drive/folders/1SykCBAbpHnUPHcCbguE34jiBfBIvOlnM?usp=sharing

Maps are in standard formats and can be used for both national applications and global scale analyses.

5.4. DISCUSSION

The methodological procedure proposed in this study resulted in reasonably accurate, spatially explicit predictions of SOCc and SOCs in peninsular Spain, achieving the best possible results with the information and methods available. This is demonstrated by the following indicators: 1) the robustness of the models based on the small differences between the information criteria for calibration and validation and showing no overfitting of models (Mulder et al., 2016); 2) the acceptable mean errors in the SOC model diagnostics (MAE) in a Mediterranean climate area with relatively poor sample distribution (Hengl et al., 2004; Zeraatpisheh et al., 2019); and 3) the agreement between the final generated maps with the expected spatial distribution and with other national SOC modelings such as those of Calvo de Anta et al. (2020) or (Rodríguez Martín et al., 2016).

However, the predictive capacity for SOC concentration and stock differed depending on the carbon variable and depth. At the same depth, i.e., at 0-30 cm, the SOCc map showed more accurate predictions than the SOCs map. This difference was also demonstrated in chapter 4. Such as it was concluded, the lack of soil databases with the basic information required to determine the carbon stock, i.e., mainly bulk density and coarse fragments, could propagate errors by the extrapolation of parameters (Durante et al., 2020; Poeplau et al., 2017). Thus, only those soil samples with complete information were used to calculate the stock in this study, which greatly decreased the number of available profiles (data) and, consequently the prediction capacity.

Data availability

The analysis of database representativeness, as well as the slight difference between the profile samples at 0-30 cm and ESD (1475 and 1499, respectively), showed the low number of representative profile samples with complete information.

Despite efforts to compile the available soil information for this study, acquiring new data that improved the representativeness of the originals, it was demonstrated that the soil information available nationwide in operative formats was clearly insufficient. A national effort would be necessary to increase the number of available profiles, and specifically, the homogeneous representativeness of the whole territory. Creation of a national database compiling all the legacy soil information in the country is recommended. Such a database would serve as a basis for analyzing the information available and designing new studies on sustainable use of soil beyond carbon. This objective has been clearly underlined by several international organizations, such as the FAO and the ISRIC, and some countries have already achieved this aim successfully, such as Chile (Pfeiffer et al., 2020) or Ecuador (Armas et al., 2021).

Final modeling could have been influenced by the relatively low number of soil samples, specifically in certain data intervals, since it was the stock values with low profile representation that were modeled least accurately (Brungard et al., 2015; Hengl et al., 2007) . In our case, 40% of the data were below 12 kg/m^2. Likewise, although the number of profiles used in SOC concentration modeling was considerably higher (8322 samples), 70% of the

samples had less than 15 g/kg. Therefore, improved SOC sample representativeness would increase the performance of DSM models (Brungard et al., 2015).

Covariate selection

We included the selection of an appropriate set of covariates in our methodology because they are a key factor in DSM accuracy, especially when samples are limited (Liang et al., 2020). The relationship between each potential covariate and the soil property may be a linear, polynomial or even more complex function. The results of variable selection approaches used in this study (multiple linear regression, Bayesian analysis and PPR) described a very strong linear relationship with topology and temperature, specifically with the rough-magnitude and minimum temperature covariates, which was consistent in all of the selection methods. However, other covariates with strong relative influence on SOC modeling, such as slope, mean annual precipitation and the NDWI remote sensing index, were not clearly detected by the linear models. As confirmed in other studies, the use of feature selection methods to detect both linear and nonlinear relationships is critical to improving modeling of digital mapping of soil properties (Hengl et al., 2015; Zeraatpisheh et al., 2019).

The inclusion in this study of a wide variety of detailed-scale geomorphological covariates differentiated it from previous similar studies such Calvo de Anta et al. (2020) or Doblas-Miranda et al. (2013). In those studies, elevation was the only topographic variable included in the SOC estimate in peninsular Spain, and was not significantly influential in the SOC modeling.

The climate influence on soil carbon has been related to both the increase in biomass contribution and the role of microorganisms in SOC stability, since the microflora in the soil are highly sensitive to humidity (Coûteaux et al., 1991), and organic matter decomposes more rapidly at higher temperatures (Knorr et al., 2005). Although the magnitude of the relationship between decomposition ratio and temperature is controversial (Giardina and Ryan, 2000; Powlson, 2005), several national studies have corroborated an inversely proportional relationship between SOC and temperature (Doblas-Miranda et al., 2013). Calvo de Anta et al. (2020) also reported minimum SOC stock in arid and desert areas and an increase with precipitation, although the soil behavior pattern was not uniform in the humid areas. The climate factor exclusively, was not able to explain the distribution of SOC. As concluded in chapter 4, the integration of other factors such as remote-sensing indeces related to ecosystem productivity improved the prediction of spatial distribution of soil carbon even with limited soil samples (Fathololoumi et al., 2020; Liang et al., 2020; Schillaci et al., 2017b).

The most influential covariates in both SOC concentration and stock spatial distribution were the rough-magnitude, annual minimum temperature, spring precipitation, and NDVI and NDWI remote-sensing indices related to plant cover and vegetation moisture, respectively. These variables are associated with peninsular mountain ecosystems and humid climates, and these areas were precisely the best spatially represented in the maximum entropy analyses. Therefore, there could be a bias in SOC modeling associated with the spatial distribution of the soil samples (Zeraatpisheh et al., 2019).

The feature selection approaches detected different influencing covariates for SOCc and SOC, as well as for different depths. Therefore, the selection of covariates must be performed for each context.

Modeling and prediction

According to chapter 4, previous studies in SOC spatial distribution have reported wide variability in data, models and predictions for Spain (Calvo de Anta et al., 2020; Rodríguez Martín et al., 2016). However, those predictions were based on the use of a single predictive model. Due to the limitations and disadvantages of the different models (Górecki et al., 2015), a combination of them is a recommended alternative approach for improving overall model precision (Taghizadeh-Mehrjardi et al., 2019). To our knowledge, no study has applied a combination of the most accurate models for each specific area to predict SOC in peninsular Spain.

The variability in SOC estimates confirm that there is no perfect model on digital SOC mapping, especially for large areas using a legacy soil profile database compiled over long periods of time and without a unified purpose. Depending on the area (combination of forming factors) and available data, some estimations may be more appropriate than others (Guevara et al., 2018). Therefore, different approaches were used in this study for SOC spatial prediction based on model ensembles, generating the final map by selecting the most accurate model value for each pixel.

The calibration and validation parameters of model accuracy calculated in this study were very diverse and strongly influenced by the statistical method used. The results showed the highest error in model validation in the MLR approach, which was performed with the strictest validation method (spatial cross-validation). However, it was a fairly accurate model, as observed in the final map compilation of both SOCc and SOCs.

Considering the low density of soil samples used for SOC concentration and stock modeling (17 and 4 soil samples per km^2 for SOCc and SOCs, respectively) the performance of the models was acceptable, especially for the prediction of SOCc in the upper 30 cm with a mean CCval=0.5 and nMAE= 0.45 (9.5 g/kg). These results are comparable to those described by Calvo de Anta et al. (2020), which reported an R^2 and RMSE of 0.61 and 14.8 g/kg, respectively, using a database of 12,724 profiles. The residuals analysis of the final map compilation based on the linear relationship between observed and predicted values showed a significant correlation (R^2= 0.67). However, high SOCc values (above 140 g/kg), which represented only 3% of the profile database, were poorly predicted. The uncertainties for the final map were significantly low, hence, the models used were rather accurate.

In view of the great difficulty in estimating SOC due to its high intrinsic spatial variability based on the complex interaction of factors, especially in areas with complex topography and climate(Hounkpatin et al., 2018; Vaysse and Lagacherie, 2017; Wälder et al., 2008), spatially explicit uncertainty maps are indispensable to land management and planning policies.

5.5. CONCLUSIONS

The SOCc and SOCs carbon maps generated in this study achieved the best possible results with the information and methods available for peninsular Spain. For the final SOC maps, a novel methodology was proposed based on the combination of different multi-model ensemble algorithms by assigning the most accurate model prediction for each pixel. Furthermore, considering the complex relationship between SOC and environmental variables, various feature selection methods were applied to detect both linear and nonlinear relationships. These techniques improved overall model precision by avoiding the risk of statistical redundancy. The small differences between the information criteria for calibration and validation set showed the robustness and no overfitting of models.

However, the final modeling could have been influenced by the relatively low number of soil samples available in peninsular Spain. The soil information available nationwide in operative formats was clearly insufficient. A national effort would be necessary to increase the number of available profiles, and specifically, the homogeneous representativeness of the whole territory.

Capítulo 6. Soil Organic Carbon Sequestration Potential National Map under Future Scenario Projections.

ABSTRACT

There is a pressing need to support sustainable management initiatives to mitigate climate change. Different practices have been used to reduce or restore soil organic carbon (SOC) loss and to estimate and project SOC variations. However, these initiatives have mainly focused on agricultural areas and site-specific studies. A large challenge is to identify areas with the potential for SOC sequestration for different land uses (i.e., forest areas) and with a national perspective. The goal of this work was to apply standardized and interexchange approaches to produce a SOC potential sequestration map for peninsular Spain. We predict SOC and SOC sequestration rates for the years 2020-2040 using the RothC model at the national level (in 1 km grids) following the methodological guidelines recently suggested by the Global Soil Partnership of the FAO (FAO, 2020a).

Our results suggested that at the national level, to achieve SOC neutrality in Spain and to change the situation from being a sink of SOC to a source, it is necessary to implement a SOC sequestration scenario with a rate of at least 5% SOC gain in the next 20 years. The results showed that average SOC sequestration in the business as usual (BAU) scenario would decrease at a rate of -0.12 T C ha^{-1} yr^{-1} between 2020 and 2040. However, with respect to this BAU scenario, we found an average increase of 0.027 T C ha^{-1} yr^{-1} under a sustainable management scenario for the period under study. Therefore, a sequestration potential of 5.59 T C yr^{-1} can be expected under the adoption of sustainable management practices in the next 20 years.

Our results contribute as a benchmark to assess the impact of global environmental change on SOC sequestration rates from a national perspective.

Keywords: RothC, relative carbon sequestration rates, scenario projections, sustainable management practices.

6.1. INTRODUCTION

Soil organic carbon (SOC) stocks are key indicators of soil health and soil fertility (McBratney et al., 2014). The role of SOC in climate change adaptation and mitigation has been recognized widely and validated in various studies, especially in the Mediterranean region, which is extremely vulnerable to global change (Acácio et al., 2017). Future climate predictions make the estimation of SOC sequestration relevant for establishing priorities for the research and implementation of public policies. Responding to this requirement, the Global Soil Partnership (GSP) Plenary Assembly in 2020 instructed the Intergovernmental Technical Panel on Soils (ITPS) and the GSP Secretariat to develop the Global Soil Organic Carbon Sequestration Potential map (GSOCseq map). The country-driven approach proposed was similar to that developed for the Global Soil Organic Carbon map utilizing standardized procedures but using the best available local data (FAO, 2020a)

However, the magnitude and rate of carbon sequestration in soils can vary greatly, depending on land use changes and practices, soil characteristics, vegetation, topography and climate, among other soil-forming factors and processes (Florinsky, 2012). In Spain, many studies have suggested that land abandonment, the dominant form of land use change in recent decades, determines the SOC evolution during secondary succession following abandonment (Gabarrón-Galeote et al., 2015). The ongoing increase in temperature, drought periods and expected heavy rainfall events (Pachauri et al., 2014) could dramatically influence SOC stocks, especially during this abandoned land evolution (Kirschbaum, 2006). Therefore, it is a challenge to identify relevant areas in Spain with great potential for soil carbon sequestration.

A large proportion of these carbon emissions are from soils due to intensive high-yield agricultural systems, forest conversion to agricultural or pasture lands, and current climate trends (López-Garrido et al., 2014). Agricultural lands occupy 33% of Spain, with large heterogeneity in terms of agricultural classes and management, where unsuitable management practices have led them to moderate or high degradation. Several studies have demonstrated both enhanced SOC sequestration and increased soil fertility by the application of improved agricultural management practices, such as conservation tillage recommended for Mediterranean regions (Haddaway et al., 2015). Specifically, the Global Soil Partnership developed the Voluntary Guidelines for Sustainable Soil Management (VGSSM) as a reference, providing general technical and policy recommendations on sustainable soil management (SSM). The enhancement of soil organic matter content through these SSM practices is more easily implemented in agricultural systems since they refer to the use of organic amendments, managing crop residues, practicing organic farming, integrated soil fertility management, implementing crop rotations or crop mixes, or reducing- or no-tillage practice (FAO, 2017a). The application of these practices has shown more than a 50% SOC sequestration increase (Jebari et al., 2018) or rates ranging from 0.15 to 0.32 $T \cdot ha^{-1}$ $year^{-1}$ after no-tillage practices on rainfed arable land (Álvaro-Fuentes et al., 2012b).

Forests represent a resource for counteracting hydrological problems and for helping mitigate the greenhouse effect. In particular, Spanish forest ecosystems present more carbon storage per hectare than other land uses, with a soil carbon rate of approximately

40-50% of the total ecosystem (Charro et al., 2008). With more than 18 million hectares of forests, representing nearly 35% of the national surface, Spain is the third country in terms of forest area within the European Union, contributing to approximately 12% of the total European forest areas (FAO, 2020a). To enhance the C sink forest capacity, similar adoptions of SSM practices in croplands and grasslands can be adopted in forest. In this sense, in 2003, the Intergovernmental Panel on Climate Change (IPCC) published a Good Practices Guide for Land Use, Land Use Change, and Forestry (Penman et al., 2003). The report provided methods for estimating, measuring, monitoring and reporting changes in carbon stocks and anthropogenic greenhouse gas emissions from LUCLUF activities. Although the afforestation of agricultural land was included among these good practices, soil processes occurring after farmland afforestation are not well known, especially in semiarid Mediterranean environments (Muñoz-Rojas et al., 2015; Ruiz-Sinoga et al., 2011).

Other emerging initiatives are focused on sustainable forest management (SFM) related to modeling the forest carbon sequestration rate as a result of specific management actions. Actions such as the increase in biomass production through forest treatments (e.g., cuttings, pruning or thinning), appropriate management of plant residue decomposition, enhancement of vegetation structure or planning of forest production could potentially maximize forest C sequestration. Some studies argued that the rate of C sequestration was approximately 5 $T \cdot ha^{-1}$ $year^{-1}$ for different pine forests (Navarrete-Poyatos et al., 2019) or that a 60% SOC increase occurred after moderate thinning practices over 13 years (Navarro-Cerrillo et al., 2018).

Even though Spain, Poland and Germany accounted for more than half of the net reduction in greenhouse gas (GHG) emissions in absolute terms in the EU in 2019, Spain was ranked as the 5th emitter of GHG in the EU inventory in 2019 (EU NIR, 2021). In the LULUCF sector, the main increase in absorption was linked to crops (+6.2%), and there was a reduction in absorption in forest activity and produce (-9.5%). Nevertheless, global soil recarbonization initiatives in different land use areas by the implementation of both sustainable soil and forest management practices, henceforth referred to as suitable management (SM), are still needed in Spain.

In Mediterranean Spain, previous efforts have modeled SOC sequestration using process-based models (e.g., the Century model) from the adoption of SM practices in agricultural land. Several authors have already applied this method at the plot (Álvaro-Fuentes et al., 2012b; Nieto et al., 2010) and regional (Álvaro-Fuentes et al., 2012a; Pardo et al., 2017) scales but without spatially explicit information. Jebari et al. (2018) applied the RothC simulation model during the 2010 to 2100 period, including spatial mapping, to evaluate the impact of climate change on SOC sequestration at the regional level. Despite the general increase in SOC stocks is shown under all climate change scenarios, the SOC sequestration rates were smaller than those under baseline conditions. These low SOC sequestration rates were associated with the greatest decline in precipitation and rise in temperature. Related to forest areas, few studies have relied on site-specific surveys to evaluate SOC changes in afforested land when croplands are abandoned. These results suggested higher mean SOC values in forest plantations absorbing carbon for a few decades until reaching an equilibrium state after some 100 years (Segura et al., 2016; Smith, 2004).

To our knowledge, there have not been many studies investigating SOC dynamics from a national perspective. To contribute to determining the current state of global SOC sequestration, a country-driven approach using standardized procedures is necessary. This work was based on the technical specification guideline provided by GSP FAO (2020), which was based on a well-known process-based model to represent and simulate SOC change (e.g., RothC). Our main objective was to generate a baseline to benchmark SOC sequestration rates in peninsular Spain to identify areas with high potential for soil carbon sequestration and hot spots of soil carbon losses to support sustainable management practices or soil conservation efforts. We described a first attempt to model SOC with a national perspective using a process-based approach over different land uses. Specific objectives were 1) to estimate SOC evolution under business as usual (BAU) practices in the next twenty years (2040); 2) to estimate the absolute SOC sequestration of a sustainable management (SM) scenario with a 5% increase in organic matter inputs; and 3) to calculate the differences in SOC sequestration between the BAU scenario and SM scenario (relative sequestration rates - RSR), as well as the differences between the SM scenario in 2040 and the SOC stocks in 2020 (absolute sequestration rates - ASR).

6.2. MATERIAL AND METHODS

6.2.1. Study Area

Peninsular Spain has an area of 491,258 sq. km. Although the agricultural surface area used has decreased in the last 20 years, the land cover in Spain reflects a predominantly rural landscape (MAPA, 2020). Agricultural lands occupy 50% of the surface, where heterogeneous agricultural systems of annual and permanent crops coexist and are distributed throughout the territory. Forest areas with natural vegetation and open spaces occupy 47.1% of the surface (approximately 23 Mha of forest area), distributed among wooded (26.8%), not wooded (17.4%) and spaces with little or without vegetation (2.36%). Therefore, the treeless open area is almost 10 Mha. The humid zones and water surfaces account for 0.9% of Spain, and the artificial surfaces account for 2.1% of the total surface area.

Spain is one of the most climatically diverse countries in the world, with four main Köppen climatic zones on the Iberian Peninsula (AEMET IPMA, 2011) from a humid to semiarid climate (Figure 6.1). The dominant climate extends to the Mediterranean coast and southern inland areas and has a warm temperate climate with dry, warm summers and moderate, wet winters. The maximum average temperature is 30 °C, and the rainiest month is above 70 mm. The vast central plateau has a similar climate but with a more continentally influenced climate with hot, dry summers and cold winters with temperature values ranging from below 0 °C to 35 °C. The mountains surrounding the plateau have higher rainfall and often experience heavy snowfall in winter. The western and northwestern coastal areas have an oceanic climate characterized by relatively mild winters and warm summers. Rainfall is generally abundant, exceeding 1,000 mm, and temperatures vary from 9 °C to 21 °C. A semiarid climate is predominant in the southeastern part of the country, where the dry season continues beyond the end of summer and the vegetation is less dense.

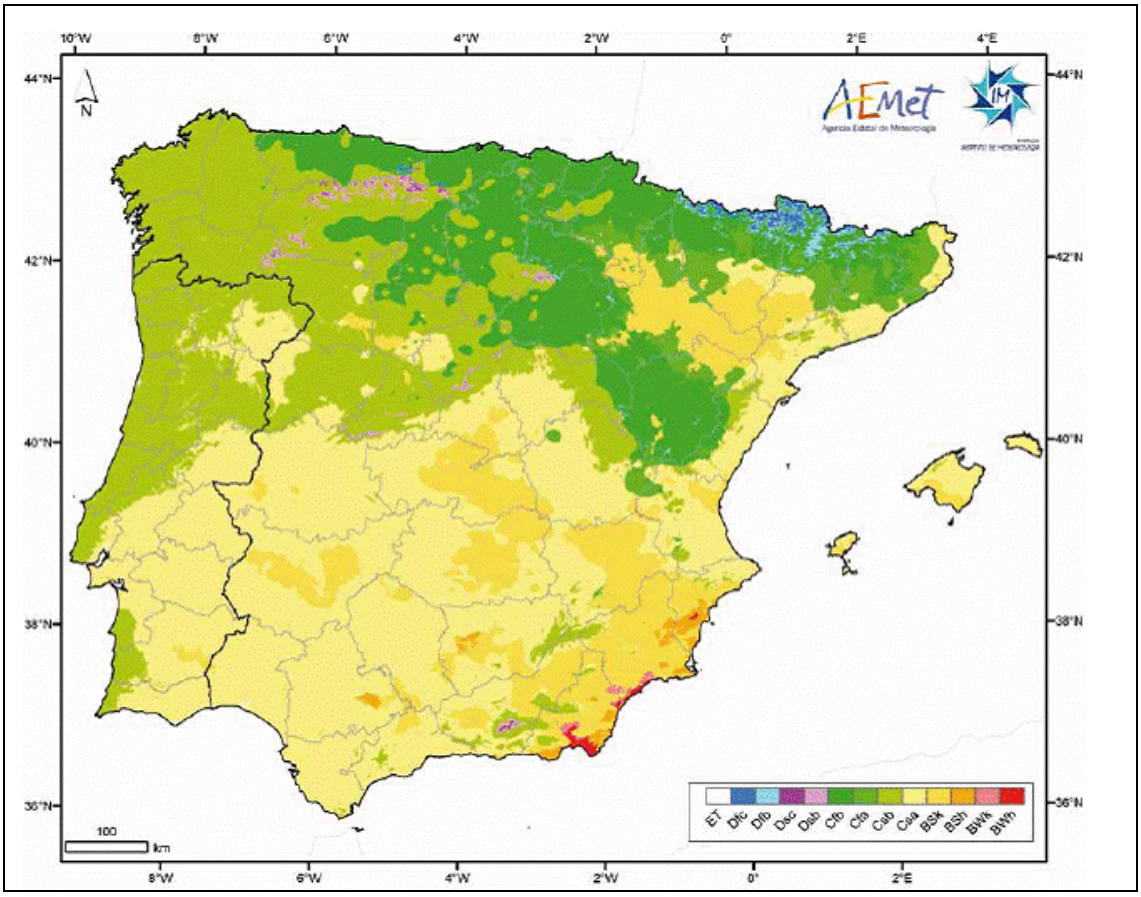


Figure 6.1. Köppen-Geiger climate classification for the Iberian Peninsula and Balearic Islands (Source AEMET 2011). ET= cold, tundra; Dfc= cold, no dry season, cold summer; Dfb= cold, no dry season, warm summer; Dsc= cold, dry season, cold summer; Dsb= cold, dry season, warm summer; Cfb= temperate, no dry season, warm summer; Cfa= temperate, no dry season, hot summer; Csb= temperate, dry season, warm summer; Csa= temperate, dry season, hot summer; BSk: arid, steppe, cold; BSh: arid, steppe, hot; BWk: arid, desert, cold; and BWh: arid, desert, hot.

Soils in Spain developed under a wide range of temperatures and moisture regimes, which implies a complex and diverse distribution. According to the "Keys to Soil Taxonomy" World Reference Base for soil resources (WRB-IUSS, 2014), the dominant soils are Calcisols Inceptisols (Calcixerepts and Haploxerepts) covering most of Spain, with the exception of the north and west of peninsula. These soils, formed over predominantly calcareous parent material, are present in areas with water stress. Due to its high mineralization rate, they are vulnerable to organic matter decline and preserving the vegetation cover is therefore essential (Bot and Benites, 2005). In the northern Spain, the most common soils are humic Dystroxerepts Umbrisols, typical of cold and humid climate conditions, with little or no moisture deficit and rich in organic matter. The west of the country is occupied by Entisols Regosols, very weakly developed mineral soils and extensive in eroded lands, in particular in arid and semi-arid areas and in mountain regions (lithic Xerorthents). Regosols Entisols are frequently found with AlfisolsLuvisols, and Inceptisols Umbrisols or Leptosols. Fluvents isols and Vertisols are predominant in fluvial areas, characterized by a high proportion of swelling clays (De la Rosa et al., 2001). Figure 6.2 depicts a selection of the dominant soil morphologies on the peninsula, which present limited depths.

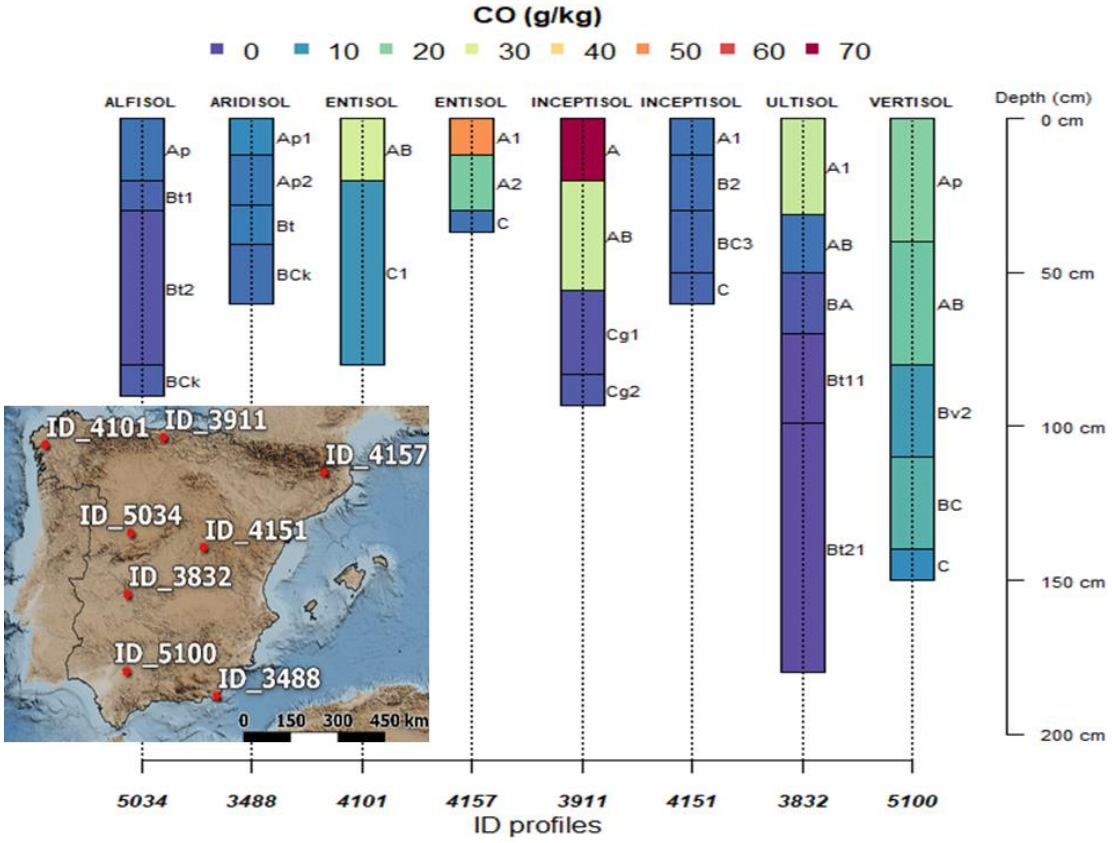


Figure 6.2. Profile plots and the spatial location of a representative soil profile selection of the peninsular Spain. Source: soil database compiled for this study.

Spain has complex topographic and orographic patterns characterized by a cluster of rugged mountain systems and high plateaus ranging from the Atlantic Ocean to the Mediterranean Sea. The center of the Iberian Peninsula is dominated by a vast plain, with elevations from 600 to 760 m, which slopes gently to the west. Undeveloped hills and high, rugged mountains extend inside and surround the plateau, reaching 3,478 amsl at the highest peak on the peninsula. This complex topography has resulted in a complex mosaic of soil types and soil covers. One emergent challenge is to accurately describe the spatial and temporal distribution of SOC across the complex variability of topographic and soil types (Verheye and la Rosa, 2005).

6.2.2. General Methodology

The basic process of SOC sequestration (SOCseq) in the terrestrial biosphere involves the transfer of atmospheric CO_2 into plant biomass and the conversion of biomass into stable SOC through the formation of organomineral complexes (Lal et al., 2018). Thus, estimates of SOC stock are the baseline used to assess SOCseq potential.

According to the FAO (2020b) methodology (Figure 6.3), we projected SOC stocks at 0–30 cm depth over a 20-year period. We assumed this period to be the default period, during which SOC stocks are approaching a new steady state, to be able to compare results among regions and countries and with other estimation methods (e.g., IPCC, 2006 Tier 1-2; IPCC, 2019). To infer SOC stocks under future conditions, process-oriented

multicompartment model such as RothC (Coleman and Jenkinson, 1996) was used to simulate changes in production systems.

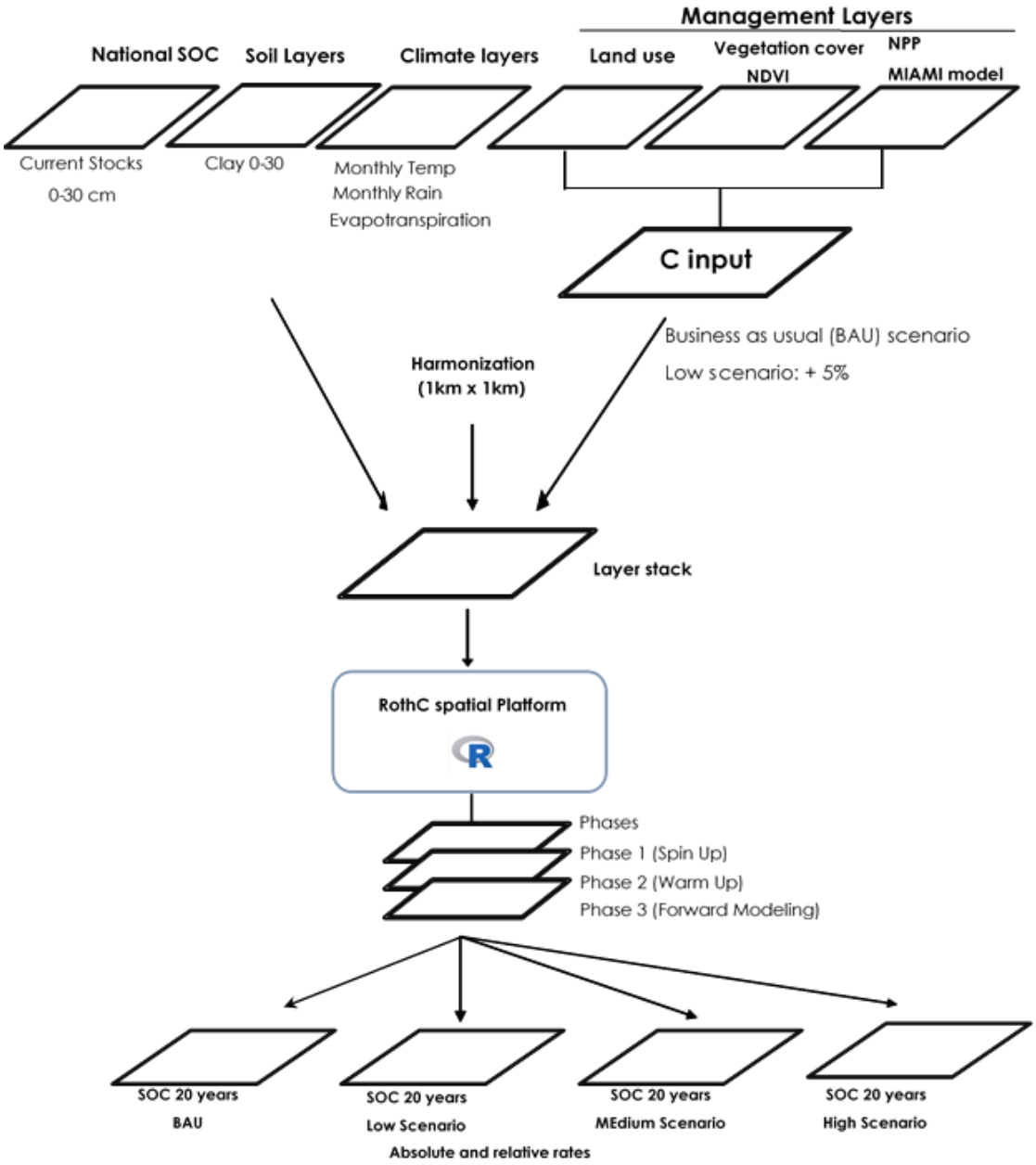


Figure 6.3 Simplified methodological outline to estimate the soil organic carbon (SOC) sequestration mpas adopted from FAO, 2020a.

This model is used to predict SOC dynamics based on four active conceptual C pools or compartments (Figure 6.4), decomposable plant material (DPM), resistant plant material (RPM), microbial biomass (BIO) and humified organic matter (HUM), that change in size via decomposition rates and stabilization mechanisms (Stockmann et al., 2013). In Spain, the use of simulation process-based models such as RothC has become popular in recent decades to predict the potential of SOCseq in different land management scenarios, site-specific levels, or periods of time (Nieto et al., 2010).

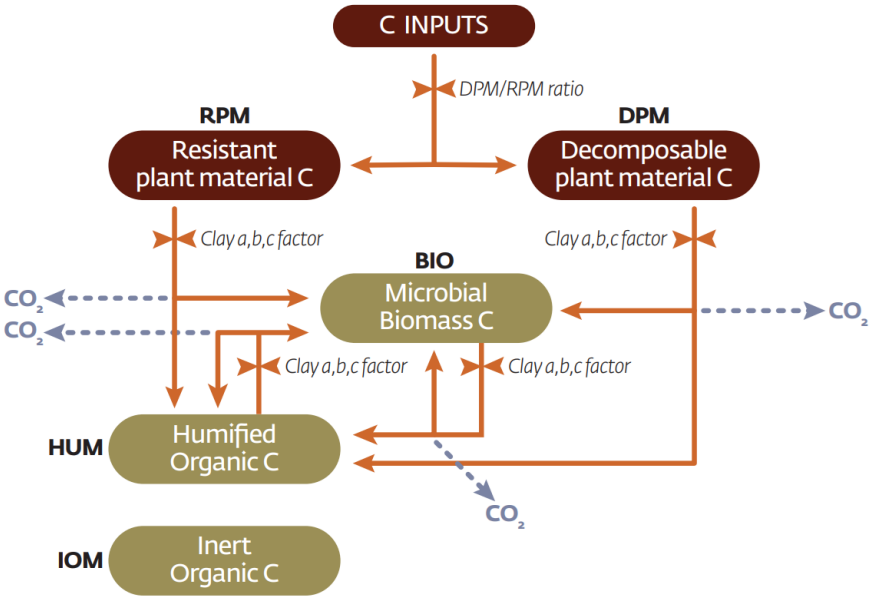


Figure 6.4. Structure pools and flows of carbon in the RothC model, including major factors controlling the fluxes (a = multiplier for effects of temperature, b = multiplier for effects of moisture, and c = multiplier for effects of soil cover; DPM/RPM = decomposable/resistant plant material ratio). *Source: FAO (2020), redrawn from Coleman and Jenkinson (1996) and Falloon and Smith (2009).*

SOC stocks were projected under two scenarios: BAU and increase in 5% C input from the adoption of a SM strategy (i.e., both SSM and SFM). The BAU scenario refers to an unchanging management of land use, applying the current implementation (i.e., time =0 or 2020). Compared to this scenario, SM practices (in both agricultural and forest systems) can potentially increase the relative effect on C inputs, considering 5% SOC gains as default values to test potential. This hypothetical scenario was considered the low SSM scenario (SSM1) in the guidelines for global sequestration potential (FAO, 2020a).

The resulting SOCseq estimates were expressed based on the definition of SOC baseline stocks and time toward a new equilibrium state. We referred to 'absolute SOC sequestration' (SOC_{ABS}) as the change in SOC stocks over time relative to a base period (or reference period, t0) and to 'relative SOC sequestration' (SOC_{REL}) as the change in SOC stocks over time relative to the BAU scenario. Both estimations were determined by the following equations (1 and 2):

$$\triangle SOC_{ABS}(T \cdot C \cdot ha^{-1}) = SOC_{SM/BAU\ t} - SOC_{t0} \quad (1)$$

where SOC SM/BAUt refers to the final SOC stocks after a 20-year period (year 2040, under the business as usual or SM practices) and SOC_{t0} refers to the initial or base period SOC stocks.

$$\triangle SOC_{REL}(T \cdot C \cdot ha^{-1}) = SOC_{SM\ t} - SOC_{BAU\ t} \quad (2)$$

where SOC_{SMt} refers to the final SOC stocks after a 20-year period of implementing SM practices and SOC_{BAUt} refers to the final SOC stocks after a 20-year period under business as usual (BAU) practices.

The projected modeling approach considered three different modeling phases: spin-up, warm-up and forward phases (Smith et al., 2006, 2005, 2007). In the spin-up phase, the initial conditions for the future simulations were determined, and the C input estimates over the years for each land use class were adjusted. Thus, an iterative process assessed the size of the four soil compartments by balancing the soil C fluxes based on the RothC model scheme. The mean climatic data (1980-2000), clay contents and land use conditions were considered constant in this process. The aim of the warm-up phase (or short spin-up) was a temporal harmonization of the SOC stock because the soil samples used to generate this map were collected over different decades (i.e., 1970s to 2000s). Since the SOC stock map referred to the year 2000, the RothC model was run for the 2001-2020 period using year-to-year climatic conditions to account for the climatic variation effect, as well as clay contents, SOC pool stocks and land use. In the third step, the forward phase, SOC stocks were simulated from 2020 (t=0) to 2040 (t = +20) for the BAU and SM scenarios.

The absolute and relative SOCseq were estimated following the previous equations for each grid cell of 1 km of spatial resolution across Spain (Figure 6.5).

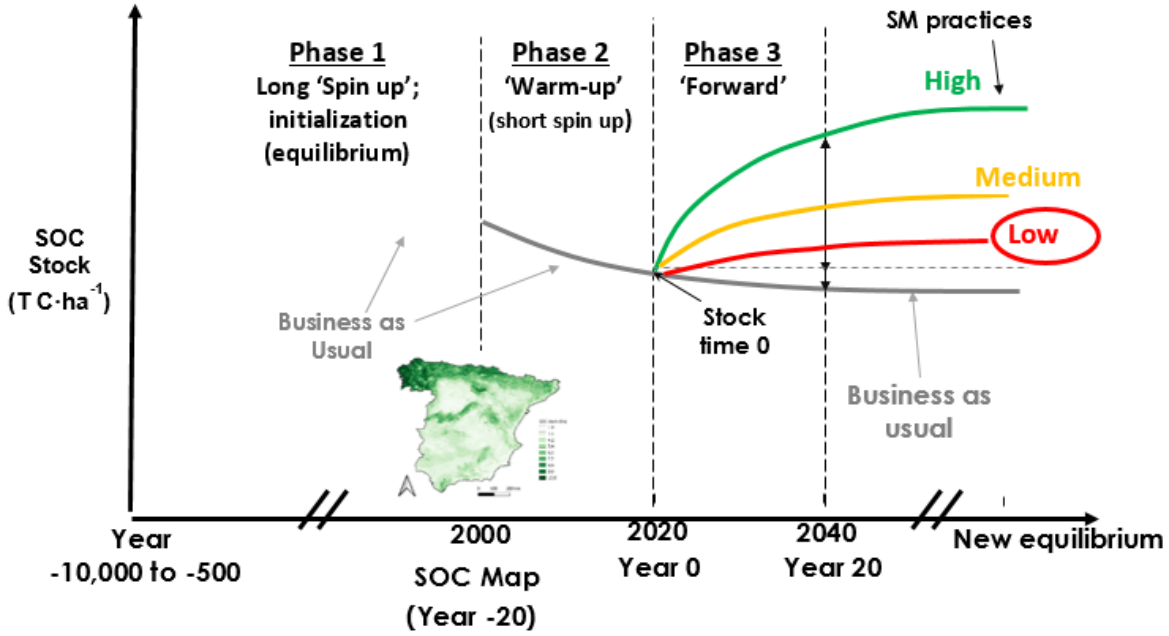


Figure 6.5 Soil organic carbon (SOC) stocks simulated in the different phases according to the proposed general modeling procedure. *Source: FAO (2020), based on Smith et al 2006; 2008; Gottschalk et al 2012.*

We augmented this methodology when needed with an automated interpolation approach based on ensemble learning for removing the presence of spatial artifacts associated with the use of contrasting granularities in the input layers (e.g., the target spatial support is 1 km, but climate datasets are provided in grids of approximately 50 km of spatial support). This interpolation approach is fully described in previous work (Hengl et al., 2021).

To generate the national SOCseq maps a spatially explicit version of the RothC model was required. We used a spatialized version of the model developed by the GSP Secretariat using open source R-environment, based on the SoilR package developed by Sierra et al. (2012). We parallelized the algorithm of the process to achieve a faster processing runtime and avoid .computational limitations.

6.2.3. Input Data Layers

The minimum data requirements for the RothC model are climatic, soil and land use datasets (Table 6.1). When national data were not available, we used globally available datasets following the recommendations by the FAO (2020a). The target resolution of the maps is 1x1 km. The current SOC pool represents the first 0-30 cm of soil depth.

Table 6.1. Input data used for the RothC model requirements.

Climate Data	**Soil Data**	**Land Use- Management Data**
1. Monthly rainfall (mm) 2. Average monthly mean air temperature (°C) 3. Monthly open pan evaporation (mm)/ evapotranspiration (mm)	4. Total initial 0-30cm SOC stocks (t C ha^{-1}) 5. Initial C stocks of the different pools (t C ha^{-1}): DPM, RPM, BIO, HUM, IOM 6. Clay content (%) at simulation depth.	7. Monthly Soil cover (binary: bare vs. vegetated) 8. Irrigation (to be added to rainfall amounts) 9. Monthly Carbon inputs from plant residue (aboveground + roots + rhizodeposition), (t C ha^{-1}) 10. Monthly Carbon inputs from organic fertilizers and grazing animals' excretion (t C ha^{-1}) 11. DPM/RPM ratio, an estimate of the decomposability of the incoming plant material

The modifications applied to the original implementation proposed by the FAO for the development of this baseline version of the map rely on:

1- Soil datasets: We used for model initialization the specific SOC stock map (T/ha) of the first 0-30 cm of soil depth estimated in chapter 5 of the thesis. We rescaled the SOC stock map from 90 m to 1 km pixel resolution to fit to the input covariates. Even though

the soil samples used to generate this map were collected mainly over the 1970s to 2000s, we assumed that the stock map corresponded to 2000, the reference base year.

2- Land use data: We used the Corine Land Cover 2012 for the SOCseq modeling, where a reclassification to FAO land use classes (Table 6.2) was performed (Figure 6.6)

3- Land management, C inputs scenarios:

- Only the standard C input increase scenario in +5% was applied (SM).
- The residue quality (decomposable plant material/resistant plant material, DPM/RPM ratio) assigned to forestland was adapted to Mediterranean values found in the literature (Vicente-Vicente et al., 2017). The adopted value was 0.20.

Table 6.2. Reclassification of Corine land cover (CLC) legend to FAO land use classes.

ID	CLC Legend	ID	FAO Legend
1	Artificial surfaces	0	No Data
2	Non-irrigated arable land	12	Tree crops (permanent crops)
3	Permanently irrigated land	2	Croplands(arable land)
4	Heterogeneous agricultural areas	2	Croplands(arable land)
5	Agro-forestry areas	8	Sparse vegetation
6	Broad-leave forest	4	Tree covered
7	Coniferous forest	4	Tree covered
8	Mixed forest	4	Tree covered
9	Sclerophyllous vegetation	4	Tree covered
10	Natural grassland	3	Grassland
11	Dense shrub cover	5	Shrubs covered
12	Scattered shrub cover	5	Shrubs Covered
13	Sparsely vegetated areas	8	Sparse Vegetation

6.2.4. Uncertainties

The uncertainties for this map were calculated following the default implementation of the technical specifications of the FAO methodology. We used one standard deviation of variation around the median value of the SOC stocks as uncertainty values, which are associated with the variance of the input data layers in the RothC simulations.

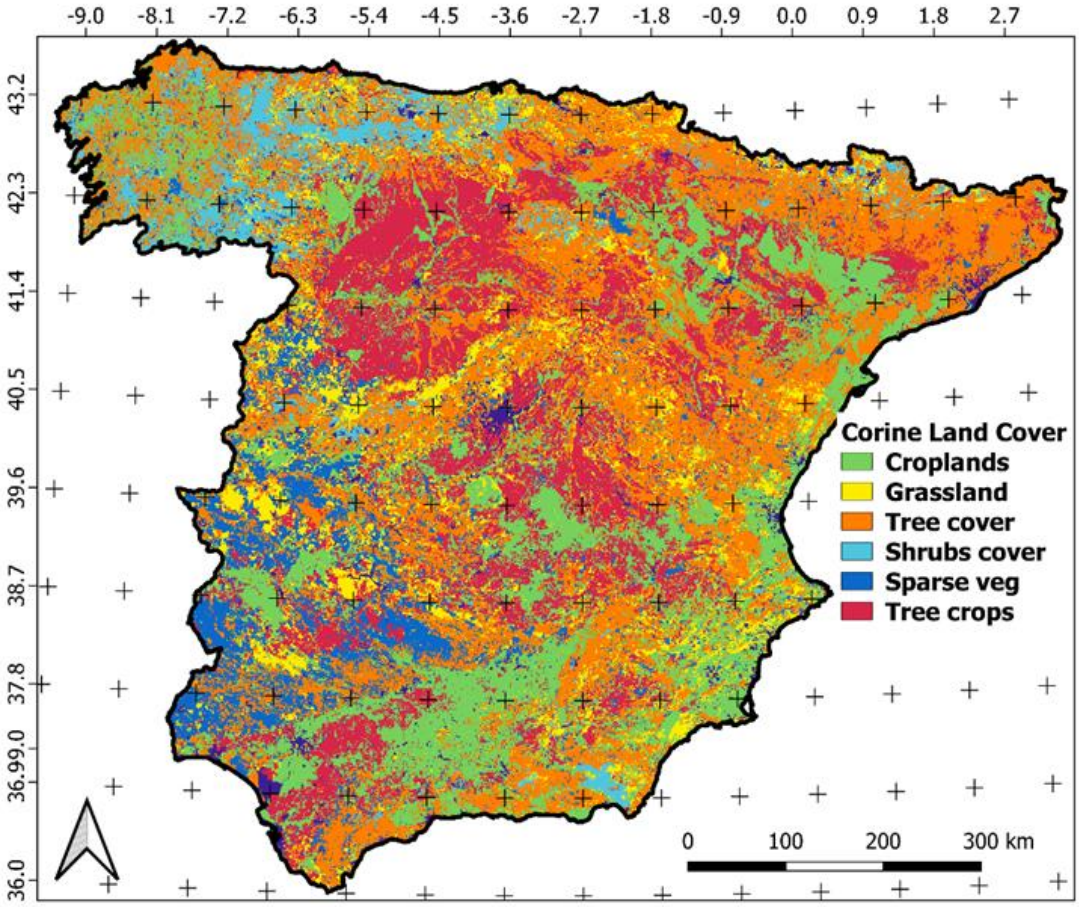


Figure 6.6 Distribution of Corine Land Cover (CLC) classes reclassified to the FAO legend on the peninsular Spain.

6.3. RESULTS

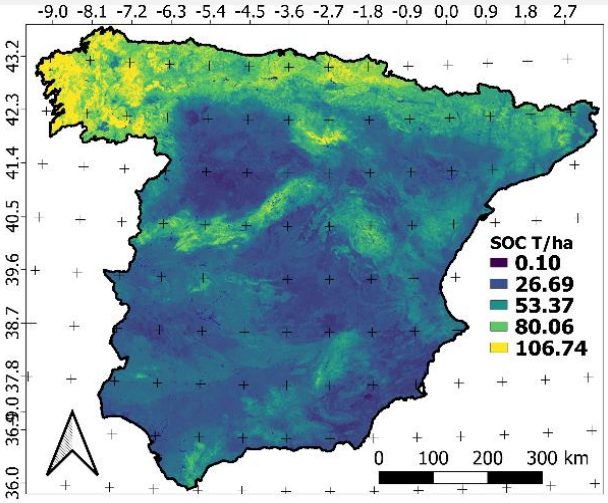
6.3.1. Summary and Spatial Prediction of SOC Sequestration Rates in Spain

We observed median values of 38.99, 38.87 and 39.42 T C/ha for the base SOC map (referred to 2020), the projection of business as usual (BAU) SOC map (referred to 2040) and the sustainable management scenario (SM) with 5% SOC gains, respectively (Figure 6.7).

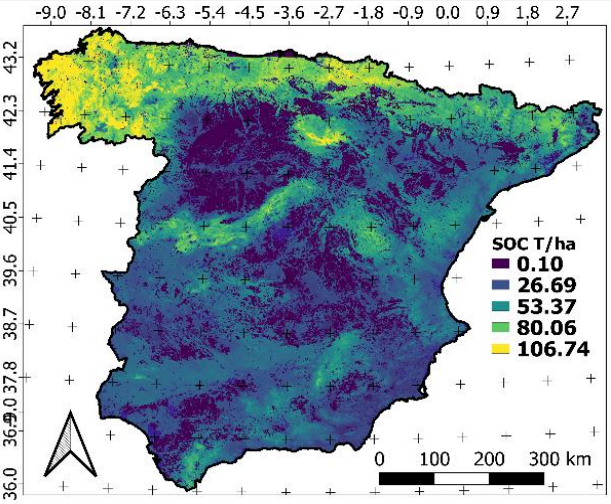
The general distribution of the spatial pattern of SOC was similar in all maps, differing in SOC storage. This general distribution depicted maximum values in the northeast and north, corresponding to the most humid zone on the peninsula, and in the main inland mountain ranges. The lowest SOC values corresponded to the vast arable areas of both the northern and southern plateaus, as well as the depression watershed in the southwest and the subarid area in the southeast.

For comparative purposes, we quantified 2.86 PgC of the base SOC map (year 2020) in the first 0-30 cm of the soil surface on the peninsular Spain at a 1 km pixel size. This number decreases to 2.79 Pg (i.e., -68,720,000 TC) of SOC in the BAU model in 2040. However, the results show that the SOC stock could recover to 2.89 Pg (i.e., +31,575,300 TC compared with the base SOC map) of SOC in the hypothetical scenario of 5% SOC gains by the adoption of SM practices.

(a) SOC base map (2020)



(b) Map of BAU scenario projection (2040)



c) Map of SM scenario projection (2040)

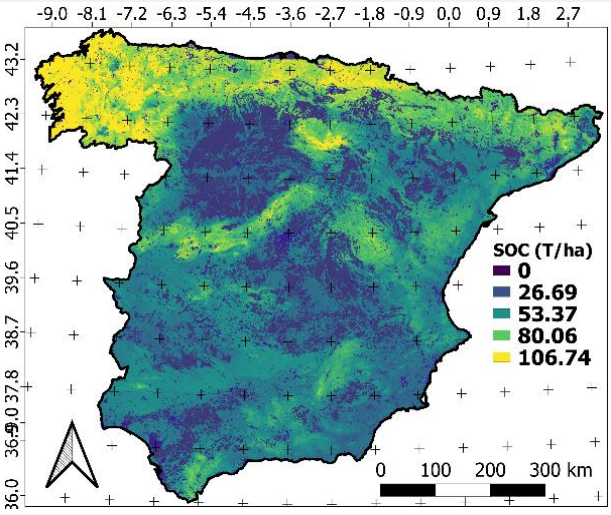


Figure 6.7 Soil organic carbon (SOC, T/ha) map at the reference period (2020) (a), SOC sequestration map of the business as usual (BAU) model (b) and the hypothetical scenario of 5% SOC gains from the adoption of a sustainable management (SM) strategy (c).

The absolute sequestration rate (ASR) was expressed as the change in SOC stocks over time relative to the base period SOC map (2020-2040). The global negative median value and the standard deviation (SD) of SOC losses showed a rate of -0.12 T C ha^{-1} yr^{-1} (0.97 SD) for the BAU projection (Figure 6.8a). This negative rate decreases to a median value of 0.43 T C ha^{-1} (0.98 SD) in 20 years from 2020 to 2040 (0.021 T C ha^{-1} yr^{-1}) considering the hypothetical scenario of low SOC gains (Figure 6.8b).

These maps depict the main results of the model, indicating areas that contribute to higher SOC losses in dark blue and, in yellow, areas with higher SOC sequestration potential to reduce C emissions from soils to the atmosphere under a low hypothetical scenario of sustainable management.

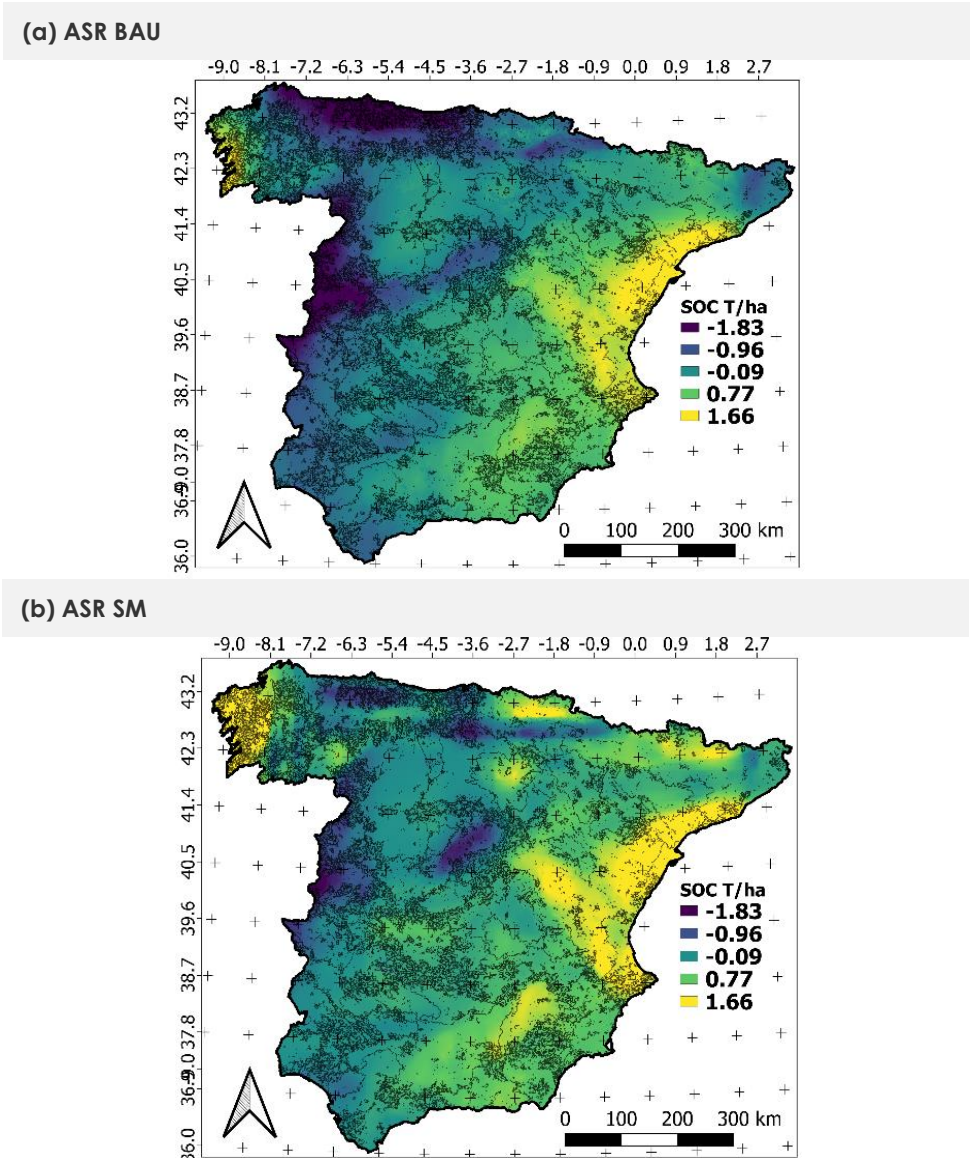


Figure 6.8 The absolute sequestration rate (ASR) of soil organic carbon (SOC, T/ha) in the business as usual (BAU) model (a) and the hypothetical scenario of 5% SOC gains from the adoption of a sustainable management (SM) strategy (b). The ASR was expressed as the change in SOC stocks over time (2020-2040). The land use cover is overlapped (gray lines).

We observed that the majority of values in the absolute sequestration rate for the map of SM scenario projection were positive versus the map of BAU scenario projection (Figure 6.9).

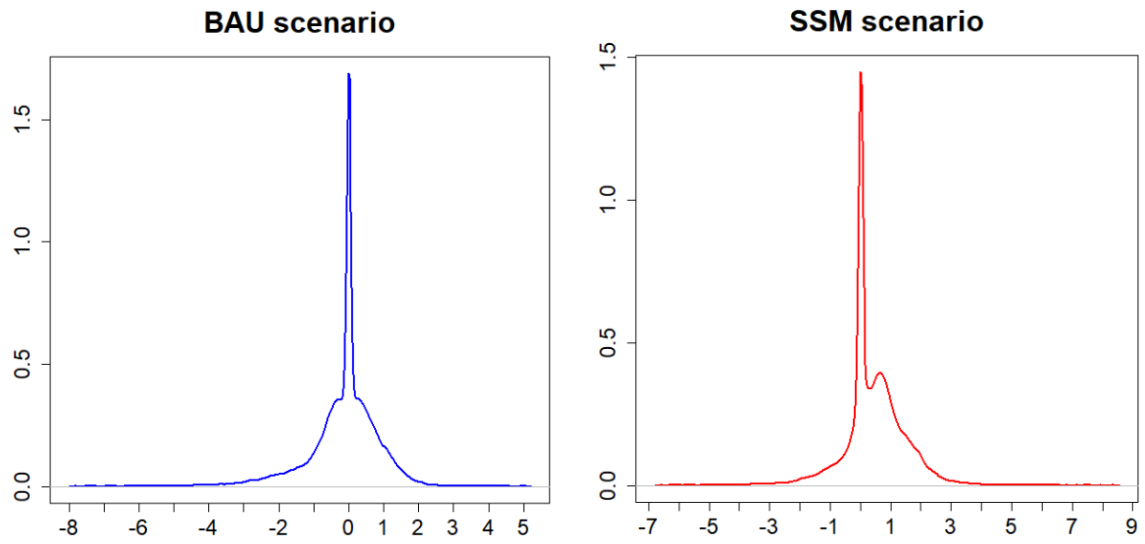


Figure 6.9 Density distribution for the absolute sequestration rate of soil organic carbon (SOC) data of the map of business as usual (BAU) projection (blue) and the hypothetical scenario of 5% SOC gains from the adoption of a sustainable management (SM) strategy (red).

The relative sequestration rate (RSR) of SOC projected under the hypothetical scenario of 5% SOC gains from the adoption of a SM was also depicted (Figure 6.10). The RSR was expressed as the change in SOC stocks over time (2020-2040) relative to the BAU scenario. The visualization of their values revealed the spatial pattern of SOC gains associated with the SM scenario. Although similar patterns on the maps of scenario projections can be observed (Figure 6.7 b, c), the eastern and central areas of the northern peninsula were clearly the areas with the highest SOC gains. These zones correspond to a mild temperature and the most humid region of Spain. The zones with the lowest SOC gains were the vast cultivated areas on both the northern and southern plateaus, as well as the main watershed depressions in the eastern and southern areas of the peninsula (Figure 6.10).

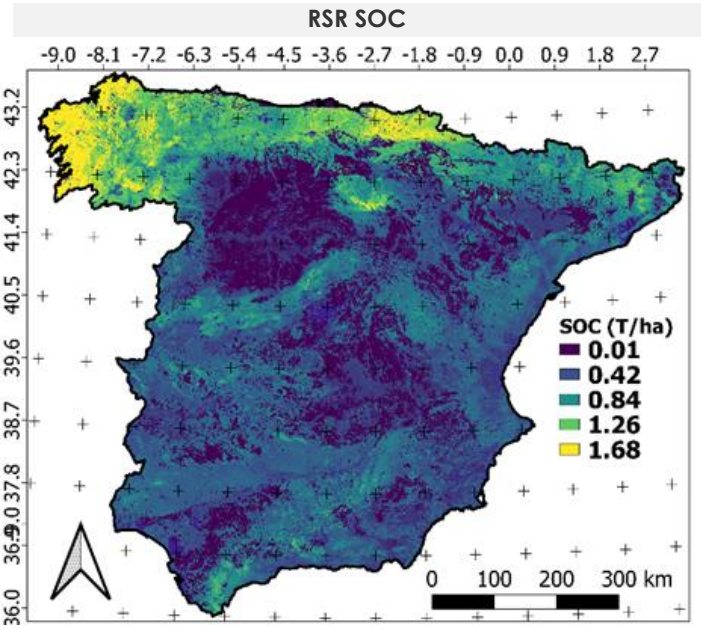


Figure 6.10 The relative sequestration rate (RSR) of soil organic carbon (SOC, T/ha) of the hypothetical scenario of 5% SOC gains from the adoption of a sustainable management (SM) strategy. The RSR was expressed as the change in SOC stocks over time (2020-2040) relative to the business as usual (BAU) scenario.

The analysis of these SOC projected values related to the main land use group, reclassified to FAO legend classes, revealed a positive absolute SOC sequestration in the BAU scenario only for croplands (0.002 T C·ha^{-1} yr^{-1}) (Table 6.3). However, sparse vegetation was the only land use with SOC losses in the SM scenario (-0.007 T C·ha-1 yr^{-1}). With regard to the relative SOC sequestration (RSR), the results showed a higher and positive rate for all land uses with respect to the ASR situation. In this situation, the mean SOC stock will increase on average by 0.82 T C·ha^{-1} in the total period, i.e., in 20 years. The highest potential SOC sequestration rate was obtained in the shrub cover (0.084 T C·ha^{-1} yr^{-1}), followed by tree cover and grassland (0.051 T C·ha^{-1} yr^{-1}) and tree crops (referred to as permanent crops) with the lowest values (0.016 T C·ha^{-1} yr^{-1}).

Table 6.3 Average absolute sequestration rate (ASR) and relative sequestration rate (RSR) of soil organic carbon (SOC) for business as usual (BAU) and sustainable management (SM) scenarios by land use groups.

		Average ASR		**Average RSR**
	Area (km^2)	**BAU** (T C·ha^{-1} yr^{-1})	**SM** (T C·ha^{-1} yr^{-1})	**SM** (T C·ha^{-1} yr^{-1})
Shrubs cover	6,480	-0.073	0.011	0.084
Grassland	23,535	-0.045	0.006	0.051
Sparse vegetation	23,949	-0.043	-0.007	0.036
Tree crops (permanent crops)	107,372	-0.004	0.012	0.016
Croplands(arable land)	114,582	0.002	0.044	0.044
Tree cover	196,834	-0.004	0.046	0.051
Average all land uses	482,751	-0.009	0.032	0.041

Table 6.4 shows the total absolute and relative SOC sequestration for each land use group and at the national scale (i.e., referred to the total surface). We observed the same situation in terms of the average ASR and RSR. Even though cropland was the only positive SOC sequestration land use in the BAU scenario projection (22 Gg C ·yr^{-1}), cropland could achieve over 500 Gg C ·yr^{-1} following the adoption of the SM strategy.

Notably, tree cover showed SOC losses at a rate of -82 Gg C ·yr^{-1} in the BAU scenario projection, achieving 987 Gg C ·yr^{-1} in the hypothetical scenario of 5% SOC gains from the adoption of the SM strategy (Table 6.4).

Table 6.4 Total soil organic carbon (SOC) sequestration (absolute and relative) for business as usual (BAU) and sustainable management (SM) scenarios by land use groups.

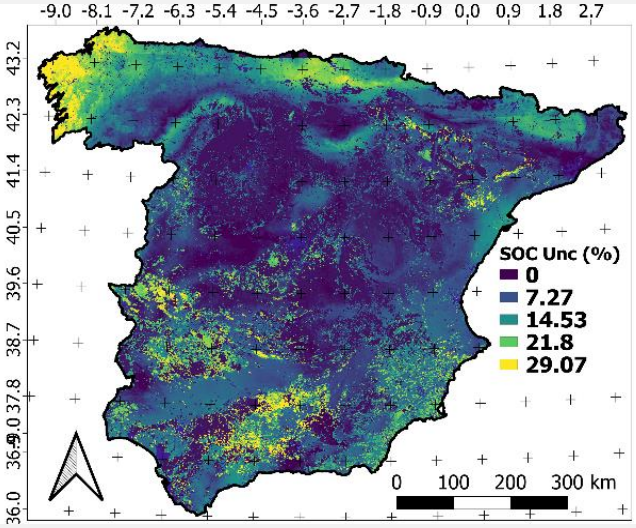
		Absolute sequestration		**Relative sequestration**
	Area (km^2)	**BAU** (GgTC ·yr^{-1})	**SM** (Gg C ·yr^{-1})	**SM** (Gg C ·yr^{-1})
Shrubs cover	6,480	-121	18	139
Grassland	23,535	-107	13	120
Sparse vegetation	23,949	-103	-16	87
Tree crops (permanent crops)	107,372	-40	127	167
Croplands(arable land)	114,582	22	478	500
Tree cover	196,834	-82	906	987
Total sum	482,751	-429	1,548	1,977

6.3.2. Uncertainties

We provided the uncertainties for the modeled SOC sequestration rates (Figure 6.11). We observed a general uncertainty in the ASR varying from 5 to 35%. A greater area of low uncertainty values was depicted in the SM scenario map. This suggests, as expected, that any adoption of a SM practice from the BAU soil management will have a positive impact on SOC sequestration and low uncertainties.

We highlighted that from these maps of scenario projections, some areas were governed by uncertainties near or over 50%, mainly corresponding to the very low SOC gain area in the southwestern region of the peninsular Spain.

a) ASR Uncertainties (BAU)



(a) ASR Uncertainties (SM)

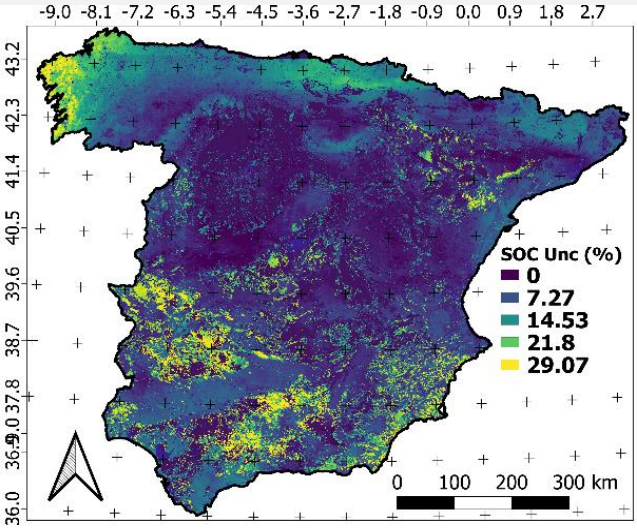


Figure 6.11 Uncertainties of absolute soil organic carbon (SOC) sequestration rates (ASR) expressed by a percentage for the business as usual (BAU) model (top) and the hypothetical scenario of SOC gains from the adoption of a sustainable management (SM) strategy (bottom).

6.4. DISCUSSION

We developed a national strategy to map SOC sequestration rates using a process-based approach selected by the FAO (2020) to generate a country-specific global SOC sequestration baseline. Although the forest area is growing in Spain with an increase of 1.2% between 2004 and 2018 (MITECO, 2019), the main interpretation of our results is that Spain is currently a SOC source, with an expected trend of SOC losses up to 2.5% of the total SOC stock in the first 30 cm of soil depth by 2040. Our results suggest that at the national level, a SOC sequestration scenario with a rate of at least 5% SOC gain in the next 20 years is needed to achieve SOC neutrality in Spain and to change the SOC source to a sink. This conclusion

was deduced from the analysis of the ASR for the map of SM scenario projection, in which the values were mostly positive (Figure 6.8).

These values were modeled using a single model (RothC model), however, we highlight that multiple country-specific characteristics (e.g., social and economic conditions) potentially affecting the SOC pool were not considered. Thus, the aforementioned results should not be considered as absolute numbers but rather a first approach to enable SOC frameworks from a national perspective, as well as to detect land uses, sites and regions with greater potential to increase SOC stocks after the use of SM. Based on the results of this first approach to calculating the SOC sequestration of peninsular Spain, the analyses showed that Spanish soils presented great potential for carbon absorption capacity by suitable management.

In previous studies, the amount of SOC was analyzed by soil type or land use (Calvo de Anta et al., 2020; Rodríguez Martín et al., 2016), and SOC sequestration rates were calculated for specific crops or natural landscapes (Navarro-Cerrillo et al., 2018; Nieto et al., 2010). While there has been interesting research quantifying SOC emissions (Antón et al., 2021; Bell et al., 2021), fewer studies have focused on predicting the potential for SOC sequestration at the regional, or even national, level (Jebari et al., 2018). The maps generated here are meant to provide a national perspective on the potential for SOC and a reproducible benchmark to improve our capacity to periodically generate SOC sequestration reports, including uncertainty.

As a preliminary approach, our results are consistent with those reported in the literature on SOC sequestration across Spanish soils. Álvaro-Fuentes et al. (2012) found an average stock interval of 60-80 T C ha^{-1} over northeastern Spain agroecosystems at 0-30 cm soil depth, and SOC stock changes ranged from 0.15-0.32 Tg C yr^{-1} (i.e., 0.04-0.08 T C ha^{-1} yr^{-1}). In this study, the SOC sequestration rates were calculated using the Century model approach calibrated for site-specific conditions. Additionally, studies across tree forest species showed different SOC sequestration rates in plots subjected to different thinning intensities than in control plots, increasing an average rate ranging from 20.41 to 100 T C ha^{-1} (Navarro-Cerrillo et al., 2018). Thus, our results represent a reliable contribution to support the development of SOC modeling and monitoring frameworks.

Modeling SOC increase scenarios, as in this study, represents a valuable tool for decision-making when developing SOC strategies. On the one hand, we highlight the pressing need to increase SOC sequestration rates to mitigate the potential negative effects of climate change (Jebari et al., 2018). On the other hand, this study will be the basis for identifying areas of intervention opportunity for supporting sustainable management practices focused on different land uses. In that sense, our results also contribute to identifying the limits of SOC sequestration due to the soil saturation of carbon (e.g., using clay content as a limiting factor for SOC storage), which can complement current and future digital SOC mapping and modeling efforts.

We observed that there is a high potential for boosting SOC sequestration in the humid regions of the north and northwest in tree cover areas and non irrigated croplands, which depend solely on rainfall as a source of water. The tree cover areas in the eastern part of

the peninsula were another great potential SOC sequestration area (Figure 6.8). While these wooded forest areas showed SOC losses in the BAU scenario projection (Table 6.4), they achieved over 980 Gg C·yr^{-1} in the hypothetical scenario of 5% SOC gains from the adoption of the SM strategy. These results can be used to establish priorities for the research and implementation of public policies since these crop areas present a high risk of future abandonment (Perpiña Castillo et al., 2020), and the forested areas represent a great source of SOC sinks.

We highlight the low sequestration rates in some of the northern and northwestern areas, with zones with a large storage of SOC in the maps of scenario projections (Figure 6.7). This is possibly because these soils in general may have a high initial carbon concentration and are located in optimal areas to sequester C due to their edaphoclimatic characteristics (% clay, temperature and humidity), and the SOC sequestration limit arises due to the saturation of carbon in the soil (Frolla et al., 2021). Moreover, the presence of spatial artifacts in the resulting maps associated with the use of contrasting granularities in the input layers must be considered. The target spatial support was 1 km, but climate datasets were provided in grids of approximately 50 km of spatial support. Even though these spurious results were removed (Hengl et al., 2021; Yigini et al., 2018), they could be an important source of uncertainty in the spatial trend of SOC sequestration.

Regarding the model evaluation, in the absence of SOC sequestration data for validating projections, the alternative presented by other countries was used an independent current SOC stock database (e.g., Frolla et al., 2021 in Argentina, or Reynoso et al., 2021 in México). As data required to quantify the accuracy of the estimates do not yet exist (projected SOC stocks for 2020-2040), the model could be evaluated by the trend of past events (ex post validation) compared to future events. The difference between a legacy SOC stock database and its trend projected 20 years in the future following current land use conditions can be estimated. This alternative provides a reproducible workflow and a benchmark in which to monitor the statistical accuracy of the SOC sequestration map in its future versions (FAO, 2020a). However, for the case of Spain, the lack of an updating, independent soil database limited the evaluation analysis of the model.

We emphasize that considering projected values derived from this work may be taken with caution, since more validation and calibration efforts are required for assessing the reliability of these SOC sequestration rates. This is a first step to explore the SOC sequestration rates of Spain from an institutional perspective. The next steps toward improving further versions must consider including more specific data in the simulation of sequestration processes. Related to soil properties and saturation in soil carbon, a deeper knowledge of mineralization rates and new input data (target and covariates) at a more suitable spatial resolution are needed. In addition, multiple country-specific characteristics (e.g., social and economic conditions), land use change data and different climatic scenarios should be included to achieve more realistic SOCseq maps.

6.5. CONCLUSIONS

The SOCseq maps generated in this study are a benchmark to explore the carbon sequestration at national level in the next 20 years, as well as to detect land uses, sites and regions with greater potential to absorb SOC stocks. The analyses suggest that peninsular Spain will be a source of carbon rather than a sink and, to prevent that situation, it will be necessary to enable management policies aimed at increasing SOCseq by at least 5% SOC gain as soil input. The modeling SOC increase scenario provided is a valuable tool for decision-making to develop SOC strategies.

This approximation has demonstrated that Spanish soils present a great potential for carbon absorption capacity as long as there is suitable management and there arer stable environmental conditions for at least the next 20 years. However, priorities need to be established since crop areas present a high risk of future abandonment and the forested areas represent a great source of SOC sinks.

Future considerations for exploring the SOC sequestration rates in Spain must improve the spatial resolution through more detailed spatial resolution and country-specific input layer in the estimation model. Moreover, multiple characteristics (e.g., social and economic conditions), land use change data, saturation in soil carbon and different climatic scenarios should be included to achieve more realistic SOCseq maps.

The implementation of the current SOCseq approach in RothC model by SoilR package enabled a multi-model ensemble approach. The parallelization of the process avoided computational limitations and achieved a faster processing runtime.

Capítulo 7. DISCUSIÓN GENERAL

Estrategias metodológicas para la modelización de la distribución espacial del carbono orgánico

Dada la importancia del carbono orgánico para la mitigación y regulación del clima, la planificación adaptativa resulta clave para su evaluación, seguimiento y gestión en los ecosistemas terrestres (Jobbágy et al., 2000; Wiesmeier et al., 2019). Por ello diversas iniciativas están surgiendo en los últimos años enfocadas a la contabilización precisa de los stocks de carbono en los ecosistemas terrestres (vegetación y suelo, principalmente), tanto a nivel global, europeo, nacional o regional. Sin embargo, existen varias limitaciones en la estimación de los stocks que dan lugar a una gran disparidad de métodos de cuantificación, monitoreo y documentación de esta información (Thamo and Pannell, 2016). Uno de los objetivos de esta es analizar diferentes estrategias de modelización del carbono para mejorar la calidad y precisión de la actual información disponible a escala nacional.

La mayoría de las técnicas actuales para la estimación del stock de carbono en la vegetación (biomasa aérea) se aplica en áreas pequeñas y con un alto nivel de resolución que dificulta su extrapolación a otras áreas o superficies más extensas. Esta tesis utiliza un enfoque novedoso para la extrapolación a escala nacional en el que combina la técnica "up-scaling" en dos etapas, integrando datos de sensores remotos de alta resolución espacial (LiDAR) con sensores ópticos hipertemporales de resolución espacial moderada (MODIS) y de cobertura global (capítulo 3). Así mismo, se comprobó que la incorporación de datos de sensores remotos en la modelización del carbono orgánico del suelo (COS) incrementó su capacidad predictiva (capítulos 4 y 5). Esto resultó muy relevante especialmente en zonas como el caso en estudio de topografía y climatología complejas con información muestral limitada, como también se demostró en estudios anteriores (Fathololoumi et al., 2020; Schillaci et al., 2017a). De igual forma, la integración de los resultados de la estimación de biomasa aérea a la modelización del COS mejoró la cuantificación del aporte de materia orgánica, resultando ser una covariable significativa en la modelización de su distribución espacial (capítulo 3 y 4).

La escasez de bases de datos disponibles para la estimación del stock de carbono tanto de la vegetación como del suelo resulta evidente en la mayoría de los estudios peninsulares (Domingo et al., 2018; Llorente et al., 2018). Este déficit de información muestral afecta también al proceso de validación ya que, dada la gran diversidad de métodos existentes, resulta necesaria su evaluación con información externa e independiente a la utilizada en la modelización. Sin embargo, la realización de muestreos de campo sigue siendo un reto debido al alto coste económico y complejidad operacional que dificulta la actualización de las bases de datos a lo largo del tiempo y, por tanto, su interoperabilidad (Smith et al., 2020; Vargas et al., 2017). En línea con investigaciones recientes, los resultados de esta tesis evidencian que uno de los actuales desafíos a los que se enfrenta la modelización del carbono es el diseño de la distribución de muestreo para optimizar su toma en campo (Zeraatpisheh et al., 2019).

En la fase de toma de datos en estudios de stock de vegetación basados en datos LiDAR combinados con muestreos sistemáticos para su calibración, a la vista de los resultados de esta tesis se considera esencial las siguientes consideraciones, observadas también en

trabajos anteriores (Fernández-Landa et al., 2018; Magnussen and Boudewyn, 1998). Estas consideraciones hacen referencia a una mínima diferencia temporal entre ambas informaciones (LiDAR-datos campo), una alta precisión en las coordenadas del centro de la parcela y que el área de la parcela de muestreo sea representativa del tamaño de píxel LiDAR (capítulo 3). Para el diseño de los muestreos del COS, los resultados obtenidos en esta tesis indican la necesidad de equilibrar el número de muestras de los intervalos de stock en base a su distribución normal, así como su representatividad espacial (capítulos 4 y 5). Mientras que el rango de valores de COS para el perfil medio varía entre 0.017 y 257.95 g/kg, con un valor medio de 23 g/kg, sólo el 8% de los datos representan valores entre 3-9 g/kg. Los valores por encima de 140g/kg tan solo representan el 2% de la base de datos de este estudio. La mejora en esta representatividad espacial y de los valores de COS incrementará la precisión de los modelos, ya que intervalos con baja representación de perfiles son modelados con menor precisión (Brungard et al., 2015; Hengl et al., 2007).

Otro factor clave en la modelización del carbono es la adecuada escala espacial de los mapas resultantes, esencial en la evaluación de impactos potenciales y la adopción de políticas de gestión (Vargas-Rojas et al., 2019). Sin embargo, los resultados de los análisis de la comparación de diferentes productos de COS, con diferentes resoluciones de píxel, no mostraron una clara consistencia respecto al tamaño mínimo de información (capítulo 4). La influencia de la escala resulta algo más evidente en los análisis de modelización del COS bajo los mismos supuestos pero a diferentes escalas, donde resoluciones con más detalle presentaron valores de precisión del modelo ligeramente superiores. Por lo que, no siempre una mayor resolución espacial se traduce en una mejor calidad de los modelos, estando influenciado por otros factores como la distribución y cantidad de los datos, estimación de covariables o métodos de modelización. Esto evidencia la dificultad en la estimación del COS, cuya heterogeneidad espacial, de escalas métricas o submétricas, no llega a ser totalmente representada en las estimaciones de superficie amplias (Minasny et al., 2017; Xiong et al., 2016).

La elección de los métodos para la generación de mapas de carbono desarrolladas en esta tesis ha sido condicionada por la limitación en la disponibilidad de datos. De entre estas, las técnicas de aprendizaje automático (machine learning) basadas en modelos no paramétricos mostraron una mayor capacidad para manejar un gran número de variables, minimizando el número de muestras necesarias para generar las predicciones (McBratney et al., 2003; Rodriguez-Galiano et al., 2015). En esta tesis se ha demostrado la necesidad de detallar los valores de las métricas de evaluación interna de los modelos, ya que se ha comprobado que la precisión de los mismos no está necesariamente relacionada con buenos resultados de calibración (capítulo 4). Para la evaluación de los modelos, además de los resultados de la validación interna, es especialmente importante detallar también el método estadístico empleado. La robustez de estos modelos viene determinada por la diferencia entre los criterios de información de calibración y validación. Dado que existe una amplia gama de algoritmos que resultan óptimos para la predicción espacial del carbono ecosistémico, en esta tesis se ha optado por la utilización de ensambles de modelos, planteando un enfoque novedoso para la selección del modelo final. En función de la variedad ambiental, resultante de las diversas combinaciones de factores predictores, y de los datos disponibles, la capacidad de predicción de los modelos resulta muy variable espacialmente respecto a la calidad de sus estimaciones dentro del territorio

modelado (Guevara et al., 2018). Por este motivo, el mapa final resultante se generó por combinación de varios ensambles de modelos, seleccionando las estimaciones más precisas obtenidas para cada píxel (capítulo 5).

Además, la implementación de estas técnicas de aprendizaje automático también ha permitido generar mapas continuos espacialmente explícitos de incertidumbre. Estos mapas proporcionan información esencial para interpretar los resultados y conocer la calidad de los mismos (Coulston et al., 2016). Permiten identificar patrones entre los valores de incertidumbre y las predicciones, estableciendo la aplicabilidad de los productos obtenidos. Así mismo, el uso de varias técnicas de selección de covariables ha permitido enriquecer todavía más la cuantificación del carbono, evitando el riesgo de redundancia estadística con el consecuente potencial de sobreajuste (Gregorutti et al., 2017). Las técnicas de selección planteadas en el capítulo 5 facilitaron una selección más objetiva de las covariables más relevantes para la predicción del carbono.

Los estudios relacionados con la estimación del carbono ecosistémico a nivel nacional son escasos, siendo más frecuentes los realizados para el suelo. De entre estos, la mayoría se centran en los 30 cm superficiales y en la variable de stock (TC/ha). En esta tesis, sin embargo, se incluye además la variable de concentración del COS (g/kg) y se considera la profundidad efectiva del perfil edáfico (capítulos 4 y 5). La importancia de analizar en primer lugar la variable concentración del COS es debida a que en las bases de datos disponibles este es el parámetro vinculado al carbono del suelo que se recoge originalmente. Además, la mayoría de las bases de datos presentan una carencia importante respecto a parámetros fundamentales para el cálculo de la variable stock de carbono. Esto puede llevar a la interpretación sesgada del stock en los distintos ecosistemas debido a la necesidad de extrapolar dichos parámetros, con el consiguiente riesgo de propagación de errores en la modelización final (Durante et al., 2019; Poeplau et al., 2017). En estos casos, y a la vista de los resultados de esta tesis, a menos que sea esencial para adaptar determinadas políticas medioambientales, se recomienda desarrollar mapas de concentración de COS en lugar de los datos frecuentemente sobreestimados de stock.

En el desarrollo del enfoque metodológico planteado en esta tesis ha prevalecido la utilización de programas de código abierto y datos públicos gratuitos, lo que fomenta la rentabilidad y extrapolación a otras zonas. Uno de los limitantes de estos procesos a escala nacional y alta resolución espacial es la capacidad computacional necesaria debido a la gran cantidad de datos empelados. En el diseño de los procesos realizados se ha priorizado reducir la limitación computacional y los tiempos de ejecución, desarrollando metodologías y procesamientos paralelos y distribuidos.

La dinámica del carbono en los ecosistemas y la integración de los stocks vegetación-suelo

Las predicciones climáticas futuras hacen que la estimación del secuestro de carbono en los ecosistemas sea esencial para establecer prioridades en políticas de reducción de GEI y gestión sostenible. Establecer modelos dinámicos de carbono capaces de predecir resultados espacialmente localizados bajo diferentes escenarios de gestión resulta fundamental (López-Garrido et al., 2014). En este sentido, desde la FAO se ha instado a los

países a la generación del mapa de secuestro de carbono orgánico del suelo (Soil Organic Carbon Sequestration Potential National Map, SCOseq) utilizando métodos estandarizados pero adaptando datos locales (capítulo 6). La integración de datos locales en el estudio dinámico, tal y como el mapa de COS stock (capítulo 5), de usos y coberturas de suelo o ratios de descomposición orgánica, mejoró la predicción del mapa final de SOCseq. Los análisis muestran que, suponiendo constantes las condiciones ambientales actuales durante los próximos 20 años, España Peninsular será una fuente de emisión de carbono en lugar de sumidero. Por lo que, para revertir esta situación será necesario establecer prácticas de gestión encaminadas a incrementar los niveles de entrada (inputs) de carbono en la dinámica del COS. Estas prácticas están más ampliamente detalladas para los sistemas agrícolas (FAO, 2017b), siendo la propuesta de esta tesis la incorporación de prácticas sostenibles encaminadas a maximizar la fijación de carbono en sistemas forestales.

Para la estimación del SOCseq se plantea la cuantificación del incremento de entrada de biomasa al suelo mediante la metodología propuesta en el capítulo 3. De esta forma, se puede generar información espacialmente continua mediante datos LiDAR. Esta información resulta detallada, fiable y fácilmente actualizable para ser incorporada a los datos de partida de la dinámica del carbono en la predicción de escenarios futuros.

Líneas futuras de investigación

Uno de los objetivos de esta tesis es contribuir a la generación de la mejor estrategia metodológica de la cuantificación del carbono terrestre a nivel nacional, asegurando su interoperabilidad global. Concretamente, se ha validado que la combinación de actualizados métodos estadísticos de aprendizaje automático con técnicas digitales y de teledetección, así como la automatización de tareas, permitió mejorar el conocimiento actual del carbono orgánico a una alta resolución espacial. Especialmente para una zona de clima y topografía complejos, y con una limitada distribución de datos como nuestro caso de estudio (Fathololoumi et al., 2020).

Sin embargo, existen limitaciones metodológicas respecto a las técnicas y datos utilizados. Los mapas de incertidumbre obtenidos en esta tesis ratifican que las zonas con mayor incertidumbre son aquellas con contenido de stock de carbono más bajo. Para el caso de la vegetación, estas zonas correspondieron a coberturas de bajo porte (es decir, cultivos, herbáceas, pastizales, matorral bajo), las cuales fueron peor estimadas debido a la propia limitación en las mediciones de los datos LiDAR utilizados respecto a este tipo de coberturas (Hancock et al., 2017). Respecto al stock del COS, las zonas de mayor incertidumbre correspondieron a aquellas con procesos de degradación más avanzados. Como se ha mostrado en trabajos anteriores (Brungard et al., 2015), la baja precisión en la modelización de estas zonas puede ser debido a la alteración en los patrones de relación de las covariables con el carbono en los suelos degradados. Así mismo, las diferencias de precisión entre las modelizaciones de las dos variables de carbono de suelo estimadas en esta tesis (es decir, concentración y stock) corroboraron la posible propagación de errores y la sobreestimación de los valores de stock (Durante et al., 2019; Poeplau et al., 2017).

Futuros trabajos deberían focalizar los esfuerzos en aumentar la información muestral nacional y mejorar la base de datos en estas zonas de mayor incertidumbre en la predicción del carbono ecosistémico y conseguir una representatividad homogénea del territorio. Concretamente para el carbono en suelo, a nivel nacional existe una gran cantidad de información fragmentada, dispersa y sin digitalizar. La recopilación, estandarización y puesta a disposición pública de dichos datos ayudaría a una mejor evaluación y estimación del carbono. La recuperación de estos datos permitiría completar la importante carencia respecto a los parámetros fundamentales para la estimación del stock del COS, tal y como se han demostrado en otros países como Chile (Pfeiffer et al., 2020) o Ecuador (Armas et al., 2021).

La integración de los reservorios de carbono de la vegetación y el suelo a nivel nacional es otro de los principales retos a los que se enfrenta el estudio del stock de carbono en los ecosistemas terrestres en España, todavía sin resolver. En esta tesis, el principal escollo que ha limitado la realización de esta integración ha sido la disponibilidad de datos de biomasa aérea, a nivel nacional, con una mínima diferencia temporal entre la información de datos LiDAR y datos de campo. Esto, unido a la gran capacidad computacional necesaria para el procesado de datos de todo el territorio peninsular, dio lugar a la imposibilidad de obtener un mapa final de carbono ecosistémico de España.

Los diferentes enfoques de los modelos de integración de los reservorios de carbono terrestres se pueden sintetizar principalmente en dos: los que enfatizan la determinación de las áreas de distribución potencial de la vegetación (modelos biogeográficos), o los que simulan los ciclos de carbono y nutrientes en los ecosistemas (modelos biogeoquímicos). De entre estos últimos, el modelo RothC (Coleman and Jenkinson, 1996) utilizado en esta tesis para la simulación de la dinámica del secuestro potencial del COS, resultó óptimo para satisfacer esta demanda de integración por su enfoque en el sistema planta-suelo (capítulo 6). A pesar de la propuesta de cuantificación de la entrada de biomasa al suelo mediante la metodología descrita en el capítulo 3, existen serias limitaciones respecto a su estimación nacional, tal y como se ha mencionado anteriormente. Por ello, más allá de estimaciones globales, actualmente no existe un mapa de biomasa nacional. Lo que supone un reto de futuro clave aún sin resolver en España. La integración de otras fuentes de información con mayor resolución espacial y temporal, como datos del satélite SENTINEL o GEDI, podrían ayudar a enfocar los esfuerzos de investigación en esta línea.

Las futuras líneas de investigación respecto al modelo RothC deberían focalizarse en adaptar su parametrización a nuestra zona de estudio. Tasas de mineralización específicas y de saturación del suelo, así como datos locales de entrada con una resolución más adecuada, mejorarán las predicciones del modelo. Además, se deben incluir múltiples características específicas del país (por ejemplo, condiciones sociales y económicas), datos de cambio de uso de la tierra y diferentes escenarios climáticos para lograr mapas de secuestro potencial de carbono ecosistémico más realistas.

Finalmente, tal y como sugiere Searle et al. (2021), los próximos pasos han de ir encaminados hacia la incorporación de estas evaluaciones de carbono terrestre en el análisis de los problemas de gestión. Esta información tiene la capacidad de anticiparse, con la consiguiente capacidad de transformación en el seguimiento y la previsión de los

impactos de las prácticas de gestión territorial. Para ello resulta fundamental la disposición pública de la información. Por lo que se trabajará en la visibilidad de los productos resultantes y de sus versiones mejoradas mediante su disponibilidad a través de servidores de libre acceso.

Capítulo 8. CONCLUSIONES GENERALES

8.1. CONCLUSIONES GENERALES

1. Los resultados de esta tesis aportan avances metodológicos significativos para la cuantificación del carbono orgánico en ecosistemas terrestres, mejorando la calidad y precisión de la actual información disponible a escala nacional. Estos avances hacen referencia especialmente a técnicas estandarizadas a diferentes escalas espaciales, así como a la aplicación de modelos dinámicos capaces de predecir los cambios de los stocks bajo diferentes escenarios de gestión. La limitada disponibilidad de los datos ha condicionado la elección de los métodos de modelización, así como la calidad final de los productos en la generación de mapas de carbono ecosistémico.

2. En esta tesis se generaron mapas de carbono basados en las mejores estimaciones posibles dentro de los límites del conocimiento científico y los recursos disponibles. Los productos estimados incluyen mapas de biomasa aérea (AGB) y de carbono orgánico del suelo (SOC). Respecto a la vegetación, se generaron dos mapas de biomasa (t/ha) para la Región de Murcia en base a diferentes fuentes de datos y a diferentes escalas espaciales: 25 m y 250 m generados con datos LiDAR y predicciones MODIS, respectivamente. Respecto al SOC, se generaron cuatro mapas a 90 m de resolución, más sus correspondientes mapas de incertidumbre asociadas, para España peninsular: mapas de concentración de SOC a 0-30 y 30-100 cm de profundidad; y mapas de stock de SOC a 0-30 cm y a su profundidad efectiva. Además, se generaron mapas con las proyecciones (2020-2040) del secuestro potencial del stock de SOC a 1 km de resolución, bajo diferentes escenarios de gestión: mapas de tasas de secuestro absoluto y sus incertidumbres (tC/ha), y mapa de tasas de secuestro relativo (tC/ha).

Bajo el compromiso de datos abiertos y accesibles, estos mapas se encuentran en un servidor de libre acceso en la siguiente dirección web: https://drive.google.com/drive/folders/1SykCBAbpHnUPHcCbguE34jiBfBIvOlnM?usp=sharing

3. La principal dificultad se ha encontrado en la limitada disponibilidad de datos de carbono a nivel nacional. En el caso de la biomasa, la diferencia temporal entre los datos LiDAR y, a su vez, con los datos de campo, dificultó la extrapolación a nivel nacional. Respecto a la información de los suelos, la falta de equilibrio en el número de muestras en determinados intervalos de stock, así como la representatividad espacial de los datos, resultaron ser más influyente en la precisión de las estimaciones que la densidad muestral.

Además, las bases de datos de suelos presentaron carencias respecto a los parámetros necesarios para el cálculo de stock. Los mejores resultados en la calidad de los modelos del SOCc respecto al SOCs, así como la sobrestimación del stock, se atribuyen a la propagación del error en la modelización debido a la extrapolación de dichos parámetros. Así mismo, la mayor parte de la información del SOC en España está actualmente normalizada a los 30 cm superficiales, comprobándose en esta tesis que se estaría subestimando el 25% del reservorio del carbono en los horizontes subsuperficiales (>30 cm).

4. La comparación de la modelización del SOC a diferentes escalas espaciales realizada en esta tesis reveló una mayor precisión en modelos de resoluciones con mayor detalle, si bien las diferencias entre las mismas no fueron muy notables. Esta falta de consistencia entre la resolución de la escala y la calidad de los modelos evidencia la dificultad en la

estimación del SOC. Esto puede estar relacionada con la alta heterogeneidad espacial del carbono, difícilmente captada en las estimaciones de amplias superficies incluso de alta resolución. Además, la inclusión en los modelos de covariables representadas a muy diversas escalas podría ser otra de las causas que influyen en la precisión de los mismos.

5. Las técnicas de aprendizaje automático (machine learning) utilizadas en esta tesis mostraron su capacidad para manejar un gran número de variables. La integración de estas técnicas junto con datos de sensores remotos incrementó la capacidad predictiva de los modelos de estimación del carbono, especialmente en estudios con información muestral limitada. La baja diferencia entre los criterios de información de calibración y validación obtenida en los modelos finales de estimación del carbono corroboró la robustez de los mismos.

6. El aporte de mapas de incertidumbre continuos y espacialmente explícitos resultó esencial para la correcta interpretación y fiabilidad de los resultados. La identificación de patrones entre los valores de incertidumbre y las predicciones permitió detectar áreas con estimaciones menos consistentes que precisan de una mayor intensidad de muestreo.

7. Se ha generado el primer mapa de secuestro potencial del SOC para España peninsular, bajo diferentes escenarios de gestión. Estos análisis manifiestan que, suponiendo constantes las condiciones ambientales actuales durante los próximos 20 años, los suelos serán una fuente de emisión más que un sumidero. No obstante, se demuestra que la realización de prácticas sostenibles que incrementen los niveles de ganancia de carbono orgánico en el suelo en al menos un 5%, revertirán esta situación para convertirse en sumideros, especialmente las zonas boscosas arboladas.

8. La integración de los reservorios de carbono de la vegetación y el suelo a nivel nacional es uno de los retos a los que se enfrenta el estudio del stock de carbono en los ecosistemas terrestres en España, todavía sin resolver. Las futuras líneas de trabajo han de focalizar los esfuerzos en la extrapolación nacional de las técnicas de estimación de biomasa aérea mediante la integración de otras fuentes multiemporales y de más alta resolución espacial. Así mismo, para el caso del carbono en suelo, sería necesario un esfuerzo nacional para la obtención de una base de datos patrimonial completa y armonizada. Hace falta incrementar el número de perfiles para una representatividad espacial adecuada del conjunto del territorio, completando la información necesaria para el cálculo del stock.

8.2. GENERAL CONCLUSIONS

1. The results of this thesis provide significant methodological advances for the organic carbon quantification in terrestrial ecosystems, thus improving the quality and precision of the information that is currently available at the national level. These advances are based on standardized techniques on different spatial and management scales, as well as the characterization of dynamic models to predict stock changes under different management scenarios. The limited availability of data has determined the selection of modeling methods and the quality of the final products to generate the carbon ecosystem maps.

2. In this thesis, the carbon maps were generated based on the best possible estimates within the limits of scientific knowledge and available resources. The estimated carbon products include aboveground biomass (AGB) maps and soil organic carbon (SOC) maps. Regarding the vegetation, two biomass maps (t/ha) were generated for the 'Región de Murcia' based on different data sources and at different spatial scales: 25 m and 250m for LIDAR data and MODIS-based prediction, respectively. Regarding SOC, four maps were generated at 90 m spatial resolution, as well as their corresponding associated uncertainties maps, for peninsular Spain: SOC concentration (SOCc) maps at 0-30 cm and 30-100 cm depths; and SOC stock (SOCs) maps at 0-30 cm and at its effective soil depth. In addition, projections of SOC stock sequestration potential map (SOCseq) were generated at 1 km spatial resolution under different management scenarios: absolute sequestration rate maps and its uncertainties (tC/ha), and relative sequestration rate map (tC/ha).

In agreement with the open free data commitment, these maps are available on an open-access server at https://drive.google.com/drive/folders/1SykCBAbpHnUPHcCbguE34jiBfBIvOlnM?usp=sharing

3. The main limitation was the low density of carbon data samples at the national level. In the case of the biomass, the time lag among LIDAR data, and also between the field-pots, complicated the extrapolation at the national level. Regarding the soil information, the lack of samples in certain values intervals, as well as the data spatial representativeness, became more influential than the sample density.

Furthermore, the soil databases showed a significant lack of basic parameters for the stock estimation. The best quality of the SOCc models with respect to SOCs, as well as the stock overestimation, are associated with the modeling error propagation due to the extrapolation of the aforementioned parameters. In addition, most of the SOC databases in Spain are currently standardized to the upper 30 cm. , proving in this thesis that it would be underestimating at least 25% of SOC if the subsoil carbon pool (>30 cm) is not considered.

4. The comparison of SOC modeling at different spatial scales performed this thesis showed a greater modeling precision when the resolution is more detailed, although the differences among scales were very low. This lack of consistency between the spatial resolution and model quality shows the difficulty in estimating SOC. Therefore, its short-range spatial heterogeneity is hardly captured in large surfaces, even at high-precision scales. Furthermore, the integration of covariates with diverse spatial resolution in the carbon modeling could also influence the model accuracy.

5. The machine learning techniques used in this thesis showed a great capacity for managing a great number of variables. The integration of these techniques with remote sensor data increased the predictive capacity of the carbon models, especially in areas with limited samples. The slight difference between the information criteria for calibration and validation sets obtained in the final carbon estimates confirmed their robustness.

6. The contribution of spatially explicit uncertainty maps was critical for the correct interpretation and validity of the results. Identifying patterns between the uncertain and predicted values enabled the detection of areas with lower accuracy estimates, which require a higher sampling density.

7. In this thesis, the first map of SOC stock sequestration potential for peninsular Spain has been generated, under different management scenarios. These analyses demonstrate that, as long as the current environmental conditions remain constant for the next 20 years, the soils will be a source of emission rather than a sink. However, the adoption of sustainable practices, which increase the carbon input rate into soils by at least 5%, will prevent this situation from becoming a sink, especially in wooded areas.

8. The integration of both vegetation and soil carbon reservoirs at the national level is a challenge for stock carbon studies in the terrestrial ecosystems, which remains unsolved in Spain. Future studies must focus on the efforts for national extrapolation of AGB estimation techniques by means of integrating other multitemporal sources with higher-precision spatial resolution. For the SOC, a national effort to harmonize and complete a soil legacy database is necessary. It is essential to increase the available number of profiles to provide a suitable spatial representativeness of the whole territory and fill in the missing information in databases to estimate SOC stock.

PUBLICACIONES RELACIONADAS CON LA TESIS

Publicaciones principales

1. **Durante P**, Algeet N, Oyonarte C. "Organic carbon stock modelling for the quantification of the carbon sinks in terrestrial ecosystems". (2017). In EGU General Assembly Conference Abstracts (Vol. 19, p.18013). http://adsabs.harvard.edu/abs/2017EGUGA..1918013D

2. **Durante, P**., Guevara, M., Vargas, R., Algeet, N., Oyonarte, C., 2020. Uncertainties in estimating the soil carbon sequestration service, in: EGU General Assembly Conference Abstracts. p. 18408.

3. **Durante, P**., Martín-Alcón, S., Gil-Tena, A., Algeet, N., Tomé, J., Recuero, L., Palacios-Orueta, A., Oyonarte, C., 2019. Improving Aboveground Forest Biomass Maps: From High-Resolution to National Scale. Remote Sens. 11, 795. https://doi.org/10.3390/rs11070795

Otras publicaciones y aportaciones a congresos relacionadas

4. Armas D., Guevara M., Alcaraz-Segura D., Rodrigo Vargas R., Soriano-Luna A., **Durante P**. y Oyonarte C. (2017). Mapa digital del perfil del carbono orgánico en los suelos de Andalucía, España. Ecosistemas 26(3), 80-88.

5. Armas D; Guevara, M; Bezares, F; Vargas, R; **Durante P**.; Oyonarte, C (2020.). Digital soil mapping: the challenge to obtain the best soil dataset and create a precise environmental model to support land use management at a national level (Ecuador). 22nd EGU General Assembly.

6. **Durante P**, Alcaraz D, Cabello J, Oyonarte C (2005). Aplicación de la teledetección para la integración de aspectos funcionales en la caracterización de unidades ambientales. Cuadernos de la Sociedad Española de Ciencias Forestales, 19, 91-96.

7. **Durante P**, Oyonarte C, Valladares F. (2009). Influence of land-use tipe and climatic variables on seasonal patterns of NDVI in Mediterranean Iberian ecosystem. Applied Vegetation Science12, 177-185.

8. **Durante P**. Algeet N., Tomé J.L., Oyonarte C. Testing Copernicus Products to estimate forest carbon. Copernicus Global Land User. MetoFrance Toulouse, Francia. 2018. (Comunicación oral).

9. **Durante P**, Algeet N, Tomé JL, Oyonarte C. "Improving forest carbon maps: modeling approaches from high-resolution to national scale". (2018) IEEE Young Professionals Conference on Remote Sensing. Comunicación oral.

10. **Durante P**. "Especial Monte y Agua. Monte y agua: mirando desde el espacio para mejorar la gestión". Revista Montes, nº 141 (2020), pag. 42-45.

11. **Durante P**, Guevara M., Oyonarte C. "Mapa Nacional del secuestro potencial del carbono en suelo (SOCseq)" (2021). Congreso Nacional de Medio Ambiente 2020. Comunicación oral.

12. Oyonarte C, Mingorance M.D., **Durante P**., Piñero G., Barahona e. (2007). Indicators of change in the organic matter in arid zones. Science of the Total Environment, 378, 133-137.

13. Oyonarte C, Aranda V, **Durante P**. (2008).Soil surface properties in Mediterranean mountain ecosystems: environmental factors control and management effect. Forest Ecology and Management, 254, 156-165.

REFERENCIAS

Abd-Elmabod, S.K., Muñoz-Rojas, M., Jordán, A., Anaya-Romero, M., Phillips, J.D., Jones, L., Zhang, Z., Pereira, P., Fleskens, L., van der Ploeg, M., De la Rosa, D., 2020. Climate change impacts on agricultural suitability and yield reduction in a Mediterranean region. Geoderma 374, 114453. https://doi.org/10.1016/j.geoderma.2020.114453

Acácio, V., Dias, F.S., Catry, F.X., Rocha, M., Moreira, F., 2017. Landscape dynamics in Mediterranean oak forests under global change: understanding the role of anthropogenic and environmental drivers across forest types. Glob. Chang. Biol. 23, 1199–1217. https://doi.org/10.1111/gcb.13487

AEMET IPMA, 2011. Atlas climático ibérico/Iberian climate atlas, Agencia Estatal de Meteorología, Ministerio de Medio Ambiente y Rural y Marino, Madrid. Instituto de Meteorologia de Portugal. Agencia Estatal de Meteorología (España) ; Instituto de Meteorología (Portugal).

Albaladejo, J., Martinez-Mena, M., Almagro, M., Ruiz-Navarro, A., Ortiz, R., Albaladejo Montoro, Juan; Martínez-Mena García, María; Almagro Costa, Mercedes; Ruiz-Navarro, Ana; Ortiz Silla, R., 2009. Factores de control en la dinámica del Carbono Orgánico de los suelos de la Región de Murcia, Advances in Studies on Desertification, Murcia (Spain). Murcia.

Alcaraz-Segura, D., Cabello, J., Paruelo, J.M., Delibes, M., 2009. Use of Descriptors of Ecosystem Functioning for Monitoring a National Park Network: A Remote Sensing Approach. Environ. Manage. 43, 38–48. https://doi.org/10.1007/s00267-008-9154-y

Alcaraz-Segura, D., Lomba, A., Sousa-Silva, R., Nieto-Lugilde, D., Alves, P., Georges, D., Vicente, J.R., Honrado, J.P., 2017. Potential of satellite-derived ecosystem functional attributes to anticipate species range shifts. Int. J. Appl. Earth Obs. Geoinf. 57, 86–92. https://doi.org/10.1016/j.jag.2016.12.009

Alias, L., Ortiz, R., 1986. Memorias y mapas de suelos de las hojas del MTN a escala 1:100.000.

Álvaro-Fuentes, J., Easter, M., Paustian, K., 2012a. Climate change effects on organic carbon storage in agricultural soils of northeastern Spain. Agric. Ecosyst. Environ. 155, 87–94. https://doi.org/10.1016/j.agee.2012.04.001

Álvaro-Fuentes, J., Morell, F.J., Plaza-Bonilla, D., Arrúe, J.L., Cantero-Martínez, C., 2012b. Modelling tillage and nitrogen fertilization effects on soil organic carbon dynamics. Soil Tillage Res. 120, 32–39. https://doi.org/10.1016/j.still.2012.01.009

Álvaro-Fuentes, J., López, M. V., Cantero-Martinez, C., Arrúe, J.L., 2008. Tillage Effects on Soil Organic Carbon Fractions in Mediterranean Dryland Agroecosystems. Soil Sci. Soc. Am. J. 72, 541–547. https://doi.org/10.2136/sssaj2007.0164

Amatulli, G., McInerney, D., Sethi, T., Strobl, P., Domisch, S., 2020. Geomorpho90m, empirical evaluation and accuracy assessment of global high-resolution geomorphometric layers. Sci. Data 7, 162. https://doi.org/10.1038/s41597-020-0479-6

Antón, R., Arricibita, F.J., Ruiz-Sagaseta, A., Enrique, A., de Soto, I., Orcaray, L., Zaragüeta, A., Virto, I., 2021. Soil organic carbon monitoring to assess agricultural climate change adaptation practices in Navarre, Spain. Reg. Environ. Chang. 21, 63. https://doi.org/10.1007/s10113-021-01788-w

Arenas-Castro, S., Regos, A., Gonçalves, J.F., Alcaraz-Segura, D., Honrado, J., 2019. Remotely Sensed Variables of Ecosystem Functioning Support Robust Predictions of Abundance Patterns for Rare Species. Remote Sens. 11, 2086. https://doi.org/10.3390/rs11182086

Armas, D.I., Guevara, M., Bezares, F., Vargas, R., Durante, P., Osorio, V.H., Jimenez, W.A., Oyonarte, C., 2021. Harmonized Soil Database of Ecuador 2021 [WWW Document]. EDI Data Portal. https://doi.org/https://doi.org/10.6073/pasta/f228e38f9e7d59b5ac1e2d219f236225

Arrouays, D., Lagacherie, P., Hartemink, A.E., 2017. Digital soil mapping across the globe, Geoderma Regional. Elsevier B.V. https://doi.org/10.1016/j.geodrs.2017.03.002

Arrouays, D., McKenzie, N., Hempel, J., Richer de Forges, A.C., McBratney, A. (Eds.), 2014. Basis of the global spatial soil information system, in: GlobalSoilMap Basis of the Global Spatial Soil Information System. Taylor & Francis Group, London, p. 868.

Asner, G.P., Hughes, R.F., Mascaro, J., Uowolo, A.L., Knapp, D.E., Jacobson, J., Kennedy-Bowdoin, T., Clark, J.K., 2011. High-resolution carbon mapping on the million-hectare Island of Hawaii. Front. Ecol. Environ. 9, 434–439. https://doi.org/10.1890/100179

Asner, G.P., Mascaro, J., 2014. Mapping tropical forest carbon: Calibrating plot estimates to a simple LiDAR metric. Remote Sens. Environ. 140, 614–624. https://doi.org/10.1016/j.rse.2013.09.023

Barahona, E., Santos, F., 1981. Estudios de correlación y regresión de diversos parámetros analíticos de 52 perfiles de suelos del sector Montiel-Alcaraz-Bienservida (Ciudad Real-Albacete). An. Edafol. y Agrobiol. 40, 761–773.

Barredo, J.I., Mauri, A., Caudullo, G., Dosio, A., 2018. Assessing Shifts of Mediterranean and Arid Climates Under RCP4.5 and RCP8.5 Climate Projections in Europe. Pure Appl. Geophys. 175, 3955–3971. https://doi.org/10.1007/s00024-018-1853-6

Bauer-Marschallinger, B., Sabel, D., Wagner, W., 2014. Optimisation of global grids for high-resolution remote sensing data. Comput. Geosci. 72, 84–93. https://doi.org/10.1016/J.CAGEO.2014.07.005

Beaudette, D.E., Roudier, P., O'Geen, A.T., 2013. Algorithms for quantitative pedology: A toolkit for soil scientists. Comput. Geosci. 52, 258–268. https://doi.org/10.1016/J.CAGEO.2012.10.020

Beaudoin, A., Bernier, P.Y., Guindon, L., Villemaire, P., Guo, X.J., Stinson, G., Bergeron, T., Magnussen, S., Hall, R.J., 2014. Mapping attributes of Canada's forests at moderate resolution through k NN and MODIS imagery. Can. J. For. Res. 44, 521–532. https://doi.org/10.1139/cjfr-2013-0401

Bell, S.M., Terrer, C., Barriocanal, C., Jackson, R.B., Rosell-Melé, A., 2021. Soil organic carbon accumulation rates on Mediterranean abandoned agricultural lands. Sci. Total Environ. 759, 143535. https://doi.org/10.1016/j.scitotenv.2020.143535

Bishop, T.F.A., McBratney, A.B., Laslett, G.M., 1999. Modeling soil attribute depth functions with equal-area quadratic smoothing splines. Geoderma 91, 27–45. https://doi.org/10.1016/S0016-7061(99)00003-8

Bispo, A., Andersen, L., Angers, D.A., Bernoux, M., Brossard, M., Cécillon, L., Comans, R.N.J., Harmsen, J., Jonassen, K., Lamé, F., Lhuillery, C., Maly, S., Martin, E., Mcelnea, A.E., Sakai, H., Watabe, Y., Eglin, T.K., 2017. Accounting for Carbon Stocks in Soils and Measuring GHGs Emission Fluxes from Soils: Do We Have the Necessary Standards? Front. Environ. Sci. 5. https://doi.org/10.3389/fenvs.2017.00041

Blackard, J., Finco, M., Helmer, E., Holden, G., Hoppus, M., Jacobs, D., Lister, A., Moisen, G., Nelson, M., Riemann, R., 2008. Mapping U.S. forest biomass using nationwide forest inventory data and moderate resolution information. Remote Sens. Environ. 112, 1658–1677. https://doi.org/10.1016/j.rse.2007.08.021

Blanco, A., 2015. Estudio de la Distribución Espacial y Cartografía Digital de Algunas Propiedades Físicas, Químicas e Hidrodinámicas de Suelos de la Cuenca del Segura.

Bohn, H.L., 1982. Estimate of Organic Carbon in World Soils: II. Soil Sci. Soc. Am. J. 46, 1118–1119. https://doi.org/10.2136/sssaj1982.03615995004600050050x

Bot, A., Benites, J., 2005. The importance of soil organic matter: Key to drought-resistant soil and sustained food production. FAO.

Bottalico, F., Chirici, G., Giannini, R., Mele, S., Mura, M., Puxeddu, M., McRoberts, R.E., Valbuena, R., Travaglini, D., 2017. Modeling Mediterranean forest structure using airborne laser scanning data. Int. J. Appl. Earth Obs. Geoinf. 57, 145–153. https://doi.org/10.1016/j.jag.2016.12.013

Bravo, S., García-Ordiales, E., García-Navarro, F.J., Amorós, J.Á., Pérez-de-los-Reyes, C., Jiménez-Ballesta, R., Esbrí, J.M., García-Noguero, E.M., Higueras, P., 2019. Geochemical distribution of major and trace elements in agricultural soils of Castilla-La Mancha (central Spain): finding criteria for baselines and delimiting regional anomalies. Environ. Sci. Pollut. Res. 26, 3100–3114. https://doi.org/10.1007/s11356-017-0010-6

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324

Brungard, C.W., Boettinger, J.L., Duniway, M.C., Wills, S.A., Edwards, T.C., 2015. Machine learning for predicting soil classes in three semi-arid landscapes. Geoderma 239, 68–83. https://doi.org/10.1016/j.geoderma.2014.09.019

Brus, D., Hengl, T., Heuvelink, G., Kempen, B., Mulder, T.V.L., Olmedo, G.F., Poggio, L., Ribeiro, E., Thine Omuto, C., Yigini, Y., others, 2017. Soil organic carbon mapping: GSOC map cookbook manual. FAO.

Brus, D.J., Kempen, B., Heuvelink, G.B.M., 2011. Sampling for validation of digital soil maps. Eur. J. Soil Sci. 62, 394–407. https://doi.org/10.1111/j.1365-2389.2011.01364.x

Calvo de Anta, R., Luís, E., Febrero-Bande, M., Galiñanes, J., Macías, F., Ortíz, R., Casás, F., 2020. Soil organic carbon in peninsular Spain: Influence of environmental factors and spatial distribution. Geoderma 370. https://doi.org/10.1016/j.geoderma.2020.114365

Cambardella, C.A., Moorman, T.B., Novak, J.M., Parkin, T.B., Karlen, D.L., Turco, R.F., Konopka, A.E., 1994. Field-Scale Variability of Soil Properties in Central Iowa Soils. Soil Sci. Soc. Am. J. 58, 1501–1511. https://doi.org/10.2136/sssaj1994.03615995005800050033x

Campbell, C.A., Paul, E.A., Rennie, D.A., McCallum, K.J., 1967. Applicability of the carbon-dating method of analysis to soil humus studies. Soil Sci. Soc. Am. J. 104, 217–224.

Camps Arbestain, M., Pinto, M., 2004. Los sumideros de carbono en el marco del Protocolo de Kioto. Edafología 11, 27–36.

Canadell, J.G., Raupach, M.R., 2008. Managing Forests for Climate Change Mitigation. Science (80-.). 320, 1456–1457. https://doi.org/10.1126/science.1155458

Carslaw, D.C., Ropkins, K., 2012. Openair - An R package for air quality data analysis. Environ. Model. Softw. 27–28, 52–61. https://doi.org/10.1016/j.envsoft.2011.09.008

Cash, D.W., Moser, S.C.S.C., 2000. Linking global and local scales: designing dynamic assessment and management processes. Glob. Environ. Chang. 10, 109–120. https://doi.org/10.1016/S0959-3780(00)00017-0

Chapin, F.S., Matson, P.A., Vitousek, P.M., 2011. Principles of Terrestrial Ecosystem Ecology. Springer New York, New York, NY. https://doi.org/10.1007/978-1-4419-9504-9

Charro, E., Hernández, S., Martín, J., Moyano, A., Ruiz, N., 2008. Estimación del secuestro de carbono en suelos bajo masas forestales de Pinus halepensis en Castilla y León (España). Cuad. la SECF 130, 125–130.

Chen, J.M., 1996. Evaluation of Vegetation Indices and a Modified Simple Ratio for Boreal Applications. Can. J. Remote Sens. 22, 229–242. https://doi.org/10.1080/07038992.1996.10855178

Chi, H., Sun, G., Huang, J., Guo, Z., Ni, W., Fu, A., 2015. National Forest Aboveground Biomass Mapping from ICESat/GLAS Data and MODIS Imagery in China. Remote Sens. 7, 5534–5564. https://doi.org/10.3390/rs70505534

Chirici, G., Mura, M., McInerney, D., Py, N., Tomppo, E.O., Waser, L.T., Travaglini, D., McRoberts, R.E., 2016. A meta-analysis and review of the literature on the k-Nearest Neighbors technique for forestry applications that use remotely sensed data. Remote Sens. Environ. 176, 282–294. https://doi.org/10.1016/j.rse.2016.02.001

Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., DeFries, R., Galloway, J., Heimann, M., others, 2014. Carbon and other biogeochemical cycles, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, pp. 465–570.

Coleman, K., Jenkinson, D.S., 1996. A Model for the turnover of carbon in soil, in: Evaluation of Soil Organic Matter Models. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 237–246. https://doi.org/10.1007/978-3-642-61094-3_17

Conant, R.T., Ryan, M.G., Ågren, G.I., Birge, H.E., Davidson, E.A., Eliasson, P.E., Evans, S.E., Frey, S.D., Giardina, C.P., Hopkins, F.M., Hyvönen, R., Kirschbaum, M.U.F., Lavallee, J.M., Leifeld, J., Parton, W.J., Megan Steinweg, J., Wallenstein, M.D., Martin Wetterstedt, J.Å., Bradford, M.A., 2011. Temperature and soil organic matter decomposition rates - synthesis of current knowledge and a way forward. Glob. Chang. Biol. 17, 3392–3404. https://doi.org/10.1111/j.1365-2486.2011.02496.x

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. Geosci. Model Dev. 8, 1991–2007. https://doi.org/10.5194/gmd-8-1991-2015

Coulston, J.W., Blinn, C.E., Thomas, V.A., Wynne, R.H., Blinn, C.E., Thomas, V.A., 2016. Approximating Prediction Uncertainty for Random Forest Regression Models 189–197. https://doi.org/10.14358/PERS.82.3.189

Coûteaux, M.-M., Mousseau, M., Célérier, M.-L., Bottner, P., Couteaux, M.-M., Celerier, M.-L., 1991. Increased Atmospheric CO_2 and Litter Quality: Decomposition of Sweet Chestnut Leaf Litter with Animal Food Webs of Different Complexities. Oikos 61, 54. https://doi.org/10.2307/3545406

Crossman, N.D., Bryan, B.A., Summers, D.M., 2011. Carbon Payments and Low-Cost Conservation. Conserv. Biol. 25, 835–845. https://doi.org/10.1111/j.1523-1739.2011.01649.x

de Brogniez, D., Ballabio, C., Stevens, A., Jones, R.J.A.A., Montanarella, L., van Wesemael, B., 2015. A map of the topsoil organic carbon content of Europe generated by a generalized additive model. Eur. J. Soil Sci. 66, 121–134. https://doi.org/10.1111/ejss.12193

De la Rosa, D., Mayol, F., Moreno, D., Rosales, A., Castillo, V., Moreno, F., Cabrera, F., Colomer, J.C., Antoine, J., Masui, S., Brinkman, R., Horn, R., Prange, N., 2001. SEIS.NET: Sistema Español de Información de Suelos en Internet. Edafología 8, 45–56.

Didan, K., 2015. MOD13Q1 MODIS/Terra vegetation indices 16-day L3 global 250m SIN grid V006.

Dixon, R.K., Solomon, A.M., Brown, S., Houghton, R.A., Trexier, M.C., Wisniewski, J., 1994. Carbon Pools and Flux of Global Forest Ecosystems. Science (80-.). 263, 185–190. https://doi.org/10.1126/science.263.5144.185

Do, T.N., Lenca, P., Lallich, S., Pham, N.K., 2010. Classifying very-high-dimensional data with random forests of oblique decision trees, in: Guillet, F., Pinaud, B., Venturini, G., Zighed, D.. (Eds.), Advances in Knowledge Discovery and Management. Springer Science & Business Media, Berlin/Heidelberg, pp. 39–55.

Doblas-Miranda, E., Martínez-Vilalta, J., Lloret, F., Álvarez, A., Ávila, A., Bonet, F.J., Brotons, L., Castro, J., Curiel Yuste, J., Díaz, M., Ferrandis, P., García-Hurtado, E., Iriondo, J.M., Keenan, T.F., Latron, J., Llusià, J., Loepfe, L., Mayol, M., Moré, G., Moya, D., Peñuelas, J., Pons, X., Poyatos, R., Sardans, J., Sus, O., Vallejo, V.R., Vayreda, J., Retana, J., 2015. Reassessing global change research priorities in mediterranean terrestrial ecosystems: how far have we come and where do we go from here? Glob. Ecol. Biogeogr. 24, 25–43. https://doi.org/10.1111/geb.12224

Doblas-Miranda, E., Rovira, P., Brotons, L., Martínez-Vilalta, J., Retana, J., Pla, M., Vayreda, J., 2013. Soil carbon stocks and their variability across the forests, shrublands and grasslands of peninsular Spain. Biogeosciences 10, 8353–8361. https://doi.org/10.5194/bg-10-8353-2013

Dobos, E., Bialkó, T., Micheli, E., Kobza, J., 2010. Legacy Soil Data Harmonization and Database Development, in: Digital Soil Mapping. Springer Netherlands, Dordrecht, pp. 309–323. https://doi.org/10.1007/978-90-481-8863-5_25

Domingo, D., Alonso, R., Lamelas, M.T., Montealegre, A.L., Rodríguez, F., de la Riva, J., 2019. Temporal Transferability of Pine Forest Attributes Modeling Using Low-Density Airborne Laser Scanning Data. Remote Sens. 11, 261. https://doi.org/10.3390/rs11030261

Domingo, D., Lamelas, M.T., Montealegre, A.L., Garcia Martin, A., J, D. la R., García-Martín, A., de la Riva, J., Garcia Martin, A., J, D. la R., 2018. Estimation of Total Biomass in Aleppo Pine Forest Stands Applying Parametric and Nonparametric Methods to Low-Density Airborne Laser Scanning Data. Forests 9, 158. https://doi.org/10.3390/f9040158

Duane, A., Aquilué, N., Gil-Tena, A., Brotons, L., 2016. Integrating fire spread patterns in fire modelling at landscape scale. Environ. Model. Softw. 86, 219–231. https://doi.org/10.1016/j.envsoft.2016.10.001

Durante, P., Guevara, M., Vargas, R., Algeet, N., Oyonarte, C., 2020. Uncertainties in estimating the soil carbon sequestration service, in: EGU General Assembly Conference Abstracts. p. 18408.

Durante, P., Martín-Alcón, S., Gil-Tena, A., Algeet, N., Tomé, J., Recuero, L., Palacios-Orueta, A., Oyonarte, C., 2019. Improving Aboveground Forest Biomass Maps: From High-Resolution to National Scale. Remote Sens. 11, 795. https://doi.org/10.3390/rs11070795

Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E., Yates, C.J., 2011. A statistical explanation of MaxEnt for ecologists. Divers. Distrib. 17, 43–57. https://doi.org/10.1111/j.1472-4642.2010.00725.x

EU NIR, 2021. European Union - 2021 National Inventory Report.

Evans, J.S., Murphy, M.A., 2018. rfUtilities. R package version 2.1-3.

FAO, 2020a. Global Soil Organic Carbon Sequestration Potential Map.

FAO, 2020b. A protocol for measurement, monitoring, reporting and verification of soil organic carbon in agricultural landscapes. FAO. https://doi.org/10.4060/cb0509en

FAO, 2017a. Unlocking the Potential of Soil Organic Carbon, in: Global Symposium on Soil Organic Carbon. p. 36.

FAO, 2017b. Voluntary Guidelines for Sustainable Soil Management. Food and Agriculture Organization of the United Nations, Rome, Italy.

Fassnacht, F.E., Latifi, H., Hartig, F., 2018. Using synthetic data to evaluate the benefits of large field plots for forest biomass estimation with LiDAR. Remote Sens. Environ. 213, 115–128. https://doi.org/10.1016/j.rse.2018.05.007

Fathololoumi, S., Vaezi, A.R., Alavipanah, S.K., Ghorbani, A., Saurette, D., Biswas, A., 2020. Improved digital soil mapping with multitemporal remotely sensed satellite data fusion: A case study in Iran. Sci. Total Environ. 721, 137703. https://doi.org/10.1016/j.scitotenv.2020.137703

Fernández-Landa, A., Fernández-Moya, J., Tomé, J.L., Algeet-Abarquero, N., Guillén-Climent, M.L., Vallejo, R., Sandoval, V., Marchamalo, M., 2018. High resolution forest inventory of pure and mixed stands at regional level combining National Forest Inventory field plots, Landsat, and low density lidar. Int. J. Remote Sens. 39, 4830–4844. https://doi.org/10.1080/01431161.2018.1430406

Ferreira, C.S.S., Seifollahi-Aghmiuni, S., Destouni, G., Ghajarnia, N., Kalantari, Z., 2022. Soil degradation

in the European Mediterranean region: Processes, status and consequences. Sci. Total Environ. 805, 150106. https://doi.org/10.1016/j.scitotenv.2021.150106

Fielding, A.H., Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ. Conserv. 24, 38–49. https://doi.org/10.1017/S0376892997000088

Filippi, P., Cattle, S.R., Pringle, M.J., Bishop, T.F.A., 2021. Space-time monitoring of soil organic carbon content across a semi-arid region of Australia. Geoderma Reg. 24, e00367. https://doi.org/10.1016/j.geodrs.2021.e00367

Florinsky, I. V., 2012. The Dokuchaev hypothesis as a basis for predictive digital soil mapping (on the 125th anniversary of its publication). Eurasian Soil Sci. 45, 445–451. https://doi.org/10.1134/S1064229312040047

Fox, J., 2015. Applied regression analysis and generalized linear models. Sage Publications.

Friedman, J.H., Stuetzle, W., 1981. Projection Pursuit Regression. J. Am. Stat. Assoc. 76, 817. https://doi.org/10.2307/2287576

Frolla, F.D., Angelini, M.E., Peralta, G.E., Di Paolo, L.E., Rodriguez, D.M., Schulz, G., Pascale Medina, C., Beltran, M.J., 2021. Argentina: Soil Organic Carbon Sequestration Potential National Map. National Report. Version 1.0. Year: 2021.

Fujisaki, K., Perrin, A.-S., Desjardins, T., Bernoux, M., Balbino, L.C., Brossard, M., 2015. From forest to cropland and pasture systems: a critical review of soil organic carbon stocks changes in Amazonia. Glob. Chang. Biol. 21, 2773–2786. https://doi.org/10.1111/gcb.12906

Gabarrón-Galeote, M.A., Trigalet, S., Wesemael, B. van, 2015. Soil organic carbon evolution after land abandonment along a precipitation gradient in southern Spain. Agric. Ecosyst. Environ. 199, 114–123. https://doi.org/10.1016/j.agee.2014.08.027

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B., 2013. Bayesian Data Analysis. Chapman and Hall/CRC. https://doi.org/10.1201/b16018

Genuer, R., Poggi, J.-M., Tuleau-Malot, C., 2015. VSURF: An R Package for Variable Selection Using Random Forests. R J. 7, 19. https://doi.org/10.32614/RJ-2015-018

Genuer, R., Poggi, J.-M., Tuleau-Malot, C., Genuer, M.R., 2019. Package 'vsurf.' Pattern Recognit. Lett.

Giardina, C.P., Ryan, M.G., 2000. Evidence that decomposition rates of organic carbon in mineral soil do not vary with temperature. Nature 404, 858–861. https://doi.org/10.1038/35009076

Gibbs, H.K., Brown, S., Niles, J.O., Foley, J.A., 2007. Monitoring and estimating tropical forest carbon stocks: making REDD a reality. Environ. Res. Lett. 2, 045023. https://doi.org/10.1088/1748-9326/2/4/045023

Glenn, N.F., Neuenschwander, A., Vierling, L.A., Spaete, L., Li, A., Shinneman, D.J., Pilliod, D.S., Arkle, R.S., McIlroy, S.K., 2016. Landsat 8 and ICESat-2: Performance and potential synergies for quantifying dryland ecosystem vegetation cover and biomass. Remote Sens. Environ. 185, 233–242. https://doi.org/10.1016/j.rse.2016.02.039

Golon, D.K., 2016. The Land Processes Distributed Active Archive Center (LP DAAC).

González-Jaramillo, V., Fries, A., Zeilinger, J., Homeier, J., Paladines-Benitez, J., Bendix, J., 2018. Estimation of Above Ground Biomass in a Tropical Mountain Forest in Southern Ecuador Using Airborne LiDAR Data. Remote Sens. 10, 660. https://doi.org/10.3390/rs10050660

Górecki, T., Krzyśko, M., Wołyński, W., 2015. Classification problems based on regression models for multi-dimensional functional data. Stat. Transit. new Ser. 16.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

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 Res. 53, 24. https://doi.org/10.1071/SR13306

Greenwell, B.M., Boehmke, B.C., 2020. Variable Importance Plots - An Introduction to the vip Package. R J. 12, 343–366. https://doi.org/10.32614/RJ-2020-013

Gregorutti, B., Michel, B., Saint-Pierre, P., 2017. Correlation and variable importance in random forests. Stat. Comput. 27, 659–678. https://doi.org/10.1007/s11222-016-9646-1

Guevara, M., Olmedo, G.F., Stell, E., Yigini, Y., Aguilar Duarte, Y., Arellano Hernández, C., Arévalo, G.E., Arroyo-Cruz, C.E., Bolivar, A., Bunning, S., Bustamante Cañas, N., Cruz-Gaistardo, C.O., Davila, F., Dell Acqua, M., Encina, A., Figueredo Tacona, H., Fontes, F., Hernández Herrera, J.A., Ibelles Navarro, A.R., Loayza, V., Manueles, A.M., Mendoza Jara, F., Olivera, C., Osorio Hermosilla, R., Pereira, G., Prieto, P., Ramos, I.A., Rey Brina, J.C., Rivera, R., Rodríguez-Rodríguez, J., Roopnarine, R., Rosales Ibarra, A., Rosales Riveiro, K.A., Schulz, G.A., Spence, A., Vasques, G.M., Vargas, R.R., Vargas, R., 2018. No silver bullet for digital soil mapping: country-specific soil organic carbon estimates across Latin America. Soil 4, 173–193. https://doi.org/10.5194/soil-4-173-2018

Haddaway, N.R., Hedlund, K., Jackson, L.E., Kätterer, T., Lugato, E., Thomsen, I.K., Jørgensen, H.B., Söderström, B., 2015. What are the effects of agricultural management on soil organic carbon in boreo-temperate systems? Environ. Evid. 4, 23. https://doi.org/10.1186/s13750-015-0049-0

Hancock, S., Anderson, K., Disney, M., Gaston, K.J., 2017. Measurement of fine-spatial-resolution 3D vegetation structure with airborne waveform lidar: Calibration and validation with voxelised terrestrial lidar. Remote Sens. Environ. 188, 37–50. https://doi.org/10.1016/j.rse.2016.10.041

Harden, J.W., Hugelius, G., Ahlström, A., Blankinship, J.C., Bond-Lamberty, B., Lawrence, C.R., Loisel, J., Malhotra, A., Jackson, R.B., Ogle, S., Phillips, C., Ryals, R., Todd-Brown, K., Vargas, R., Vergara, S.E., Cotrufo, M.F., Keiluweit, M., Heckman, K.A., Crow, S.E., Silver, W.L., DeLonge, M., Nave, L.E., Bond-Lamberty, B., Lawrence, C.R., Loisel, J., Malhotra, A., Jackson, R.B., Ogle, S., Phillips, C., Ryals, R., Todd-Brown, K., Vargas, R., Vergara, S.E., Cotrufo, M.F., Keiluweit, M., Heckman, K.A., Crow, S.E., Silver, W.L., DeLonge, M., Nave, L.E., 2018. Networking our science to characterize the state, vulnerabilities, and management opportunities of soil organic matter. Glob. Chang. Biol. 24, e705–e718. https://doi.org/10.1111/gcb.13896

Hartemink, A.E., 2006. The Future of soil science. International Union of Soil Sciences, Wageningen.

He, H.S., Ventura, S.J., Mladenoff, D.J., 2002. Effects of spatial aggregation approaches on classified satellite imagery. Int. J. Geogr. Inf. Sci. 16, 93–109. https://doi.org/10.1080/13658810110075978

Heiberger, R., Holland, B., Azuaje, F., 2005. Statistical Analysis and Data Display: An Intermediate Course with Examples in S-PLUS, R, and SAS, BioMedical Engineering OnLine. https://doi.org/10.1186/1475-925X-4-18

Heimann, M., Reichstein, M., 2008. Terrestrial ecosystem carbon dynamics and climate feedbacks. Nature 451, 289–292. https://doi.org/10.1038/nature06591

Hengl, T., Heuvelink, G.B.M., Rossiter, D.G., 2007. About regression-kriging: From equations to case studies. Comput. Geosci. 33, 1301–1315. https://doi.org/10.1016/j.cageo.2007.05.001

Hengl, T., Heuvelink, G.B.M., Stein, A., 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. Geoderma 120, 75–93. https://doi.org/10.1016/j.geoderma.2003.08.018

Hengl, T., Heuvelink, G.B.M.M., Kempen, B., Leenaars, J.G.B.B., Walsh, M.G., Shepherd, K.D., Sila, A., MacMillan, R.A., De Jesus, J.M., Tamene, L., Tondoh, J.E., Mendes de Jesus, J., Tamene, L.,

Tondoh, J.E., 2015. Mapping Soil Properties of Africa at 250 m Resolution: Random Forests Significantly Improve Current Predictions. PLoS One 10, e0125814. https://doi.org/10.1371/journal.pone.0125814

Hengl, T., MacMillan, R.A., 2019. Predictive Soil Mapping with R. Open GeoHub Foundation, Wageningen, the Nederlands.

Hengl, T., Mendes de Jesus, J., 2016. Understanding world soils: Machine Learning as a framework for analyzing global soil-landscape relationships. ISRIC - World Soil Information, Wageningen.

Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S., Kempen, B., 2017. SoilGrids250m: Global gridded soil information based on machine learning. PLoS One 12, e0169748. https://doi.org/10.1371/journal.pone.0169748

Hengl, T., Miller, M.A.E., Križan, J., Shepherd, K.D., Sila, A., Kilibarda, M., Antonijević, O., Glušica, L., Dobermann, A., Haefele, S.M., McGrath, S.P., Acquah, G.E., Collinson, J., Parente, L., Sheykhmousa, M., Saito, K., Johnson, J.-M., Chamberlin, J., Silatsa, F.B.T., Yemefack, M., Wendt, J., MacMillan, R.A., Wheeler, I., Crouch, J., 2021. African soil properties and nutrients mapped at 30 m spatial resolution using two-scale ensemble machine learning. Sci. Rep. 11, 6130. https://doi.org/10.1038/s41598-021-85639-y

Heuvelink, G.B.M., 1998. Uncertainty analysis in environmental modelling under a change of spatial scale, in: Soil and Water Quality at Different Scales. Springer Netherlands, Dordrecht, pp. 255–264. https://doi.org/10.1007/978-94-017-3021-1_24

Hiederer, R., Köchy, M., 2012. Global soil organic carbon estimates and the harmonized world soil database, JRC Scientific and Technical Reports. Luxembourg.

Hijmans, R.J., Van Etten, J., Cheng, J., Summer, M., Mattiuzzi, M., Greenberg, J.A., Lamigueiro, O.P., Bevan, A., Bivand, R., Busetto, L., others, 2018. Raster: Geographic Data Analysis and Modeling [R package raster version 2.8--4].

Hoffmann, U., Hoffmann, T., Johnson, E.A.A., Kuhn, N.J., 2014. Assessment of variability and uncertainty of soil organic carbon in a mountainous boreal forest (Canadian Rocky Mountains, Alberta). Catena 113, 107–121. https://doi.org/10.1016/j.catena.2013.09.009

Holben, B.N., 1986. Characteristics of maximum-value composite images from temporal AVHRR data. Int. J. Remote Sens. 7, 1417–1434. https://doi.org/10.1080/01431168608948945

Hollaus, M., Wagner, W., Maier, B., Schadauer, K., 2007. Airborne Laser Scanning of Forest Stem Volume in a Mountainous Environment. Sensors 7, 1559–1577. https://doi.org/10.3390/s7081559

Holmgren, J., Nilsson, M., Olsson, H., 2003. Estimation of Tree Height and Stem Volume on Plots Using Airborne Laser Scanning. For. Sci. 49, 419–428. https://doi.org/10.1093/forestscience/49.3.419

Hounkpatin, O.K.L., Op de Hipt, F., Bossa, A.Y., Welp, G., Amelung, W., 2018. Soil organic carbon stocks and their determining factors in the Dano catchment (Southwest Burkina Faso). Catena 166, 298–309. https://doi.org/10.1016/j.catena.2018.04.013

Hudak, A.T., Strand, E.K., Vierling, L.A., Byrne, J.C., Eitel, J.U.H., Martinuzzi, S., Falkowski, M.J., 2012. Quantifying aboveground forest carbon pools and fluxes from repeat LiDAR surveys. Remote Sens. Environ. 123, 25–40. https://doi.org/10.1016/j.rse.2012.02.023

Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI). Remote Sens. Environ. 25, 295–309. https://doi.org/10.1016/0034-4257(88)90106-X

Hughes, R.F., Asner, G.P., Baldwin, J.A., Mascaro, J., Bufil, L.K.K., Knapp, D.E., 2018. Estimating aboveground carbon density across forest landscapes of Hawaii: Combining FIA plot-derived

estimates and airborne LiDAR. For. Ecol. Manage. 424, 323–337. https://doi.org/10.1016/j.foreco.2018.04.053

Hui Qing Liu, Huete, A., 1995. A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. IEEE Trans. Geosci. Remote Sens. 33, 457–465. https://doi.org/10.1109/36.377946

Hyyppä, J., Hyyppä, H., Leckie, D., Gougeon, F., Yu, X., Maltamo, M., 2008. Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. Int. J. Remote Sens. 29, 1339–1366. https://doi.org/10.1080/01431160701736489

IGME, 1995. Mapa Geológico de la Península Ibérica escala 1:1.000.000 [WWW Document]. URL http://info.igme.es/cartografiadigital/geologica/Geologicos1MMapa.aspx?Id=Geologico1000_(1994)

IGN, 2012. IGN-CORINE Land Cover (España) [WWW Document]. URL https://www.idee.es:80/csw-inspire-idee/static/api/records/spaignCLC2012

IGN, 1995. Mapa de suelos de España: Escala 1:1.000.000 [WWW Document]. URL https://www.ign.es/web/catalogo-cartoteca/resources/html/030769.html

IPCC, 2014. Special report on emissions scenarios, in: International Panel on Climate Change. pp. 1–161.

Jahn, R., Blume, H.P., Asio, V.B., Spaargaren, O., Schad, P., 2006. Guidelines for soil description, 4th edition. FAO, 931, ISRIC - World Soil Information,.

Jalabert, S.S.M., Martin, M.P., Renaud, J.P., Boulonne, L., Jolivet, C., Montanarella, L., Arrouays, D., 2010. Estimating forest soil bulk density using boosted regression modelling. Soil Use Manag. 26, 516–528. https://doi.org/10.1111/j.1475-2743.2010.00305.x

Jebari, A., del Prado, A., Pardo, G., Rodríguez Martín, J.A., Álvaro-Fuentes, J., 2018. Modeling Regional Effects of Climate Change on Soil Organic Carbon in Spain. J. Environ. Qual. 47, 644–653. https://doi.org/10.2134/jeq2017.07.0294

Jenny, H., 1941. Factors of soil formation; a sytem of quantitative pedology.

Jin, Y., Yang, X., Qiu, J., Li, J., Gao, T., Wu, Q., Zhao, F., Ma, H., Yu, H., Xu, B., 2014. Remote Sensing-Based Biomass Estimation and Its Spatio-Temporal Variations in Temperate Grassland, Northern China. Remote Sens. 6, 1496–1513. https://doi.org/10.3390/rs6021496

Jobbágy, E.G., Jackson, R.B., Processes, B., Change, G., 2000. The vertical distribution of soil organic carbon and its relation to climate and vegetation. Ecol. Appl. 10, 423–436. https://doi.org/https://doi.org/10.1890/1051-0761(2000)010[0423:TVDOSO]2.0.CO;2

Jones, R.J.A., Hiederer, R., Rusco, E., Loveland, P.J., Montanarella, L., 2004. The map of organic carbon in topsoils in Europe. Eur. J. Soil Sci. 56, 655–671.

Jordan, C.F., 1969. Derivation of Leaf-Area Index from Quality of Light on the Forest Floor. Ecology 50, 663–666. https://doi.org/10.2307/1936256

Kauranne, T., Joshi, A., Gautam, B., Manandhar, U., Nepal, S., Peuhkurinen, J., Hämäläinen, J., Junttila, V., Gunia, K., Latva-Käyrä, P., Kolesnikov, A., Tegel, K., Leppänen, V., 2017. LiDAR-Assisted Multi-Source Program (LAMP) for Measuring Above Ground Biomass and Forest Carbon. Remote Sens. 9, 154. https://doi.org/10.3390/rs9020154

Kibblewhite, M.G., Jones, R.J.A., Montanarella, L., Baritz, R., Huber, S., Arrouays, D., Micheli, E., Stephens, M., 2008. Environmental Assessment of Soil for Monitoring Volume VI: Soil Monitoring System for Europe, JRC Scientific and Technical Reports. https://doi.org/10.2788/95007

Kirschbaum, M.U.F., 2006. The temperature dependence of organic-matter decomposition—still a topic of debate. Soil Biol. Biochem. 38, 2510–2518. https://doi.org/10.1016/j.soilbio.2006.01.030

Knorr, W., Prentice, I.C., House, J.I., Holland, E.A., 2005. Long-term sensitivity of soil carbon turnover to warming. Nature 433, 298–301. https://doi.org/10.1038/nature03226

Kottek, M., Grieser, J., Beck, C., Rudolf, B., Rubel, F., 2006. World Map of the Köppen-Geiger climate classification updated. Meteorol. Zeitschrift 15, 259–263. https://doi.org/10.1127/0941-2948/2006/0130

Kravchenko, A., Bullock, D.G., 1999. A Comparative Study of Interpolation Methods for Mapping Soil Properties. Agron. J. 91, 393–400. https://doi.org/10.2134/agronj1999.00021962009100030007x

Kulmatiski, A., Vogt, D.J., Siccama, T.G., Tilley, J.P., Kolesinskas, K., Wickwire, T.W., Larson, B.C., 2004. Landscape Determinants of Soil Carbon and Nitrogen Storage in Southern New England. Soil Sci. Soc. Am. J. 68, 2014–2022. https://doi.org/10.2136/sssaj2004.2014

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, 1–11. https://doi.org/10.1016/j.geoderma.2011.06.011

Lafleur, B., Fenton, N.J., Simard, M., Leduc, A., Paré, D., Valeria, O., Bergeron, Y., 2018. Ecosystem management in paludified boreal forests: enhancing wood production, biodiversity, and carbon sequestration at the landscape level. For. Ecosyst. 5, 27. https://doi.org/10.1186/s40663-018-0145-z

Lal, R., 2004. Soil carbon sequestration to mitigate climate change. Geoderma 123, 1–22. https://doi.org/10.1016/j.geoderma.2004.01.032

Lal, R., Smith, P., Jungkunst, H.F., Mitsch, W.J., Lehmann, J., Nair, P.K.R., McBratney, A.B., de Moraes Sá, J.C., Schneider, J., Zinn, Y.L., Skorupa, A.L.A., Zhang, H.-L., Minasny, B., Srinivasrao, C., Ravindranath, N.H., 2018. The carbon sequestration potential of terrestrial ecosystems. J. Soil Water Conserv. 73, 145A-152A. https://doi.org/10.2489/jswc.73.6.145A

Lamichhane, S., Kumar, L., Wilson, B., 2019. Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review. Geoderma 352, 395–413. https://doi.org/10.1016/j.geoderma.2019.05.031

LeDell, E., Poirier, S., 2020. H2O AutoML: Scalable automatic machine learning, in: Proceedings of the AutoML Workshop at ICML.

Li, A., Dhakal, S., Glenn, N., Spaete, L., Shinneman, D., Pilliod, D., Arkle, R., McIlroy, S., 2017. Lidar Aboveground Vegetation Biomass Estimates in Shrublands: Prediction, Uncertainties and Application to Coarser Scales. Remote Sens. 9, 903. https://doi.org/10.3390/rs9090903

Li, L., Guo, Q., Tao, S., Kelly, M., Xu, G., 2015. Lidar with multi-temporal MODIS provide a means to upscale predictions of forest biomass. ISPRS J. Photogramm. Remote Sens. 102, 198–208. https://doi.org/10.1016/j.isprsjprs.2015.02.007

Liang, P., Qin, C.-Z., Zhu, A.-X., Hou, Z.-W., Fan, N.-Q., Wang, Y.-J., Peng, L., Cheng Zhi, Q., Xing, Z.A., Zhi Wei, H., Nai Qing, F., Yi Jie, W., 2020. A case-based method of selecting covariates for digital soil mapping. J. Integr. Agric. 19, 2127–2136. https://doi.org/10.1016/S2095-3119(19)62857-1

Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest. R News 2, 18–22.

Llorente, M., Rovira, P., Merino, A., Rubio, A., Turrión, M., Bad\'\ia, D., Romanya, J., González, J.C.J.A., 2018. The CARBOSOL Database: a georeferenced soil profile analytical database for Spain. https://doi.org/10.1594/PANGAEA.884517

Lombardo, L., Saia, S., Schillaci, C., Mai, P.M., Huser, R., 2018. Modeling soil organic carbon with

Quantile Regression: Dissecting predictors' effects on carbon stocks. Geoderma 318, 148–159. https://doi.org/10.1016/j.geoderma.2017.12.011

López-Garrido, R., Madejón, E., Moreno, F., Murillo, J.M., 2014. Conservation Tillage Influence on Carbon Dynamics Under Mediterranean Conditions. Pedosphere 24, 65–75. https://doi.org/10.1016/S1002-0160(13)60081-8

Lorenz, K., Lal, R., 2010. Carbon Sequestration in Forest Ecosystems. Springer Netherlands, Dordrecht. https://doi.org/10.1007/978-90-481-3266-9

Lu, D., 2006. The potential and challenge of remote sensing-based biomass estimation. Int. J. Remote Sens. 27, 1297–1328. https://doi.org/10.1080/01431160500486732

Magnussen, S., Boudewyn, P., 1998. Derivations of stand heights from airborne laser scanner data with canopy-based quantile estimators. Can. J. For. Res. 28, 1016–1031. https://doi.org/10.1139/x98-078

MAGRAMA, 2012. Cuarto Inventario Forestal Nacional. Región de Murcia. Organismo Autónomo Parques Nacionales, Madrid, Spain.

Maia, S.M.F., Ogle, S.M., Cerri, C.C., Cerri, C.E.P., 2010. Changes in soil organic carbon storage under different agricultural management systems in the Southwest Amazon Region of Brazil. Soil Tillage Res. 106, 177–184. https://doi.org/10.1016/j.still.2009.12.005

Malone, B.P., McBratney, A.B., Minasny, B., Laslett, G.M., 2009. Mapping continuous depth functions of soil carbon storage and available water capacity. Geoderma 154, 138–152. https://doi.org/10.1016/j.geoderma.2009.10.007

Maltamo, M., Packalén, P., Suvanto, A., Korhonen, K.T., Mehtätalo, L., Hyvönen, P., 2009. Combining ALS and NFI training data for forest management planning: a case study in Kuortane, Western Finland. Eur. J. For. Res. 128, 305–317. https://doi.org/10.1007/s10342-009-0266-6

Manning, P., de Vries, F.T., Tallowin, J.R.B., Smith, R., Mortimer, S.R., Pilgrim, E.S., Harrison, K.A., Wright, D.G., Quirk, H., Benson, J., Shipley, B., Cornelissen, J.H.C., Kattge, J., Bönisch, G., Wirth, C., Bardgett, R.D., 2015. Simple measures of climate, soil properties and plant traits predict national-scale grassland soil carbon stocks. J. Appl. Ecol. 52, 1188–1196. https://doi.org/10.1111/1365-2664.12478

MAPA, 2021. Anuario dde Estadística de España 2020. Ministerio de Agricultura, Pesca y Alimentación.

MAPA, 2020. Anuario de Estadística de España 2019, Ministerio de Agricultura, Pesca y Alimentación. Ministerio de Agricultura, Pesca y Alimentación.

Matasci, G., Hermosilla, T., Wulder, M.A., White, J.C., Coops, N.C., Hobart, G.W., Zald, H.S.J., 2018. Large-area mapping of Canadian boreal forest cover, height, biomass and other structural attributes using Landsat composites and lidar plots. Remote Sens. Environ. 209, 90–106. https://doi.org/10.1016/j.rse.2017.12.020

Mauro, F., Valbuena, R., Manzanera, J.A., García-Abril, A., 2011. Influence of Global Navigation Satellite System errors in positioning inventory plots for tree-height distribution studiesThis article is one of a selection of papers from Extending Forest Inventory and Monitoring over Space and Time. Can. J. For. Res. 41, 11–23. https://doi.org/10.1139/X10-164

McBratney, A.. B., Mendonça Santos, M.. L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52. https://doi.org/10.1016/S0016-7061(03)00223-4

McBratney, A., Field, D.J., Koch, A., 2014. The dimensions of soil security. Geoderma 213, 203–213. https://doi.org/10.1016/j.geoderma.2013.08.013

McElreath, R., 2018. Statistical Rethinking. Chapman and Hall/CRC.

https://doi.org/10.1201/9781315372495

Mcgaughey, R.J., Carson, W.W., 2003. Fusing lidar data , photographs , and other data using 2d and 3d visualization techniques. Proc. terrain data Appl. Vis. Connect. 28–30.

McNairn, H., Protz, R., 1993. Mapping Corn Residue Cover on Agricultural Fields in Oxford County, Ontario, Using Thematic Mapper. Can. J. Remote Sens. 19, 152–159. https://doi.org/10.1080/07038992.1993.10874543

Meinshausen, N., 2017. Package 'quantregForest' Quantile Regression Forests.(R package version 1.3-7).

Meinshausen, N., 2006. Quantile Regression Forests. J. Mach. Learn. Res. 7, 983–999.

Mendoza-Ponce, A., Corona-Núñez, R., Kraxner, F., Leduc, S., Patrizio, P., 2018. Identifying effects of land use cover changes and climate change on terrestrial ecosystems and carbon stocks in Mexico. Glob. Environ. Chang. 53, 12–23. https://doi.org/10.1016/j.gloenvcha.2018.08.004

Minasny, B., Malone, B.P., McBratney, A.B., Angers, D.A., Arrouays, D., Chambers, A., Chaplot, V., Chen, Z.-S., Cheng, K., Das, B.S., Field, D.J., Gimona, A., Hedley, C.B., Hong, S.Y., Mandal, B., Marchant, B.P., Martin, M., McConkey, B.G., Mulder, V.L., O'Rourke, S., Richer-de-Forges, A.C., Odeh, I., Padarian, J., Paustian, K., Pan, G., Poggio, L., Savin, I., Stolbovoy, V., Stockmann, U., Sulaeman, Y., Tsui, C.-C., Vågen, T.-G., van Wesemael, B., Winowiecki, L., 2017. Soil carbon 4 per mille. Geoderma 292, 59–86. https://doi.org/10.1016/j.geoderma.2017.01.002

Miquelajauregui, Y., 2013. Modelos de simulación de la dinámica del carbono, in: Aplicaciones de Modelos Ecológicos a La Gestión de Recursos Naturales. pp. 15–38. https://doi.org/10.3926/oms.173

MITECO, 2019. Perfil Ambiental de España 2.018. Ministerio para la Transición Ecológica. Secretaría General Técnica.Centro de Publicaciones 2019.

Molina, P., Asner, G., Farjas Abadía, M., Ojeda Manrique, J., Sánchez Diez, L., Valencia, R., 2015. Spatially-Explicit Testing of a General Aboveground Carbon Density Estimation Model in a Western Amazonian Forest Using Airborne LiDAR. Remote Sens. 8, 9. https://doi.org/10.3390/rs8010009

Moni, C., Rumpel, C., Virto, I., Chabbi, A., Chenu, C., 2010. Relative importance of sorption versus aggregation for organic matter storage in subsoil horizons of two contrasting soils. Eur. J. Soil Sci. 61, 958–969. https://doi.org/10.1111/j.1365-2389.2010.01307.x

Montanarella, L., 2015. Agricultural policy: Govern our soils. Nature 528, 32–33. https://doi.org/10.1038/528032a

Montero, G., Pasalodos-Tato, M., López-Senespleda, E., Onrubia, R., Madrigal, G., 2013. Ecuaciones para la estimación de la biomasa en matorrales y arbustedos mediterráneos, in: 6º Congreso Forestal Español; Sociedad Española de Ciencias Forestales.

Montero, G., Ruiz-Peinado, R., Munoz, M., 2005. Producción de biomasa y fijación de CO_2 por los bosques españoles. INIA-Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria~…, Madrid.

Moore, I.D., Grayson, R.B., Ladson, A.R., 1991. Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. Hydrol. Process. 5, 3–30. https://doi.org/10.1002/hyp.3360050103

Mulder, V.L., Lacoste, M., Richer-de-Forges, A.C., Martin, M.P., Arrouays, D., 2016. National versus global modelling the 3D distribution of soil organic carbon in mainland France. Geoderma 263, 16–34. https://doi.org/10.1016/j.geoderma.2015.08.035

Muñoz-Rojas, M., Jordán, A., Zavala, L.M., De la Rosa, D., Abd-Elmabod, S.K., Anaya-Romero, M., 2012. Organic carbon stocks in Mediterranean soil types under different land uses (Southern Spain). Solid Earth 3, 375–386. https://doi.org/10.5194/se-3-375-2012

Muñoz-Rojas, M., Jordán, A., Zavala, L.M., De la Rosa, D., Abd-Elmabod, S.K., Anaya-Romero, M., 2015. Impact of Land Use and Land Cover Changes on Organic Carbon Stocks in Mediterranean Soils (1956–2007). L. Degrad. Dev. 26, 168–179. https://doi.org/10.1002/ldr.2194

Mura, M., Bottalico, F., Giannetti, F., Bertani, R., Giannini, R., Mancini, M., Orlandini, S., Travaglini, D., Chirici, G., 2018. Exploiting the capabilities of the Sentinel-2 multi spectral instrument for predicting growing stock volume in forest ecosystems. Int. J. Appl. Earth Obs. Geoinf. 66, 126–134. https://doi.org/10.1016/j.jag.2017.11.013

Murgaš, V., Sačkov, I., Sedliak, M., Tunák, D., Chudý, F., 2018. Assessing horizontal accuracy of inventory plots in forests with different mix of tree species composition and development stage. J. For. Sci. 64, 478–485. https://doi.org/10.17221/92/2018-JFS

Mutanga, O., Skidmore, A.K., 2004. Narrow band vegetation indices overcome the saturation problem in biomass estimation. Int. J. Remote Sens. 25, 3999–4014. https://doi.org/10.1080/01431160310001654923

Næsset, E., 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. Remote Sens. Environ. 80, 88–99. https://doi.org/10.1016/S0034-4257(01)00290-5

Næsset, E., Gobakken, T., 2008. Estimation of above- and below-ground biomass across regions of the boreal forest zone using airborne laser. Remote Sens. Environ. 112, 3079–3090. https://doi.org/10.1016/j.rse.2008.03.004

Navarrete-Poyatos, M.A., Navarro-Cerrillo, R.M., Lara-Gómez, M.A., Duque-Lazo, J., Varo, M. de los A., Palacios Rodriguez, G., 2019. Assessment of the Carbon Stock in Pine Plantations in Southern Spain through ALS Data and K-Nearest Neighbor Algorithm Based Models. Geosciences 9, 442. https://doi.org/10.3390/geosciences9100442

Navarro-Cerrillo, R., Duque-Lazo, J., Rodríguez-Vallejo, C., Varo-Martínez, M., Palacios-Rodríguez, G., 2018. Airborne Laser Scanning Cartography of On-Site Carbon Stocks as a Basis for the Silviculture of Pinus Halepensis Plantations. Remote Sens. 10, 1660. https://doi.org/10.3390/rs10101660

Navarro, J.A., Algeet, N., Fernández-Landa, A., Esteban, J., Rodríguez-Noriega, P., Guillén-Climent, M.L., 2019. Integration of UAV, Sentinel-1, and Sentinel-2 Data for Mangrove Plantation Aboveground Biomass Monitoring in Senegal. Remote Sens. 11, 77. https://doi.org/10.3390/rs11010077

Nelson, R., Gobakken, T., Næsset, E., Gregoire, T.G., Ståhl, G., Holm, S., Flewelling, J., 2012. Lidar sampling — Using an airborne profiler to estimate forest biomass in Hedmark County, Norway. Remote Sens. Environ. 123, 563–578. https://doi.org/10.1016/j.rse.2011.10.036

Nguyen, K.A., Chen, W., Lin, B.-S., Seeboonruang, U., 2021. Comparison of Ensemble Machine Learning Methods for Soil Erosion Pin Measurements. ISPRS Int. J. Geo-Information 10, 42. https://doi.org/10.3390/ijgi10010042

Nguyen, T., Jones, S., Soto-Berelov, M., Haywood, A., Hislop, S., 2018. A Comparison of Imputation Approaches for Estimating Forest Biomass Using Landsat Time-Series and Inventory Data. Remote Sens. 10, 1825. https://doi.org/10.3390/rs10111825

Nieto, O.M., Castro, J., Fernández, E., Smith, P., 2010. Simulation of soil organic carbon stocks in a Mediterranean olive grove under different soil-management systems using the RothC model. Soil Use Manag. 26, 118–125. https://doi.org/10.1111/j.1475-2743.2010.00265.x

Ninyerola, M., Pons, X., Roure, J., 2005. Atlas climático digital de la Península Ibérica: metodología y

aplicaciones en bioclimatología y geobotánica. Barcelona.

Ogle, S.M., Breidt, F.J., Easter, M., Williams, S., Killian, K., Paustian, K., 2010. Scale and uncertainty in modeled soil organic carbon stock changes for US croplands using a process-based model. Glob. Chang. Biol. 16, 810–822. https://doi.org/10.1111/j.1365-2486.2009.01951.x

Ometto, J.P., Aguiar, A.P., Assis, T., Soler, L., Valle, P., Tejada, G., Lapola, D.M., Meir, P., 2014. Amazon forest biomass density maps: tackling the uncertainty in carbon emission estimates. Clim. Change 124, 545–560. https://doi.org/10.1007/s10584-014-1058-7

Omran, E.-S.E., 2012. Improving the Prediction Accuracy of Soil Mapping through Geostatistics. Int. J. Geosci. 03, 574–590. https://doi.org/10.4236/ijg.2012.33058

Ozdemir, I., 2014. Linear transformation to minimize the effects of variability in understory to estimate percent tree canopy cover using RapidEye data. GIScience Remote Sens. 51, 288–300. https://doi.org/10.1080/15481603.2014.912876

Pachauri, R.K., Allen, M.R., Barros, V.R., Broome, J., Cramer, W., Christ, R., Church, J.A., Clarke, L., Dahe, Q., Dasgupta, P., Dubash, N.K., Edenhofer, O., Elgizouli, I., Field, C.B., Forster, P., Friedlingstein, P., Fuglestvedt, J., Gomez-Echeverri, L., Hallegatte, S., Hegerl, G., Howden, M., Jiang, K., Cisneroz, B.J., Kattsov, V., Lee, H., Mach, K.J., Marotzke, J., Mastrandrea, M.D., Meyer, L., Minx, J., Mulugetta, Y., O'Brien, K., Oppenheimer, M., Pereira, J.J., Pichs-Madruga, R., Plattner, G.-K., Pörtner, H.-O., Power, S.B., Preston, B., Ravindranath, N.H., Reisinger, A., Riahi, K., Rusticucci, M., Scholes, R., Seyboth, K., Sokona, Y., Stavins, R., Stocker, T.F., Tschakert, P., van Vuuren, D., van Ypserle, J.-P., 2014. Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. IPCC, Geneva, Switzerland.

Papadakis, J., 1966. Climates of the world and their agricultural pontentialities.

Pardo, G., del Prado, A., Martínez-Mena, M., Bustamante, M.A., Martín, J.A.R., Álvaro-Fuentes, J., Moral, R., 2017. Orchard and horticulture systems in Spanish Mediterranean coastal areas: Is there a real possibility to contribute to C sequestration? Agric. Ecosyst. Environ. 238, 153–167. https://doi.org/10.1016/j.agee.2016.09.034

Parker, G.G., Russ, M.E., 2004. The canopy surface and stand development: assessing forest canopy structure and complexity with near-surface altimetry. For. Ecol. Manage. 189, 307–315. https://doi.org/10.1016/j.foreco.2003.09.001

Paruelo, J.M., Epstein, H.E., Lauenroth, W.K., Burke, I.C., 1997. ANPP Estimates from NDVI for the Central Grassland Region of the United States. Ecology 78, 953. https://doi.org/10.2307/2266073

Pásztor, L., Laborczi, A., Szatmári, G., Koós, S., Bakacsi, Z., Makó, A., Tóth, B., 2019. Digital soil maps for the support of national mapping and assessment of ecosystem services. Geophys. Res. Abstr. 21, 5645.

Pecl, G.T., Araújo, M.B., Bell, J.D., Blanchard, J., Bonebrake, T.C., Chen, I.-C., Clark, T.D., Colwell, R.K., Danielsen, F., Evengård, B., Falconi, L., Ferrier, S., Frusher, S., Garcia, R.A., Griffis, R.B., Hobday, A.J., Janion-Scheepers, C., Jarzyna, M.A., Jennings, S., Lenoir, J., Linnetved, H.I., Martin, V.Y., McCormack, P.C., McDonald, J., Mitchell, N.J., Mustonen, T., Pandolfi, J.M., Pettorelli, N., Popova, E., Robinson, S.A., Scheffers, B.R., Shaw, J.D., Sorte, C.J.B., Strugnell, J.M., Sunday, J.M., Tuanmu, M.-N., Vergés, A., Villanueva, C., Wernberg, T., Wapstra, E., Williams, S.E., 2017. Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. Science (80-.). 355, eaai9214. https://doi.org/10.1126/science.aai9214

Penman, J., Gytarsky, M., Hiraishi, T., Krug, T., Kruger, D., Pipatti, R., Buendia, L., Miwa, K., Ngara, T., Tanabe, K., Wagner, F., 2003. Good Practice Guidance for Land Use, Land-Use Change and Forestry, IPCC National Greenhouse Gas Inventories Programme. Institute for Global Environmental Strategies (IGES). https://doi.org/10.1016/j.crvi.2014.11.004

Perpiña Castillo, C., Coll Aliaga, E., Lavalle, C., Martínez Llario, J.C., 2020. An Assessment and Spatial

Modelling of Agricultural Land Abandonment in Spain (2015–2030). Sustainability 12, 560. https://doi.org/10.3390/su12020560

Pettorelli, N., Vik, J.O., Mysterud, A., Gaillard, J.-M., Tucker, C.J., Stenseth, N.C., 2005. Using the satellite-derived NDVI to assess ecological responses to environmental change. Trends Ecol. Evol. 20, 503–510. https://doi.org/10.1016/j.tree.2005.05.011

Pfeiffer, M., Padarian, J.J., Osorio, R., Bustamante, N., Olmedo, G.F., Guevara, M., Aburto, F., Albornoz, F., Antilén, M., Araya, E.E., Arellano, E., Barret, M., Barrera, J., Boeckx, P., Briceño, M., Bunning, S., Cabrol, L., Casanova, M., Cornejo, P., Corradini, F., Curaqueo, G., Doetterl, S., Duran, P., Escudey, M., Espinoza, A., Francke, S., Fuentes, J.P., Fuentes, M., Gajardo, G., García, R., Gallaud, A., Galleguillos, M., Gomez, A.A., Hidalgo, M., Ivelic-Sáez, J., Mashalaba, L., Matus, F., Meza, F., Mora, M. de la L., Mora, J., Muñoz, C., Norambuena, P., Olivera, C., Ovalle, C., Panichini, M., Pauchard, A., Pérez-Quezada, J.F., Radic, S., Ramirez, J.J., Riveras, N.N., Ruiz, G.G., Salazar, O., Salgado, I.I., Seguel, O., Sepúlveda, M., Sierra, C., Tapia, Y., Tapia, F., Toledo, B., Torrico, J.M.J.M., Valle, S., Vargas, R., Wolff, M., Zagal, E., Federico Olmedo, G., Guevara, M., Aburto, F., Albornoz, F., Antilén, M., Araya, E.E., Arellano, E., Barret, M., Barrera, J., Boeckx, P., Briceño, M., Bunning, S., Cabrol, L., Casanova, M., Cornejo, P., Corradini, F., Curaqueo, G., Doetterl, S., Duran, P., Escudey, M., Espinoza, A., Francke, S., Pablo Fuentes, J., Fuentes, M., Gajardo, G., García, R., Gallaud, A., Galleguillos, M., Gomez, A.A., Hidalgo, M., Ivelic-Sáez, J., Mashalaba, L., Matus, F., Meza, F., De La Luz Mora, M., Mora, J., Muñoz, C., Norambuena, P., Olivera, C., Ovalle, C., Panichini, M., Munoz, C., Pérez-Quezada, J.F., Radic, S., Ramirez, J.J., Riveras, N.N., Ruiz, G.G., Salazar, O., Salgado, I.I., Seguel, O., Sepúlveda, M., Sierra, C., Tapia, Y., Tapia, F., Toledo, B., Torrico, J.M.J.M., Valle, S., Vargas, R., Wolff, M., Zagal, E., 2020. CHLSOC: the Chilean Soil Organic Carbon database, a multi-institutional collaborative effort. Earth Syst. Sci. Data 12, 457–468. https://doi.org/10.5194/essd-12-457-2020

Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. Ecol. Modell. 190, 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Piao, S., Friedlingstein, P., Ciais, P., de Noblet-Ducoudre, N., Labat, D., Zaehle, S., 2007. Changes in climate and land use have a larger direct impact than rising CO_2 on global river runoff trends. Proc. Natl. Acad. Sci. 104, 15242–15247. https://doi.org/10.1073/pnas.0707213104

Pike, R.J., Wilson, S.E., 1971. Elevation-Relief Ratio, Hypsometric Integral, and Geomorphic Area-Altitude Analysis. GSA Bull. 82, 1079–1084. https://doi.org/10.1130/0016-7606(1971)82[1079:ERHIAG]2.0.CO;2

Poeplau, C., Vos, C., Don, A., 2017. Soil organic carbon stocks are systematically overestimated by misuse of the parameters bulk density and rock fragment content. SOIL 3, 61–66. https://doi.org/10.5194/soil-3-61-2017

Post, W.M., Mann, L.K., 2005. Changes in Soil Organic Carbon and Nitrogen as a Result of Cultivation. Biol. Environ. Res. https://doi.org/10.3334/CDIAC/TCM.006

Powlson, D., 2005. Will soil amplify climate change? Nature 433, 204–205. https://doi.org/10.1038/433204a

Qi, J., Chehbouni, A., Huete, A.R., Kerr, Y.H., Sorooshian, S., 1994. A modified soil adjusted vegetation index. Remote Sens. Environ. 48, 119–126. https://doi.org/10.1016/0034-4257(94)90134-1

R developement Core Team, 2013. R: A language and environment for statistical computing.

Reed, B.C., Schwartz, M.D., Xiao, X., 2009. Remote Sensing Phenology, in: Phenology of Ecosystem Processes. Springer New York, New York, NY, pp. 231–246. https://doi.org/10.1007/978-1-4419-0026-5_10

Reynoso, V., Frolla, F., Ortiz, S., Angelini, M., Cerón, A., Beltrán, M., Bunge, V., Peralta, G., Di Paolo, L., Rodriguez, D., Schulz, G., Juan, V.L., Medina, C., Guevara, M., 2021. Mapping soil carbon sequestration across Argentina and Mexico using Roth C, in: WGs Digital Soil Mapping - Global Soil Map. https://doi.org/10.13140/RG.2.2.31488.97286

Robinson, T.P., Metternicht, G., 2006. Testing the performance of spatial interpolation techniques for mapping soil properties. Comput. Electron. Agric. 50, 97–108. https://doi.org/10.1016/j.compag.2005.07.003

Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., Chica-Rivas, M., 2015. Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. Ore Geol. Rev. 71, 804–818. https://doi.org/10.1016/j.oregeorev.2015.01.001

Rodríguez, F., Fernández, A., Tomé, J., 2014. Resultados y reflexiones tras cinco años de inventario forestal con tecnología LiDAR. Foresta 28–33.

Rodríguez Martín, J.A., Álvaro-Fuentes, J., Gonzalo, J., Gil, C., Ramos-Miras, J.J., Grau Corbí, J.M., Boluda, R., 2016. Assessment of the soil organic carbon stock in Spain. Geoderma 264, 117–125. https://doi.org/10.1016/j.geoderma.2015.10.010

Rosell, R.A., Gasparoni, J.C., Galantini, J.A., 2001. Soil organic matter evaluation, in: Assessment Methods for Soil Carbon. Lewis Publishers Boca Raton, pp. 311–322.

Rossel, R.A.V., Behrens, T., 2010. Using data mining to model and interpret soil diffuse reflectance spectra. Geoderma 158, 46–54. https://doi.org/10.1016/j.geoderma.2009.12.025

Roujean, J.-L., Breon, F.-M., 1995. Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. Remote Sens. Environ. 51, 375–384. https://doi.org/10.1016/0034-4257(94)00114-3

Ruiz-Sinoga, J.D., Martínez-Murillo, J.F., Gabarrón-Galeote, M.A., García-Marín, R., 2011. The effects of soil moisture variability on the vegetation pattern in Mediterranean abandoned fields (Southern Spain). Catena 85, 1–11. https://doi.org/10.1016/j.catena.2010.11.004

Rumpel, C., Kögel-Knabner, I., 2011. Deep soil organic matter—a key but poorly understood component of terrestrial C cycle. Plant Soil 338, 143–158. https://doi.org/10.1007/s11104-010-0391-5

Saarela, S., Holm, S., Grafström, A., Schnell, S., Næsset, E., Gregoire, T.G., Nelson, R.F., Ståhl, G., 2016. Hierarchical model-based inference for forest inventory utilizing three sources of information. Ann. For. Sci. 73, 895–910. https://doi.org/10.1007/s13595-016-0590-1

Saatchi, S.S., Houghton, R.A., Dos Santos Alvalá, R.C., Soares, J. V., Yu, Y., 2007. Distribution of aboveground live biomass in the Amazon basin. Glob. Chang. Biol. 13, 816–837. https://doi.org/10.1111/j.1365-2486.2007.01323.x

Sanchez, P.A., Ahamed, S., Carre, F., Hartemink, A.E., Hempel, J., Huising, J., Lagacherie, P., McBratney, A.B., McKenzie, N.J., Mendonca-Santos, M. d. L., Minasny, B., Montanarella, L., Okoth, P., Palm, C.A., Sachs, J.D., Shepherd, K.D., Vagen, T.-G., Vanlauwe, B., Walsh, M.G., Winowiecki, L.A., Zhang, G.-L., 2009. Digital Soil Map of the World. Science (80-.). 325, 680–681. https://doi.org/10.1126/science.1175084

Savin, I.Y., Zhogolev, A. V., Prudnikova, E.Y., 2019. Modern Trends and Problems of Soil Mapping. Eurasian Soil Sci. 52, 471–480. https://doi.org/10.1134/S1064229319050107

Schillaci, C., Acutis, M., Lombardo, L., Lipani, A., Fantappiè, M., Märker, M., Saia, S., 2017a. 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. Sci. Total Environ. 601–602, 821–832. https://doi.org/10.1016/j.scitotenv.2017.05.239

Schillaci, C., Lombardo, L., Saia, S., Fantappiè, M., Märker, M., Acutis, M., 2017b. Modelling the topsoil carbon stock of agricultural lands with the Stochastic Gradient Treeboost in a semi-arid Mediterranean region. Geoderma 286, 35–45. https://doi.org/10.1016/j.geoderma.2016.10.019

Schlesinger, W.H., Bernhardt, E.S., 2013. Biogeochemistry: An Analysis of Global Change (3rd edition). Elsevier. https://doi.org/10.1016/C2010-0-66291-2

Searle, R., McBratney, A., Grundy, M., Kidd, D., Malone, B., Arrouays, D., Stockman, U., Zund, P., Wilson, P., Wilford, J., Van Gool, D., Triantafilis, J., Thomas, M., Stower, L., Slater, B., Robinson, N., Ringrose-Voase, A., Padarian, J., Payne, J., Orton, T., Odgers, N., O'Brien, L., Minasny, B., Bennett, J.M., Liddicoat, C., Jones, E., Holmes, K., Harms, B., Gray, J., Bui, E., Andrews, K., 2021. Digital soil mapping and assessment for Australia and beyond: A propitious future. Geoderma Reg. 24, e00359. https://doi.org/10.1016/j.geodrs.2021.e00359

Segura, C., Jiménez, M.N., Nieto, O., Navarro, F.B., Fernández-Ondoño, E., 2016. Changes in soil organic carbon over 20 years after afforestation in semiarid SE Spain. For. Ecol. Manage. 381, 268–278. https://doi.org/10.1016/j.foreco.2016.09.035

Serrano, F.S., 2000. Deformaciones Recientes en el Centro de la Península Ibérica.

SGE-IGME, 2004. Mapa Geológico de España a escala 1:2.000.000 [WWW Document]. Geol. España. URL https://info.igme.es/cartografiadigital/geologica/Geologicos1MMapa.aspx?Id=Geologico2000_(2004)

Shangguan, W., Dai, Y., Duan, Q., Liu, B., Yuan, H., 2014. A global soil data set for earth system modeling. J. Adv. Model. Earth Syst. 6, 249–263. https://doi.org/10.1002/2013MS000293

Shangguan, W., Hengl, T., Mendes de Jesus, J., Yuan, H., Dai, Y., 2017. Mapping the global depth to bedrock for land surface modeling. J. Adv. Model. Earth Syst. 9, 65–88. https://doi.org/10.1002/2016MS000686

Sierra, C.A., Müller, M., Trumbore, S.E., 2012. Models of soil organic matter decomposition: the SoilR package, version 1.0. Geosci. Model Dev. 5, 1045–1060. https://doi.org/10.5194/gmd-5-1045-2012

Silatsa, F.B.T., Yemefack, M., Tabi, F.O., Heuvelink, G.B.M., Leenaars, J.G.B., 2020. Assessing countrywide soil organic carbon stock using hybrid machine learning modelling and legacy soil data in Cameroon. Geoderma 367. https://doi.org/10.1016/j.geoderma.2020.114260

Smith, J., Smith, P., Meyer, J., Wattenbach, M., Zaehle, S., Lindner, M., Jones, R.J.A., Hiederer, R., Rounsevell, M., Montanarella, L., Reginster, I., Kankaanpää, S., 2006. Projected changes in mineral soil carbon of European forests, 1990–2100. Can. J. Soil Sci. 86, 159–169. https://doi.org/10.4141/S05-078

Smith, J., Smith, P., Wattenbach, M., Zaehle, S., Hiederer, R., Jones, R.J.A., Montanarella, L., Rounsevell, M.D.A., Reginster, I., Ewert, F., 2005. Projected changes in mineral soil carbon of European croplands and grasslands, 1990-2080. Glob. Chang. Biol. 11, 2141–2152. https://doi.org/10.1111/j.1365-2486.2005.001075.x

Smith, P., 2004. How long before a change in soil organic carbon can be detected? Glob. Chang. Biol. 10, 1878–1883. https://doi.org/10.1111/j.1365-2486.2004.00854.x

Smith, P., Smith, J.U., Franko, U., Kuka, K., Romanenkov, V.A., Shevtsova, L.K., Wattenbach, M., Gottschalk, P., Sirotenko, O.D., Rukhovich, D.I., Koroleva, P. V., Romanenko, I.A., Lisovoi, N. V., 2007. Changes in mineral soil organic carbon stocks in the croplands of European Russia and the Ukraine, 1990–2070; comparison of three models and implications for climate mitigation. Reg. Environ. Chang. 7, 105–119. https://doi.org/10.1007/s10113-007-0028-2

Smith, P., Soussana, J.F., Angers, D., Schipper, L., Chenu, C., Rasse, D.P., Batjes, N.H., Egmond, F., McNeill, S., Kuhnert, M., Arias-Navarro, C., Olesen, J.E., Chirinda, N., Fornara, D., Wollenberg, E., Álvaro-Fuentes, J., Sanz-Cobena, A., Klumpp, K., van Egmond, F., McNeill, S., Kuhnert, M., Arias-Navarro, C., Olesen, J.E., Chirinda, N., Fornara, D., Wollenberg, E., Álvaro-Fuentes, J., Sanz-Cobena, A., Klumpp, K., 2020. How to measure, report and verify soil carbon change to realize the potential of soil carbon sequestration for atmospheric greenhouse gas removal. Glob.

Chang. Biol. 26, 219–241. https://doi.org/10.1111/gcb.14815

Stockmann, U., Adams, M.A., Crawford, J.W., Field, D.J., Henakaarchchi, N., Jenkins, M., Minasny, B., McBratney, A.B., Courcelles, V. de R. de, Singh, K., Wheeler, I., Abbott, L., Angers, D.A., Baldock, J., Bird, M., Brookes, P.C., Chenu, C., Jastrow, J.D., Lal, R., Lehmann, J., O'Donnell, A.G., Parton, W.J., Whitehead, D., Zimmermann, M., 2013. The knowns, known unknowns and unknowns of sequestration of soil organic carbon. Agric. Ecosyst. Environ. 164, 80–99. https://doi.org/10.1016/J.AGEE.2012.10.001

Stockmann, U., Padarian, J., McBratney, A., Minasny, B., de Brogniez, D., Montanarella, L., Hong, S.Y., Rawlins, B.G., Field, D.J., 2015. Global soil organic carbon assessment. Glob. Food Sec. 6, 9–16. https://doi.org/10.1016/j.gfs.2015.07.001

Sun, W., Liu, X., 2020. Review on carbon storage estimation of forest ecosystem and applications in China. For. Ecosyst. 7, 4. https://doi.org/10.1186/s40663-019-0210-2

Taghizadeh-Mehrjardi, R., Minasny, B., Toomanian, N., Zeraatpisheh, M., Amirian-Chakan, A., Triantafilis, J., 2019. Digital Mapping of Soil Classes Using Ensemble of Models in Isfahan Region, Iran. Soil Syst. 3, 37. https://doi.org/10.3390/soilsystems3020037

Thamo, T., Pannell, D.J., 2016. Challenges in developing effective policy for soil carbon sequestration: perspectives on additionality, leakage, and permanence. Clim. Policy 16, 973–992. https://doi.org/10.1080/14693062.2015.1075372

Thenkabail, P.S., Smith, R.B., De Pauw, E., 2000. Hyperspectral Vegetation Indices and Their Relationships with Agricultural Crop Characteristics. Remote Sens. Environ. 71, 158–182. https://doi.org/10.1016/S0034-4257(99)00067-X

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, 57–70. https://doi.org/10.1016/j.geoderma.2006.03.037

Tian, H., Lu, C., Yang, J., Banger, K., Huntzinger, D.N., Schwalm, C.R., Michalak, A.M., Cook, R., Ciais, P., Hayes, D., Huang, M., Ito, A., Jain, A.K., Lei, H., Mao, J., Pan, S., Post, W.M., Peng, S., Poulter, B., Ren, W., Ricciuto, D., Schaefer, K., Shi, X., Tao, B., Wang, W., Wei, Y., Yang, Q., Zhang, B., Zeng, N., 2015. Global patterns and controls of soil organic carbon dynamics as simulated by multiple terrestrial biosphere models: Current status and future directions. Global Biogeochem. Cycles 29, 775–792. https://doi.org/10.1002/2014GB005021

Tobler, W.R., 1988. Resolution, resampling, and all that, in: Mounsey, H., Tomlinson, R.F. (Eds.), Building Databases for Global Science: The Proceedings of the First Meeting of the International Geographical Union Global Database Planning Project. Taylor and Francis, Hampshire, U.K., pp. 129–137.

Trnka, M., Olensen, J.E., Kersebaum, K.C., Skjelvag, A.O., Eitzinger, J., SEGUIN, B., Peltonen-Sainio, P., Rötter, R., Iglesias, A., Orlandini, S., Dubrovský, M., Hlavinka, P., Balek, J., Eckersten, H., Cloppet, E., Calanca, P., Gobin, A., Vucetc, V., Nejedlik, P., Kumar, S., Lalic, B., Mestre, A., Rossi, F., Kozyra, J., Alexandrov, V., Semerádová, D., Žalud, Z., 2011. Agroclimatic conditions in Europe under climate change. Glob. Chang. Biol. 17, 2298–2318. https://doi.org/10.1111/j.1365-2486.2011.02396.x

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens. Environ. 8, 127–150. https://doi.org/10.1016/0034-4257(79)90013-0

Urbazaev, M., Thiel, C., Cremer, F., Dubayah, R., Migliavacca, M., Reichstein, M., Schmullius, C., 2018. Estimation of forest aboveground biomass and uncertainties by integration of field measurements, airborne LiDAR, and SAR and optical satellite data in Mexico. Carbon Balance Manag. 13, 5. https://doi.org/10.1186/s13021-018-0093-5

USDA, 1987. Module 3. USDA Textural Classification Study Guide, in: Soil Mechanics Level 1. National Employee Development Staff, Soil Conservation Service, U.S. Department of Agriculture.,

Washington DC.

Vargas-Rojas, R., Cuevas-Corona, R., Yigini, Y., Tong, Y., Bazza, Z., Wiese, L., 2019. Unlocking the Potential of Soil Organic Carbon: A Feasible Way Forward, in: International Yearbook of Soil Law and Policy. pp. 373–395. https://doi.org/10.1007/978-3-030-00758-4_18

Vargas, R., Alcaraz-Segura, D., Birdsey, R., Brunsell, N.A., Cruz-Gaistardo, C.O., de Jong, B., Etchevers, J., Guevara, M., Hayes, D.J., Johnson, K., Loescher, H.W., Paz, F., Ryu, Y., Sanchez-Mejia, Z., Toledo-Gutierrez, K.P., 2017. Enhancing interoperability to facilitate implementation of REDD+: case study of Mexico. Carbon Manag. 8, 57–65. https://doi.org/10.1080/17583004.2017.1285177

Vaysse, K., Lagacherie, P., 2017. Using quantile regression forest to estimate uncertainty of digital soil mapping products. Geoderma 291, 55–64. https://doi.org/10.1016/j.geoderma.2016.12.017

Verheye, W., la Rosa, D., 2005. Mediterranean soils. In land use and land cover, from Encyclopedia of life support systems (EOLSS), Developed under the Auspices of the UNESCO.

Vermote, E., 2015. MOD09Q1 MODIS/Terra Surface Reflectance 8-Day L3 Global 250 m SIN Grid V006.

Vicente-Vicente, J.L., Gómez-Muñoz, B., Hinojosa-Centeno, M.B., Smith, P., Garcia-Ruiz, R., 2017. Carbon saturation and assessment of soil organic carbon fractions in Mediterranean rainfed olive orchards under plant cover management. Agric. Ecosyst. Environ. 245, 135–146. https://doi.org/10.1016/j.agee.2017.05.020

Villarreal, S., Guevara, M., Alcaraz-Segura, D., Brunsell, N.A., Hayes, D., Loescher, H.W., Vargas, R., 2018. Ecosystem functional diversity and the representativeness of environmental networks across the conterminous United States. Agric. For. Meteorol. 262, 423–433. https://doi.org/10.1016/j.agrformet.2018.07.016

Vitharana, U.W.A., Mishra, U., Mapa, R.B., 2019. National soil organic carbon estimates can improve global estimates. Geoderma 337, 55–64. https://doi.org/10.1016/j.geoderma.2018.09.005

Wälder, K., Wälder, O., Rinklebe, J., Menz, J., 2008. Estimation of soil properties with geostatistical methods in floodplains. Arch. Agron. Soil Sci. 54, 275–295. https://doi.org/10.1080/03650340701488485

Wang, B., Waters, C., Orgill, S., Cowie, A., Clark, A., Li Liu, D., Simpson, M., McGowen, I., Sides, T., 2018a. Estimating soil organic carbon stocks using different modelling techniques in the semi-arid rangelands of eastern Australia. Ecol. Indic. 88, 425–438. https://doi.org/10.1016/j.ecolind.2018.01.049

Wang, B., Waters, C., Orgill, S., Gray, J., Cowie, A., Clark, A., Liu, D.L., 2018b. High resolution mapping of soil organic carbon stocks using remote sensing variables in the semi-arid rangelands of eastern Australia. Sci. Total Environ. 630, 367–378. https://doi.org/10.1016/j.scitotenv.2018.02.204

Webster, R., Oliver, M.A., 2001. Geostatistics for Environmental Scientists, second. ed. John Wiley & Sons. Ltd.

White, J.C., Coops, N.C., Wulder, M.A., Vastaranta, M., Hilker, T., Tompalski, P., 2016. Remote Sensing Technologies for Enhancing Forest Inventories: A Review. Can. J. Remote Sens. 42, 619–641. https://doi.org/10.1080/07038992.2016.1207484

Wiesmeier, M., Urbanski, L., Hobley, E., Lang, B., von Lützow, M., Marin-Spiotta, E., van Wesemael, B., Rabot, E., Ließ, M., Garcia-Franco, N., Wollschläger, U., Vogel, H.-J., Kögel-Knabner, I., 2019. Soil organic carbon storage as a key function of soils - A review of drivers and indicators at various scales. Geoderma 333, 149–162. https://doi.org/10.1016/j.geoderma.2018.07.026

Wilks, D.S., 2019. Statistical Methods in the Atmospheric Sciences. 4th Edition. (International Geophysics). Elsevier. https://doi.org/10.1016/C2017-0-03921-6

Willaarts, B.A., Oyonarte, C., Muñoz-Rojas, M., Ibáñez, J.J., Aguilera, P.A., 2016. Environmental Factors Controlling Soil Organic Carbon Stocks in Two Contrasting Mediterranean Climatic Areas of Southern Spain. L. Degrad. Dev. 27, 603–611. https://doi.org/10.1002/ldr.2417

Wilson, D.S., Jamieson, S.S.R., Barrett, P.J., Leitchenkov, G., Gohl, K., Larter, R.D., 2012. Antarctic topography at the Eocene–Oligocene boundary. Palaeogeogr. Palaeoclimatol. Palaeoecol. 335–336, 24–34. https://doi.org/10.1016/j.palaeo.2011.05.028

WRB-IUSS, 2014. World Reference Base for Soil Resources. 2014, International Soil Classification System for Naming Soils and Creating Legends for Soil Maps. FAO, Rome, Italy.

Wulder, M.A., White, J.C., Nelson, R.F., Næsset, E., Ørka, H.O., Coops, N.C., Hilker, T., Bater, C.W., Gobakken, T., 2012. Lidar sampling for large-area forest characterization: A review. Remote Sens. Environ. 121, 196–209. https://doi.org/10.1016/j.rse.2012.02.001

Xiao, J., Moody, A., 2004. Photosynthetic activity of US biomes: responses to the spatial variability and seasonality of precipitation and temperature. Glob. Chang. Biol. 10, 437–451. https://doi.org/10.1111/j.1365-2486.2004.00745.x

Xiong, X., Grunwald, S., Corstanje, R., Yu, C., Bliznyuk, N., 2016. Scale-dependent variability of soil organic carbon coupled to land use and land cover. Soil Tillage Res. 160, 101–109. https://doi.org/10.1016/j.still.2016.03.001

Xu, B., Huang, J.Z., Williams, G., Wang, Q., Ye, Y., 2012. Classifying Very High-Dimensional Data with Random Forests Built from Small Subspaces. Int. J. Data Warehous. Min. 8, 44–63. https://doi.org/10.4018/jdwm.2012040103

Xu, L., Yu, G., He, N., Wang, Q., Gao, Y., Wen, D., Li, S., Niu, S., Ge, J., 2018. Carbon storage in China's terrestrial ecosystems: A synthesis. Sci. Rep. 8, 2806. https://doi.org/10.1038/s41598-018-20764-9

Yan, F., Wu, B., Wang, Y., 2015. Estimating spatiotemporal patterns of aboveground biomass using Landsat TM and MODIS images in the Mu Us Sandy Land, China. Agric. For. Meteorol. 200, 119–128. https://doi.org/10.1016/j.agrformet.2014.09.010

Yan, X., Su, X., 2009. Linear Regression Analysis: Theory and Computing. World Scientific.

Yigini, Y., Olmedo, G.F., Reiter, S., Baritz, R., Viatkin, K., Vargas, R., 2018. Soil organic carbon mapping: Cookbook 2nd Edition. FAO, Rome.

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. https://doi.org/10.1016/j.geoderma.2018.09.006

Zhang, C., Ma, Y., 2012. Ensemble Machine Learning: Methods and applications. Springer US, Boston, MA. https://doi.org/10.1007/978-1-4419-9326-7

Zhao, M., Yang, J., Zhao, N., Liu, Y., Wang, Y., Wilson, J.P., Yue, T., 2019. Estimation of China's forest stand biomass carbon sequestration based on the continuous biomass expansion factor model and seven forest inventories from 1977 to 2013. For. Ecol. Manage. 448, 528–534. https://doi.org/10.1016/j.foreco.2019.06.036

Zhou, G., Liu, S., Li, Z., Zhang, D., Tang, X., Zhou, C., Yan, J., Mo, J., 2006. Old-Growth Forests Can Accumulate Carbon in Soils. Science (80-.). 314, 1417–1417. https://doi.org/10.1126/science.1130168

UNIVERSIDAD
DE ALMERÍA

CAESCG