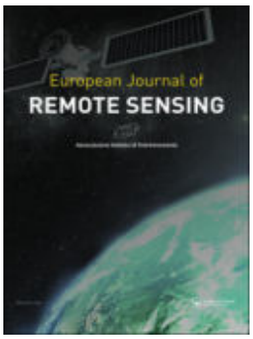


European Journal of Remote Sensing

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tejr20

Digital mapping of soil organic carbon stocks in the forest lands of Dominican Republic

Efraín Duarte, Erick Zagal, Juan A. Barrera, Francis Dube, Fabio Casco & Alexander J. Hernández

To cite this article: Efraín Duarte, Erick Zagal, Juan A. Barrera, Francis Dube, Fabio Casco & Alexander J. Hernández (2022) Digital mapping of soil organic carbon stocks in the forest lands of Dominican Republic, European Journal of Remote Sensing, 55:1, 213-231, DOI: 10.1080/22797254.2022.2045226

To link to this article: https://doi.org/10.1080/22797254.2022.2045226

Published online: 20 Mar 2022.

EUROPEAN JOURNAL OF REMOTE SENSING
2022, VOL. 55, NO. 1, 213–231
https://doi.org/10.1080/22797254.2022.2045226

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS Check for updates

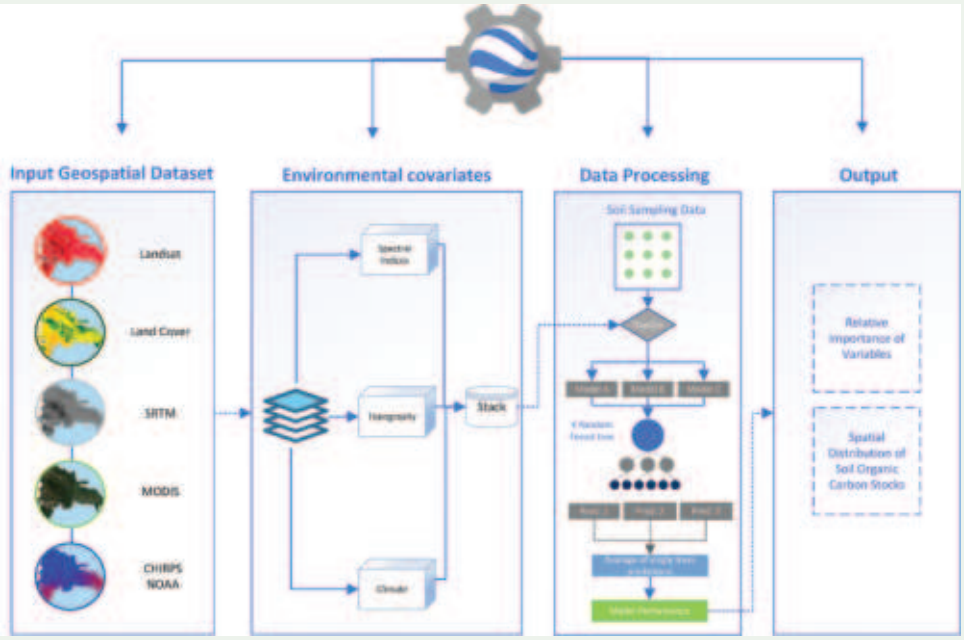
Digital mapping of soil organic carbon stocks in the forest lands of Dominican Republic

Efraín Duarte[a,b], Erick Zagal[a], Juan A. Barrera[a], Francis Dube[c], Fabio Casco[d] and Alexander J. Hernández[e]

[a]Department of Soils and Natural Resources, Faculty of Agronomy, University of Concepcion, Chillán, Chile; [b]Faculty of Agronomy, University of Concepcion, Chillán, Chile; [c]Department of Silviculture, Faculty of Forest Sciences, University of Concepcion, Concepción, Chile; [d]IR3 initiative, Food and Agriculture Organization (FAO) of the United Nations, Tegucigalpa, Honduras; [e]United States Department of Agriculture, Agricultural Research Service., Utah State University, Logan, Utah, USA

ABSTRACT

Mapping the spatial distribution of soil organic carbon (SOC) in lands covered by tropical forests is important to understand the relationship and dynamics of SOC in this type of ecosystem. In this study, the Random Forest (RF) algorithm was used to map SOC stocks of topsoil (0–15 cm) in forest lands of the Dominican Republic. The methodology was developed using geospatial datasets available in the Google Earth Engine (GEE) platform combined with a set of 268 soil samples. Twenty environmental covariates were analyzed, including climate, topography, and vegetation. The results indicate that Model A (combining all 20 covariates) was only marginally better than Model B (combining topographic and climatic covariates), and Model C (only combining multispectral remote sensing data derived from Landsat 8 OLI images). Model A and Model B yielded SOC mean values of 110.35 and 110.87 Mg C ha^{-1}, respectively. Model A reported the lowest prediction error and uncertainty with an R^2 of 0.83, an RMSE of 35.02 Mg C ha^{-1}. There was a strong dependence of SOC stocks on multispectral remote sensing data. Therefore, multispectral remote sensing proved accurate to map SOC stocks in forest ecosystems in the region.



ARTICLE HISTORY
Received 9 June 2021
Revised 18 February 2022
Accepted 18 February 2022

KEYWORDS
Random forest; landsat; machine learning; tropical forest; environmental covariates; Google Earth Engine

Introduction

Soils hold the largest carbon (C) pool on Earth after the oceans, with an estimated total of 1,500–2,400 Pg C up to 1 m depth (Scharlemann et al., 2014; Tifafi et al., 2018). Soil organic carbon (SOC) directly influences the physicochemical properties nutrient retention capacity and infiltration rate of the soil (Scholten et al., 2017; Viscarra Rossel et al., 2016). In addition, SOC has the potential to mitigate the adverse impacts of current and future climate change (Edenhofer et al., 2014) and help improve the primary productivity of the biosphere (Grinand et al., 2017).

The Global Climate Observing System (GCOS) has identified 54 Essential Climate Variables (ECVs),

CONTACT Erick Zagal ezagal@udec.cl Department of Soils and Natural Resources, Faculty of Agronomy, University of Concepcion, Vicente Méndez 595, Casilla 537, Chillán 3812120, Chile

including components of the cryosphere, biosphere, and hydrosphere that are "needed to understand and predict the evolution of climate, to guide mitigation and adaptation measures, to assess risks and enable attribution of climate events to underlying causes, and to underpin climate services" (WMO, 2020). The present work focuses on one of these ECVs, SOC stocks, particularly those stored in lands covered by tropical forests.

The global C stocks in forest biomass and their spatial distribution are relatively well documented and have been estimated with reasonable accuracy compared to SOC stocks (Baccini et al., 2012; Harris et al., 2012; Ruesch & Gibbs, 2008; Saatchi et al., 2011). Most local and international policies for climate change mitigation have focused on conserving and studying the C stored in forests. In addition to this, SOC is of great importance because the Earth's soils store two-to-three times as much carbon in organic form as there is C in the atmosphere globally (Trumbore, 2009). In this sense, the construction of a robust and transparent system to measure, report and verify (MRV) SOC changes represents a key tool to support compliance with the Sustainable Development Goals (SDG), specifically the SDG indicator 15.3.1 "Proportion of land that is degraded over the total land area" (FAO, 2020; Jan & Jeffrey, 2018).

At present, there is still uncertainty about the amount of global SOC stocks and their spatial distribution, mainly due to the little attention given by decision makers at the local, national and international levels (Gianelle et al., 2010). In an analysis and review of 27 studies that estimated SOC globally, it was found that the SOC mean value is approximately 1,460.5 Pg C, ranging from 504 to 3,000 Pg C (Scharlemann et al., 2014). One of the main reasons for the uncertainties found in these estimates is the large number of factors that interfere in SOC dynamics combined with all the uncertainties leading to error propagation associated with the difficulty in assessing C and soil bulk density (Köchy et al., 2015).

Even though there is scientific interest in monitoring forests and soils, there is a lack of data to carry out efficient monitoring and determine the current state of these resources (Liang et al., 2016). In 2017, the Food and Agriculture Organization (FAO), the Intergovernmental Technical Panel on Soils (ITPS), the Intergovernmental Panel on Climate Change (IPCC), the Global Soil Partnership (GSP), the Science-Policy Interface of the United Nations Convention to Combat Desertification (UNCCD-SPI), and the World Meteorological Organization (WMO) jointly organized a Global Symposium on Soil Organic Carbon. This symposium provided guidelines for developing efficient systems and protocols for measuring SOC with higher accuracy (FAO, 2020a). In the last decade, digital soil mapping (DSM) approaches have focused on mapping SOC using remote sensing techniques as the main emerging tool to improve spatial estimates of SOC (Mahmoudzadeh et al., 2020; Padarian et al., 2019).

DSM allows quantifying the spatial variation of SOC stocks using environmental covariates (Zhang et al., 2017), which describe the relationship of a soil attribute and its spatially implicit forming factors (Jenny, 1941). The environmental auxiliary variables of SOC can be obtained from digital elevation models (DEM) (Farr et al., 2007; B. Wang et al., 2018), remote sensing data (Duarte et al., 2020; Wulder et al., 2016; Xiao et al., 2019) and climatic data (Ermida et al., 2020; Veronesi & Schillaci, 2019). The easy accessibility of satellite images combined with Machine Learning (ML) techniques has significantly improved the accuracy of SOC mapping. In a review of 120 studies on SOC mapping, in which different ML techniques were applied, it was found that the Random Forest (RF) algorithm has optimum performance in the selection of environmental covariates for SOC mapping. At the same time, it also behaves better than other ML techniques and Multiple Linear Regression (MLR) (Lamichhane et al., 2019).

There are few studies on SOC mapping of Dominican Republic lands. The main report comes from the Global Soil Organic Carbon Map (GSOCmap) launched by FAO. In fact, GSOCmap represents the first global estimation of SOC content carried out with a participatory approach to compile all the available data on soils at the national level (FAO, 2020). With regard to the tropics, SOC estimates are very limited on forest lands since most of the research has focused on estimating SOC from agricultural lands.

The objective of this study was to estimate SOC content in forest lands of the Dominican Republic and their spatial distribution by applying ML techniques, a dataset of environmental covariates obtained from remote sensing (RS) and field data. We compared the influence of three groups of predictive variables for SOC mapping: (1) multispectral remote sensing variables, (2) topographic variables, and (3) climatic variables. The performance of the ML model was also evaluated. Our model was implemented in the Google Earth Engine (GEE) cloud-based computing platform (Gorelick et al., 2017).

Materials and methods

The overall structure of the method (Figure 3) consisted of five stages such as: (i) selection of a geospatial dataset; (ii) data pre-processing; (iii) model building/development; (iv) evaluation of the model performance; and (v) mapping of SOC.

Study area

The study was carried out in the Island of Hispaniola (central region of the Caribbean), Dominican Republic. It corresponds to forest lands located between 17°36′ and 19°58′ latitude north, and 68°19′ and 72°01′ longitude west, belonging to the Greater Antilles. The territory of the country covers 48,198 km^2 (Figure 1).

The Dominican Republic has a tropical climate, which is altered only by the Alyssian winds of the Atlantic and topographical factors of the Island. The average annual temperature is 25°C, with August being the hottest month and January the coldest. Precipitation is distributed in two seasons: a rainy season, which goes from April to June and from September to November, with precipitation of 2,500 mm yr^{-1}; and a dry season, which goes from December to March, with precipitation of 450 mm yr^{-1}. On the Island, the areas with the highest humidity are in the north because they are influenced by the Atlantic Ocean, while the driest areas are found to the south, along the Caribbean coast (Cano-Ortiz et al., 2015).

The country's native forests (Figure 2) include pine, broadleaf, dry and mangrove forests. Pine forests are primarily made up of *Pinus occidentalis*, a species endemic to the Island (Kennedy et al., 2005). The composition of the broadleaf forests is diverse, with species such as *Swietenia mahagoni, Ocotea spp., Sloanea berteriana, Didymopanax tremulus*, and *Clusia rosea*. Dry forests include various species such as *Guaiacum officinale, Phylostilum braziliensis*, and *Prosopis julifora*, while mangrove forests are composed of *Avicennia germinans, Laguncularia racemosa, Conocarpus erectus* and *Rizophora mangle* (MARN, 2019). Shrubland and herbaceous vegetation can also be found (Martin & Fahey, 2006).

The soils are divided into 10 classes based on characteristics such as depth, slope, and drainage. Soil classes correspond to savannah, non-calcareous clay and calcareous soils as well as soils derived from igneous rocks, soils of volcanic and metamorphic origins, recent alluvial soils, organic soils, wetlands, coastal beaches and dunes (MARN, 2012). Our study focused on soils covered by pine, broadleaf, dry and mangrove forests, covering an area of 14,499 km^2.

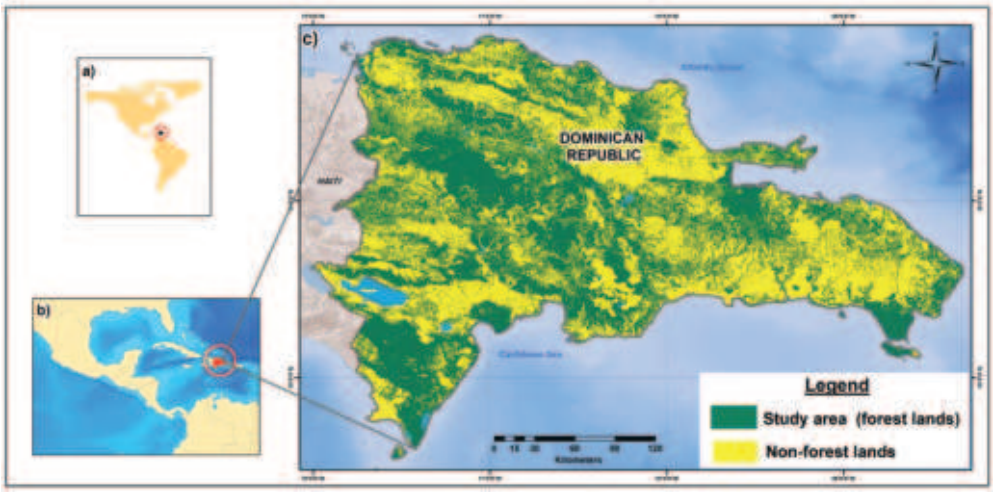


Figure 1. Study area: (a) General location of the Dominican Republic, (b) regional location, (c) study area (forest lands).

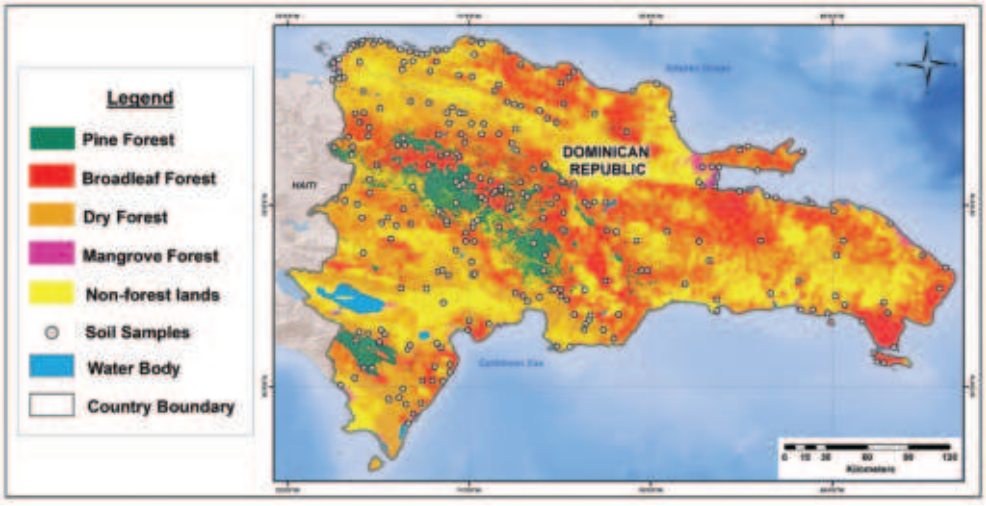


Figure 2. Map of soil sample location and forest types of the Dominican Republic.

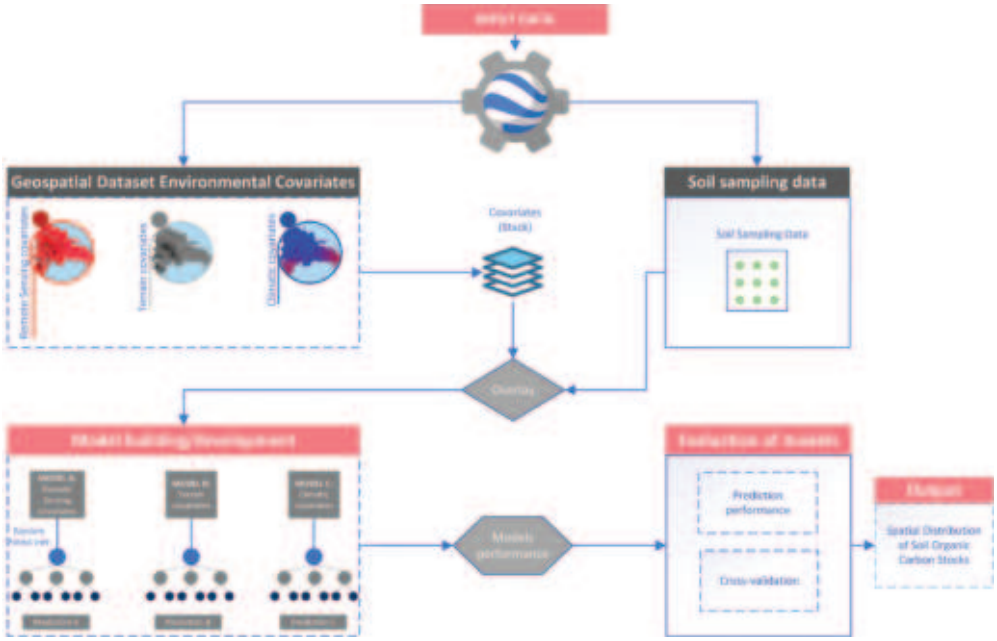


Figure 3. Flowchart: Dataset of environmental covariates obtained from remote sensing and combined with soil samples to define different predictive models using machine learning for mapping the spatial distribution of SOC.

Reference data

Forest mask

To estimate SOC stocks, the study area was defined with a forest mask. For this, we used the land cover map prepared in 2015 by the Ministry of the Environment and Natural Resources (MARN) of the Dominican Republic. On this map, the concept of forest was applied in accordance with the FAO Forest Resources Assessment (FRA-FAO) defined by the country as *"Natural or planted ecosystem of at least 0.5 hectares covered by trees higher than 5 meters and with a canopy cover of more than 40%"* (FAO, 2020b). Forest types (Figure 2) that are part of this study are defined as follows (MARN, 2012):

Broadleaf forests. represented by trees where the combination of broad-leaved species predominates; it comprises plant communities from semi-humid in transition to cloudy. It is the type of forest with the largest existence in the country. It is classified as cloudy broadleaf forest, located in areas with elevations from 600 to 2,300 m above sea level (m.a.s.l.); moist broadleaf forest in areas with elevations from 300 to 1,500 m.a.s.l. and semi-humid broadleaf forest located in areas with elevations up to 900 meters above sea level, or m.a.s.l.

Dry forests. mostly secondary forest; they are composed of semi-deciduous trees that develop at elevations below 500 m.a.s.l. The greatest presence of these forests is located in the lowlands, both south-southwest and northwest of the country.

Pine forests. composed of pine species. The authentic Dominican pine forest is located especially in the highlands and the dominant species is *Pinus occidentalis*, which is found in the large mountain ranges with elevations above 2,000 m.a.s.l. The pine forest has two types of cover: high and low canopy cover density.

Mangrove forests. Coastal and wet ecosystems found in swampy and flooded regions; they mainly belong to the Rhizophoraceae family, with exposed supporting roots. This forest is located at elevations below 20 m.a.s.l.

Soil organic carbon database

In this study, 268 soil samples from the National Forest Inventory (NFI) were used. The NFI was collected in 2018 by the MARN of the Dominican Republic with the support of the REDD/CCAD-GIZ program and the World Bank's Forest Carbon Partnership Facility (FCPF) (available in: https://www.sica.int/documentos/inventario-forestal-nacional-de-republica-dominicana_1_126744.html) (MARN, 2021). The sampling design adopted by the NFI corresponds to stratified sampling for each forest stratum. For building the NFI, the methodology proposed by the REDD/CCAD-GIZ program was used (Feliz et al., 2019; MARN, 2021).

Data used were taken from 268 plots located in the different forest types (Figure 2; Table 1). Each soil sample was collected from the NFI at a depth of 0 to 15 cm and their geographic coordinates were recorded with a global positioning system device (GPS). In the NFI, SOC data were reported with an extrapolation from depth of 0 to 15 cm.

Table 1. Descriptive statistics of soil organic carbon (SOC) stocks (Mg C ha^{-1}) (0–15 cm depth) collected from the National Forest Inventory (NFI).

	SOC (Mg C ha^{-1})							
Description	n	Mean	Median	Min	Max	SD	Skewness	Kurtosis
Dry forest	52	126.47	124.97	18.88	282.38	64.02	0.49	−0.12
Pine forest	43	69.68	57.64	19.78	187.81	41.53	1.03	0.34
Broadleaf forest	129	105.57	94.64	15.95	274.49	61.93	0.65	−0.30
Mangrove forest	44	145.06	140.53	27.92	261.87	63.43	−0.07	−0.88
Total forest	*268*	*110.35*	*101.34*	*15.95*	*282.38*	*63.78*	*0.57*	*−0.46*

Min: minimum; Max: maximum; SD: standard deviation

The samples were numbered, bagged, and brought back to the Laboratory of Soils and Water of the Dominican Institute of Agricultural and Forestry Research Agricultural Technology Center (CENTA) and the Dominican Agribusiness Laboratory (LAD). After drying, the samples were weighed and passed through a 2-mm sieve. The determination of SOC content (Mg C ha^{-1}) is based on the Walkley & Black chromic acid wet oxidation method (Walkley & Black, 1934).

Bulk density (BD) was determined on subsamples dried at 105°C as described by (Dane Topp et al., 2002). Results were reported as g cm^{-3} on an oven-dry basis and SOC was reported as g (100 g)$^{-1}$. Soil organic carbon stock (SOCS) was computed as the product of three variables, organic carbon content (C), bulk density (BD), and thickness (D). SOCS was calculated according to Equation 1:

$$SOC_{stock} = C \times BD \times D \times \left(1 - \frac{gravel[\%]}{100}\right) \quad (1)$$

Where C is the concentration of soil carbon (g C (100 g)$^{-1}$); BD is bulk density (g cm^{-3}), D is the thickness of the layer (cm), gravel [%] is the percentage of gravel in the soil sample.

Table 1 shows the descriptive statistics of SOC (0–15 cm depth) samples collected from the NFI. SOC contents ranged from 15.95 to 282.38, with a mean value of 110.35 and a median of 101.34 Mg C ha^{-1}. The coefficient of skewness is −0.46 Mg C ha^{-1}. The sampling point´s standard deviation (SD) is 63.78 Mg C ha^{-1} and is lower than the mean value.

Once the forest mask and environmental variables was defined, a method to estimate SOC content in forest soils with a geospatial dataset was developed, they were combined with the soil samples collected in the field, and a regression model was applied for each of the three models; the covariates were divided using the RF algorithm in the GEE platform; finally, the models were evaluated and the spatial distributions and SOC stock map was built (Figure 3).

We iterated the model A, B, and C and calculated the average standard deviation (SDs) to analyze the uncertainty of each model in predicting topsoil SOC (Figure 6).

Environmental predictors

For the digital mapping of SOC, we selected 20 accessible and commonly used predictive environmental dataset covariates, which represent key factors for the spatial distribution and formation of SOC content such as: vegetation, soil, topography, and climate (McBratney et al., 2003). These covariates represent factors of soil formation according to (Jenny, 1941). Further spectral vegetation indices (SVIs) were calculated using Landsat-8 images (Table 2). From the combination of these dataset covariates with data soil samples, three models with different combinations of predictive variables were built, using the RF algorithm for the digital mapping of SOC. The models were as follows:

- **Model A**: Multispectral remote sensing variables + topographic variables + climatic variables.
- **Model B**: Topographic and climatic variables.
- **Model C**: Multispectral remote sensing variables.

For each of the models, the RF algorithm was applied, and its accuracy was evaluated. The relative importance of the variables in the model was also assessed. All the datasets were obtained and processed in the GEE cloud-based platform.

Multispectral imagery. We used Landsat 8 collection 1 Tier 1 Operational Land Imager (OLI) surface reflectance data with 16 days and 30 m resolutions, available in the GEE platform, using the Landsat Digital Number (DN) values, representing scaled, calibrated at-sensor radiance; GEE collection snippet: ee. ImageCollection("LANDSAT/LO08/C01/T1_RT").

All Landsat-8 OLI surface reflectance data from the year 2018 ± 0.5 (available in the GEE platform) were used in this study: a total of 92 images from path 123 and row 32; 89 images from path 125 and row 34 and 94 images from path 125 and row 33. Landsat surface reflectance data were atmospherically corrected using the Landsat Surface Reflectance Corrected (LaSRC) (OLI) algorithms (Masek et al., 2013); We used methods provided by Earth Engine for filtering image collections using the code "imageCollection.filterDate()" and we built a composite mosaic multiband and multi-

Table 2. Predictive covariates derived from Landsat 8 OLI.

Dataset (Covariates)	Abbr.	Formula	References
Remote Sensing-derived covariates			
Normalized Difference Vegetation Index	NDVI	$NDVI = \frac{(NIR-\mathrm{Red})}{(NIR+Red)}$	(Sobrino & Raissouni, 2000)
Enhanced Vegetation Index	EVI	$EVI = 2.5 * \frac{(NIR+Red)}{(NIR+6*Red-7.5*Blue+L)}$	(Huete et al., 1997)
Soil-Adjusted Vegetation Index	SAVI	$SAVI = \frac{(NIR-Red)}{(NIR+Red+L)} * (1+L)$	(Huete, 1988)
Index-Based built-up Index	IBI	$IBI = \frac{NDBI-(SAVI+MNDWI)/2}{NDBI+(SAVI+MNDWI)/2}$	(Xu, 2008)
Bare Soil Index	BSI	$BSI = \frac{(SWIR1+Red)-(NIR+Blue)}{(SWIR1+Red)+(NIR+Blue)}$	(Piyoosh & Ghosh, 2018)
Green Normalized Difference Vegetation Index	GNDVI	$GNDVI = \frac{(NIR-Green)}{(NIR+Green)}$	(Alba et al., 2017)
Near-infrared reflectance of vegetation	NIRv	$NIRv = (NDVI_{medianmonthly} - 0.08) xNIR_{medianmonthly}$	(Badgley et al., 2017)
Band 2 Blue	B_{BLUE}	Wavelength of 0.450–0.515 µm	
Band 3 Green	B_{GREEN}	Wavelength of 0.525–0.600 µm	
Band 4 Red	B_{RED}	Wavelength of 0.630–0.680 µm	
Band 5 Near Infrared	NIR	Wavelength of 0.845–0.885 µm	
Band 6 Shortwave Infrared-1	SWIR1	Wavelength of 1.560–1.660 µm	
Band 7 Shortwave Infrared-2	SWIR2	Wavelength of 2.100–2.300 µm	

date formed by combining spatially overlapping images into a single image based on a function of multiple spectral and temporal aggregation ranges; compositing in GEE refers to the process of combining spatially overlapping images into a single image based on an aggregation function, and mosaicking refers to the process of spatially assembling image datasets to produce a spatially continuous image. (For more details see GEE codes developed in this study available in Appendix A or visit https://developers.google.com/earth-engine/guides/ic_composite_mosaic)

The CFmask algorithm was used to mask cloud and shadow produced, as well as a per-pixel saturation mask (Zhu & Woodcock, 2012). In addition, the empirical Earth rotation model (ERM) was used as a basis to perform a terrain illumination correction algorithm (Tan et al., 2013), which allowed conducting the topographic correction for each image. For reflectance images, we used the medoid method (Flood, 2013).

Following radiometric, geometric, and atmospheric corrections, digital numbers for the blue (B1), green (B2), red (B3), near-infrared (B4), and shortwave IR-2 bands (B6) were extracted. Several spectral indices

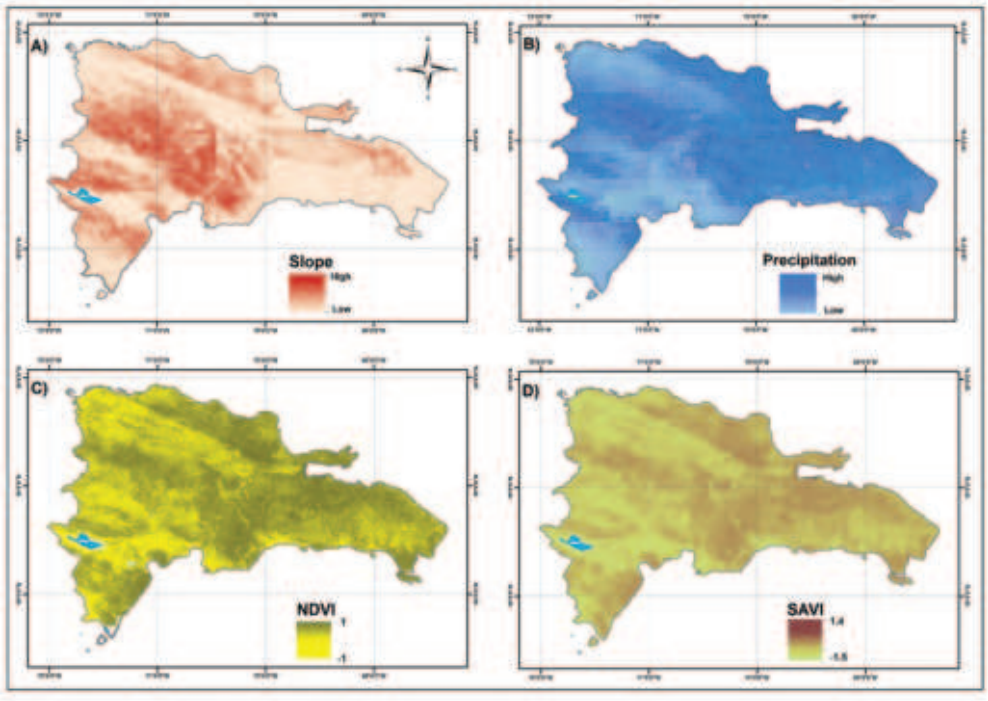


Figure 4. Example of the geospatial dataset used and computed on the Google Earth Engine platform. (a) slope; (b) precipitation; (c) Normalized Difference Vegetation Index (NDVI); (d) Soil-Adjusted Vegetation Index (SAVI).

were then calculated: The Bare Soil Index (BSI), the Normalized Difference Vegetation Index (NDVI), the Soil-Adjusted Vegetation Index (SAVI), the Index-Based built-up Index (IBI), the Enhanced Vegetation Index (EVI) and the Green Normalized Difference Vegetation Index (GNDVI) with a spatial resolution of 30 m. (Table 2) and (Figure 4).

Climatic variables. We used the climatic datasets available in the GEE platform closest to the date soil sampling, such as Moderate Resolution Imaging Spectroradiometer (MODIS) MOD11A1 V6 product, which provides daily land surface temperature (LST) and emissivity Daily Global 1 km. *GEE collection snippet*: *ee.ImageCollection("MODIS/006/MOD11A1")*. The temperature value is derived from the MOD11_L2 swath product (Wan, 2014).

Relative humidity data (2 m above ground) were obtained from the Global Forecast System (GFS). This is a weather forecast model produced by the National Centers for Environmental Prediction (NCEP) and National Aeronautics and Space Administration (NASA). The GFS is a coupled model, composed of an atmosphere model, an ocean model, a land/soil model, and a sea ice model, which work together to provide an accurate picture of weather conditions (Saha et al., n.d.). *GEE Collection snippet*: *ee.ImageCollection ("NOAA/GFS0P25")*.

For the precipitation data, we used the Climate Hazards Group InfraRed Precipitation with Stations (CHIRPS) dataset, which builds on previous approaches to "smart" interpolation techniques and high resolution, long period of record precipitation estimates based on infrared Cold Cloud Duration (CCD) observations (Funk et al., 2015). *GEE collection snippet*: *ee.ImageCollection("UCSB-CHG/CHIRPS/DAILY")*.

Topographic variables. Terrain analysis is crucial for modeling environmental systems. Specifically, the topography is considered as a variable that can largely explain SOC changes. In fact, models that take topographic attributes into account can provide better estimates of SOC stocks (McBratney et al., 2003). We used the digital elevation model (DEM) derived from NASA's Shuttle Radar Topography Mission (SRTM DEM) (Farr et al., 2007). We calculated the topographic slope, aspect and elevation from this SRTM V3 product (SRTM Plus) at a resolution of 1 arc-second (approximately 30 m). GEE collection Snippet: *ee.Image("USGS/SRTMGL1_003")*. The number of terrain and climate-based covariates used within each dataset is shown in Table 3, and an example of its geographical representation is shown in Figure 4.

Table 3. Terrain and climate-based covariates.

Dataset (Covariates)	Abbreviation	Definition
Terrain-based covariates		
(1) Elevation	Elev	Height above sea level (m)
(1) Slope	Slo	Average gradient above flow path
(1) Aspect	Asp	The compass direction of the maximum rate of change
(1) Topographic Wetness index	TWI	Combined local upslope contributing area and slope
Climatic-based covariates		
(1) Temperature	Temp	It is derived from the daily temperature values
(1) Precipitation	Prec	It is derived from the daily precipitation values
(1) Relative Humidity	RH	Water vapor in the air, compared to how much it could hold a specific temperature

Data processing

Google earth engine (GEE) platform

The model for mapping SOC stocks in forests developed in the present study was built in the GEE cloud-based computing platform. GEE is a platform designed for scientific analysis at the petabyte (PB) scale and has an extensive public data catalog for earth observation (Gorelick et al., 2017). One way to use this platform is using an online tool called The Code Editor, which lets the user access the platform using a scripting language (JavaScript).

GEE has hosted historical images of the Earth for more than forty years. The images collected daily are made available to the public for data mining on a global scale. GEE allows processing massive data of a raster format for large areas and with high volumes of information. In our case, topographic, climatic and vegetational variables were analyzed (Tables 2 and 3) with high performance and minimum user involvement in the processing. The algorithm used in GEE to estimate SOC contents was the Random Forest (RF). The codes developed in this study using the GEE cloud-based computing platform are available in Appendix A.

Random forest (RF) modelling

In this study, the RF algorithm was selected to predict the SOC stocks in the forest ecosystem of the Dominican Republic. RF is one of the most popular and most powerful supervised ML algorithms that can perform both regression and classification tasks. As the name suggests, this algorithm builds a set of regression trees. Each of the trees predicts the result in each pixel, while the final prediction is obtained averaging these values (Breiman, 2001). We used the RF to estimate the relative importance of the predictive variables.

For the prediction accuracy, the 268 SOC samples were randomly divided into 2 sets: 70% of the total samples were used as model training data (n = 188), and the remaining 30% for model validation and accuracy assessment (n = 80). RF modelling was performed using the GEE cloud computing platform applying the following line of code: *ee.Classifier.smileRandomForest.setOutputMode("REGRESSION")*. The principal parameters of the algorithm were: number of decision trees = 100 and default values to min leaf population (1), variables per split (square root of the number of variables), bag fraction (0.5), max nodes (defaults to no limit), seed (0) and set output mode = regression.

Model evaluation and uncertainty analysis

To evaluate the performance of the SOC model, five indexes were calculated using the following formulas: coefficient of determination (R^2), Lin's concordance correlation coefficient (LCCC) (Lin, 1989), root-mean-square error (RMSE), mean absolute percentage error (MAPE), and mean absolute deviation (MAD).

$$R^2 = \frac{\left(\sum_{i=1}^{n}\left(y_i-\bar{y}_i\right)\left(f_i-\bar{f}_i\right)\right)^2}{\sum_{i=1}^{n}\left(y_i-\bar{y}_i\right)^2\sum_{i=1}^{n}\left(f_i-\bar{f}_i\right)^2} \quad (2)$$

$$LCCC = \frac{2r\partial_x\partial_y}{\partial_y^2+\partial_f^2+\left(\bar{y}+\bar{f}\right)^2} \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i-f_i\right)^2} \quad (4)$$

$$MAPE = \frac{\sum\left|\frac{y_i-f_i}{y_i}\right|}{n}\times 100 \quad (5)$$

$$MAD = \frac{\sum|y_i-f_i|}{n} \quad (6)$$

where $n(i=1,2,\ldots,n)$ is the number of samples used for the ML model, y_i is the value observed (Mg C ha^{-1}), $\bar{y}_i$ is the corresponding mean value, f_i is the predicted value (Mg C ha^{-1}), $\bar{f}_i$ is mean value. ∂_x and ∂_y are the variances of the predicted and measured values; r is the correlation coefficient between the predicted value and the measured value.

The prediction algorithm with the lowest MAD, MAPE and RMSE, and highest R^2 and LCCC values are determined as the best model for SOC prediction. We iterated model A, B, and C and calculated the average standard deviation (SDs) to analyze the uncertainty of each model in predicting topsoil SOC (Figure 6).

Results

Exploratory data analysis

Geospatial environmental predictive datasets

The summary statistics for each of the 20 environmental covariates at each sampling site used in the present study are shown in Table 4. To describe the

Table 4. Descriptive statistics of remote sensing-derived environmental variables at sample sites.

Dataset	Unit	Mean	Median	Min	Max	SD	Skewness	Kurtosis
Remote Sensing imagery								
NDVI	digital number	0.79	0.80	0.43	0.93	0.08	−1.02	1.22
EVI	digital number	2.36	2.35	0.90	3.56	0.44	−0.20	0.09
SAVI	digital number	1.18	1.21	0.65	1.40	0.13	−1.02	1.22
IBI	digital number	−0.27	−0.28	−0.54	0.02	0.10	0.00	−0.06
BSI	digital number	−0.29	−0.30	−0.58	0.07	0.12	0.28	−0.19
GNDVI	digital number	0.72	0.73	0.49	0.85	0.07	−0.76	0.40
NIRv	digital number	2,195.10	2,131.65	788.13	4,209.52	623.93	0.28	−0.29
B_{BLUE}	digital number	303.68	301.00	125.00	600.00	81.57	0.55	0.65
B_{GREEN}	digital number	497.71	488.00	245.00	1,007.00	114.20	0.63	1.22
B_{RED}	digital number	351.11	327.00	115.00	1,048.00	131.60	1.52	4.17
NIR	digital number	3,057.91	2,987.00	1,445.00	5,235.00	624.17	0.28	0.08
SWIR1	digital number	1,510.44	1,537.00	583.00	2,698.00	404.05	0.11	0.16
SWIR2	digital number	648.53	609.00	178.00	1,606.00	256.88	0.93	1.45
Terrain-based covariates								
Elevation	m	438.60	279.75	−1.17	2,359.85	499.42	1.65	2.43
Slope	degree	10.51	8.90	0.00	45.32	9.38	0.84	0.14
Aspect	degree	187.18	185.99	0.09	359.61	116.74	−0.07	−1.29
TWI	digital number	9.35	8.99	0.00	19.19	2.56	0.37	3.29
Climatic-based covariates								
Temperature means	degree Celsius	26.66	27.85	0.00	33.88	5.79	−3.31	12.76
Precipitation means	mm/day	3.70	3.58	0.00	7.41	1.45	0.37	−0.62
Relative humidity Means	%	77.15	76.36	62.47	86.69	5.66	0.06	−1.06

NDVI, Normalized Difference Vegetation Index; EVI, Enhanced Vegetation Index; SAVI, Soil Adjust Vegetation Index; IBI, Index-Based built-up Index; BSI, Bare Soil Index; GNDVI, Green Normalized Difference Vegetation Index; NIRv, Near-infrared reflectance of vegetation; B_{BLUE}, Landsat 8 blue band reflectance; B_{GREEN}, Landsat 8 green band reflectance; B_{RED}, Landsat 8 red band reflectance; NIR, Landsat 8 band 5 near-infrared; SWIR1, Landsat 8 band 6 Shortwave infrared-1; SWIR2, Landsat 8 band 7 Shortwave infrared-2; TWI, Topographic Wetness Index.

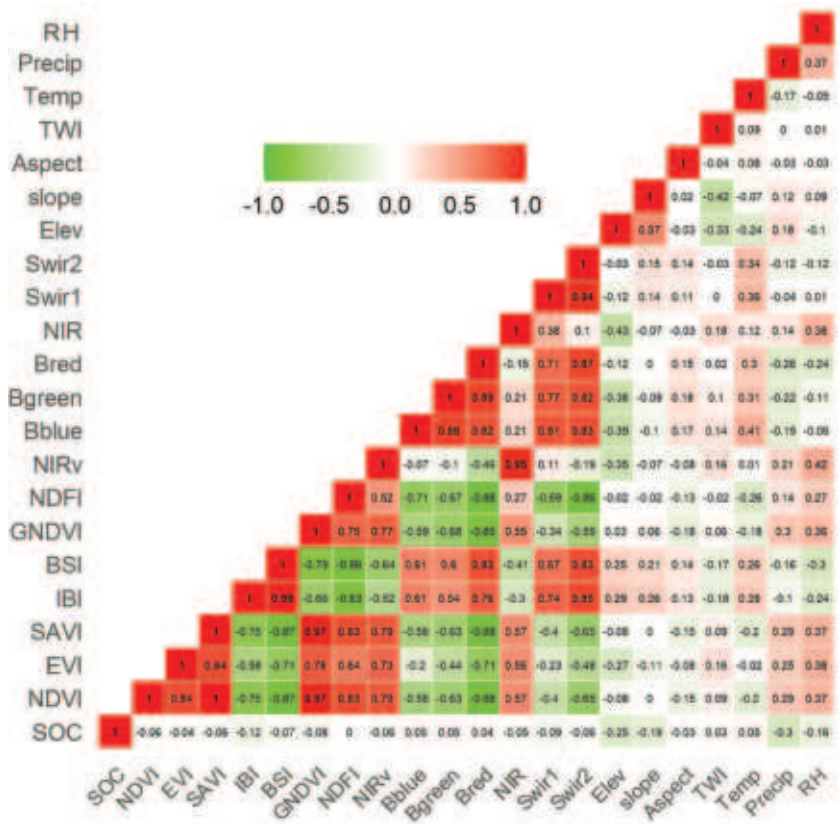


Figure 5. Pearson`s correlation coefficient analysis between observed SOC stocks and all environmental variables based on 268 sample sites.

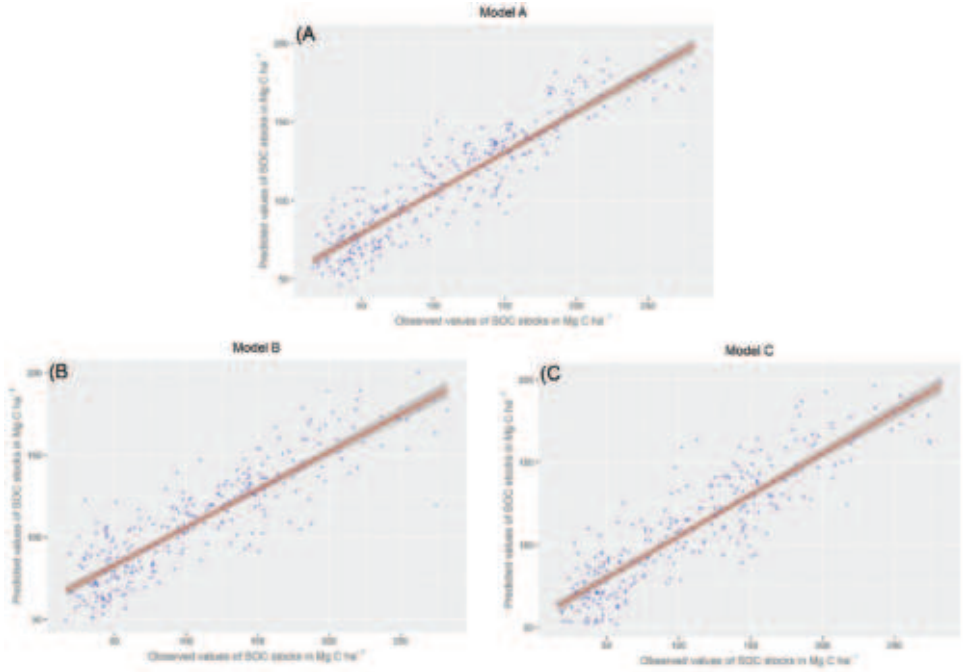


Figure 6. Scatter plot of observed vs. predicted values (Mg C ha^{-1}) for SOC content (0–15 cm) using the Random Forest Algorithm. (a) The predicted data are derived from Model A: Multispectral remote sensing variables + topographic variables + climatic variables; (b) Model B: Topographic and climatic variables and (c) Model C: Multispectral remote sensing variables.

environmental covariates, they were divided into three groups that fit different models in our study: (i) imagery remote sensing data set, (ii) terrain-based covariate data set, and (iii) climatic-based covariates.

The relationships between SOC content and each predictive covariate is of great importance to know information about the contribution of each covariate in the model. Pearson`s correlation analysis between

Table 5. Descriptive statistics of predicted soil organic carbon (Mg C ha^{-1}) models.

Model	Mean	Median	Min	Max	SD	SE	Lower	Upper	Skewness	Kurtosis
Model A: Multispectral remote sensing variables + topographic variables + climatic variables	110.35	107.74	46.46	191.33	36.10	2.21	105.78	114.92	0.31	−0.85
Model B: Topographic and climatic variables	110.87	108.86	49.54	200.33	33.53	2.05	106.47	115.27	0.30	−0.59
Model C: Multispectral remote sensing variables	110.45	108.04	51.76	198.28	35.96	2.20	106.23	114.67	0.36	−0.73

Min: minimum; Max: maximum; SD: standard deviation; SE: standard error; Lower and Upper: the lower and upper limits of the mean at 95% probability.

SOC and predictive covariates was derived as shown in Figure 5. SOC stock was positively correlated with temperature, TWI, $B_{blue,}$ B_{red} and B_{green} band, but negatively correlated with precipitation, relative humidity, elevation, and slope. Interestingly, the correlations with the images were the most significant. Finally, we found that there was multicollinearity between the vegetation indexes derived from remote sensing and SOC.

Soil organic carbon models

SOC content in forest-covered soils of the Dominican Republic were estimated using three different models that grouped a series of geospatial datasets. Table 5 shows the descriptive statistics of soil organic carbon (values in Mg C ha^{-1}) for each model. Model A recorded the highest mean value of SOC (110.35 Mg C ha^{-1}). The analysis of variance applied showed no significant differences ($p < 0.05$) between the three models evaluated. The use of climatic and topographic covariates helped improve the model, but not significantly. Using only multispectral imaging can produce good results in digital mapping of SOC at a depth of 0–15 cm.

Spatial model performance

Comparisons of the performance for the three dataset models by cross-validation is shown in Table 6, Figures 6 and 7. We found that *Model A* is the model with the best performance, explaining 83% of the spatial variation of SOC. In general, the more predictors, the better the model. This model groups 20 predictive variables in a dataset: multispectral remote sensing variables + topographic variables + climatic variables. Model B (topographic and climatic variables) yielded an R^2 = 0.77 and an RMSE = 38.57 Mg C ha^{-1}. Interestingly, Model C (only multispectral remote sensing-derived variables) yielded an R^2 = 0.79 and an RMSE = 35.69 Mg C ha^{-1}. These results are consistent with the significant correlations between covariates and SOC (see, Figure 7).

Table 6. Comparison and evaluation of predicted model performance by cross-validation.

Model	Mg C ha^{-1}	R^2	RMSE (Mg C ha^{-1})	MAD	MAPE	LCCC
Model A: Multispectral remote sensing variables + topographic variables + climatic variables	110.35	0.83	35.02	27.90	40.18	0.78
Model B: Topographic and climatic variables	110.87	0.77	38.57	30.00	44.13	0.72
Model C: Multispectral remote sensing variables	110.45	0.79	35.69	28.76	41.00	0.76

R^2: coefficient of determination; RMSE: root mean square error; MAD: mean absolute deviation; MAPE: mean absolute percentage error and, LCCC: Lin's concordance correlation coefficient.

We iterated the model A, B, and C and calculated the average standard deviation (SDs) to analyze the uncertainty of each model in predicting topsoil SOC (Figure 6). We found the highest SD in Model C for all forest types and Model A with the lower uncertainty compared to models B and C.

Covariates relative importance

The average relative importance of each covariate derived from the geospatial dataset to estimate SOC was calculated. To facilitate the analysis for each model, we combined the relative importance of all environmental covariates to 100% (Figure 8). For the dataset grouped in Model A, the 3 most important covariates were slope, temperature and NDVI (34% of the total relative importance). The vegetation indices were ranked at different levels. In Model B, which groups climatic and topographic covariates, the covariates of elevation and precipitation recorded the highest relative importance (48% of the total relative importance). For model C, which only groups covariates derived from Landsat 8 satellite images, the Index-Based built-up Index (IBI) was the covariate with the highest relative importance in the model. For the IBI index, three thematic indices were used: the Modified Normalized Difference Water Index (MNDWI), the Soil Adjusted Vegetation Index (SAVI), and the Normalized Difference Built-up Index (NDBI; Xu, 2008). Table 2 provides further details of the indices used and Figure 9 provides further details of the indices per each type of forest.

SOC stock spatial distribution

A spatially explicit SOC map was created using the GEE cloud computing platform. The results obtained with the three spatial distribution models to predict SOC content in forest-covered areas of the Dominican Republic are shown in Figure 10. Non-forest areas were excluded from the analysis and are shown without color on the map.

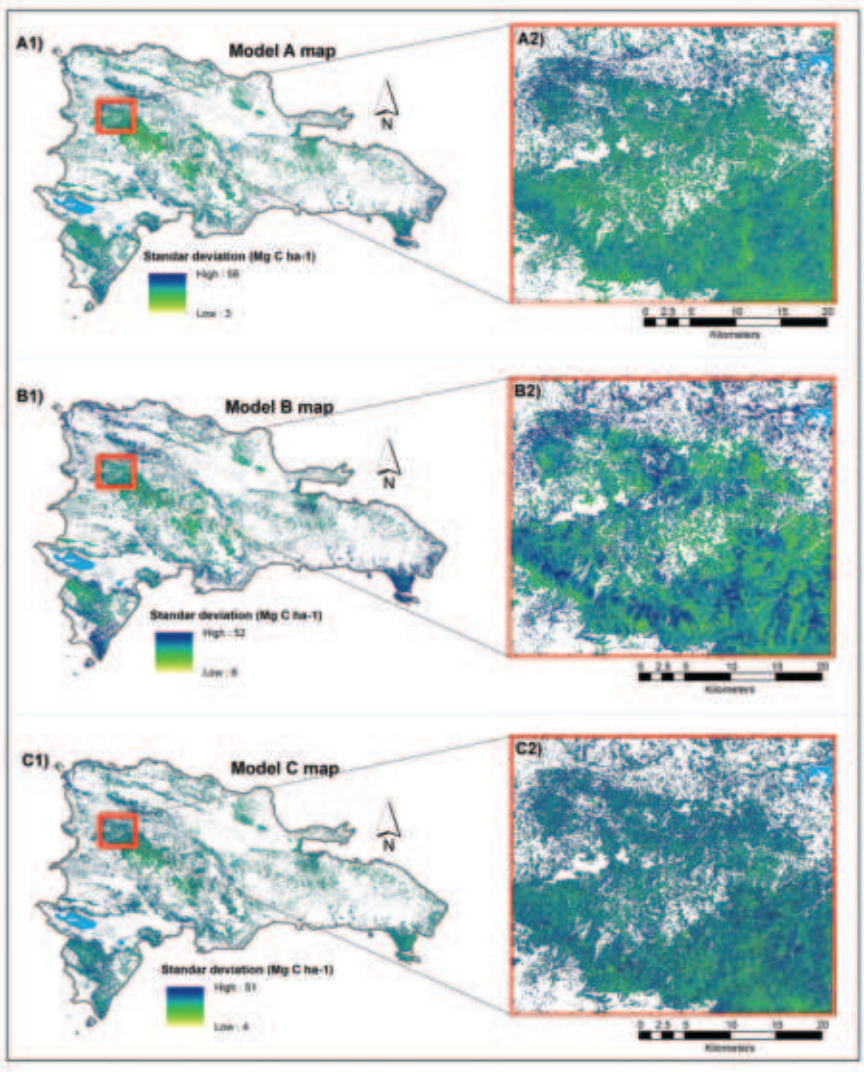


Figure 7. Standard deviation (SD) of SOC stock at 0–15 cm predicted from the random forest (RF) model.

Analyzing the best model obtained in our study (Model A), identified that the spatial patterns of SOC are closely related to the type of forest. The highest SOC (0–15 cm depth) contents are found in the mangrove forests that are located in the coastal areas of the country, with average estimates of 131.87 Mg C ha^{-1}, and maximum and minimum values of 193.09 Mg C ha^{-1} and 63.91 Mg C ha^{-1}, respectively (Table 7). The lowest SOC content is found in the soils covered by pine forests, especially those located in the elevated and steep slopes, in the central and southern region of the country. These soils have a SOC mean value of 89.06 Mg C ha^{-1} and a minimum value of 44.76 Mg C ha^{-1}. Most of these soils are dominated by degraded forests with low productivity and dry shrub vegetation. Their low SOC content is attributed to steeper slopes, which make soils more susceptible to erosion and greater water discharge.

We found that in the soils covered by forests in the Dominican Republic, a total of 144,051,831 Mg C is stored in the topsoil (0–15 cm depth), with 52.1% corresponding to soils covered by broadleaf forests, 31.2% covered by dry forests, 14.3% covered by pine forests, and 2.4% covered by mangrove forests. Table 7 shows more details of the results of SOC obtained in the three models for forest type.

Discussion

Method for measuring and monitoring SOC stocks: a contribution to regional and global initiatives

SOC stocks have acquired great relevance due to the role they play in climate regulation and as an important indicator of soil quality. International organizations such as The United Nations Framework Convention on Climate Change (UNFCCC), the United Nations Convention to Combat Desertification (UNCCD) and the Convention on Biodiversity (CBD) have widely recognized the importance of SOC in the international framework of climate change mitigation. In this sense, there have been emerging regional initiatives aiming at the sustained production of soil information, such as the

Figure 8. Relative importance in the Random Forest (RF) models trained for different geospatial datasets.

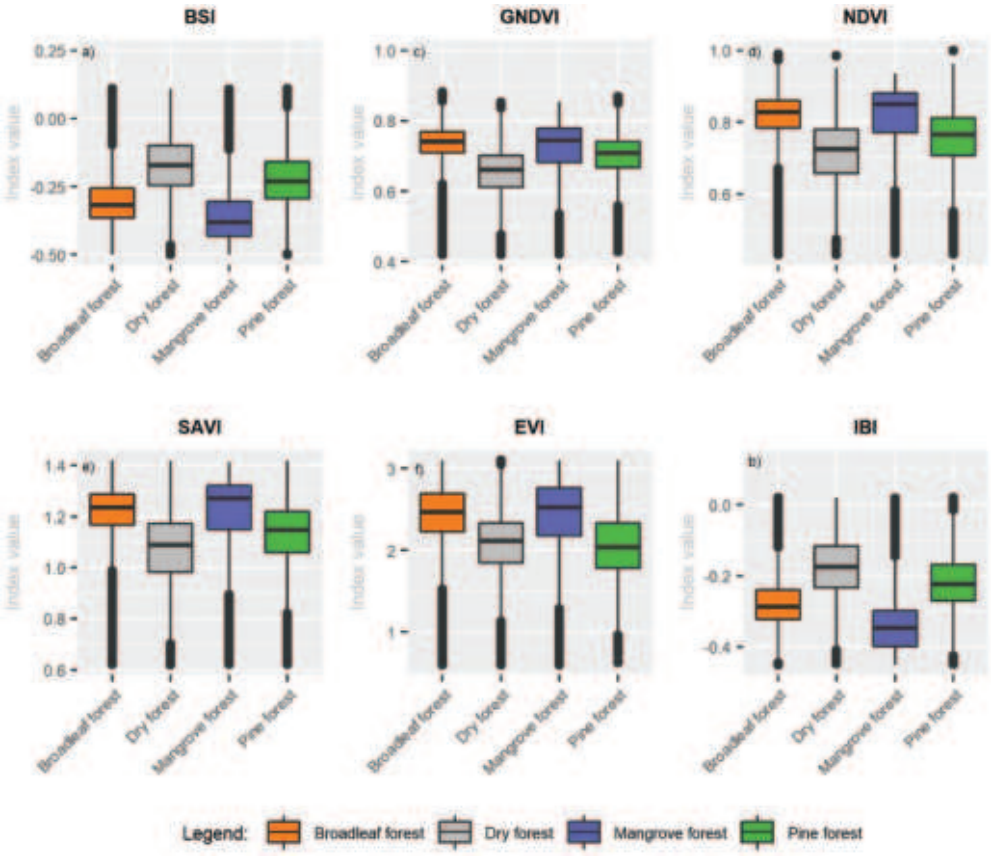


Figure 9. Distribution of the spectral indices for each type of forest derived from Landsat 8.

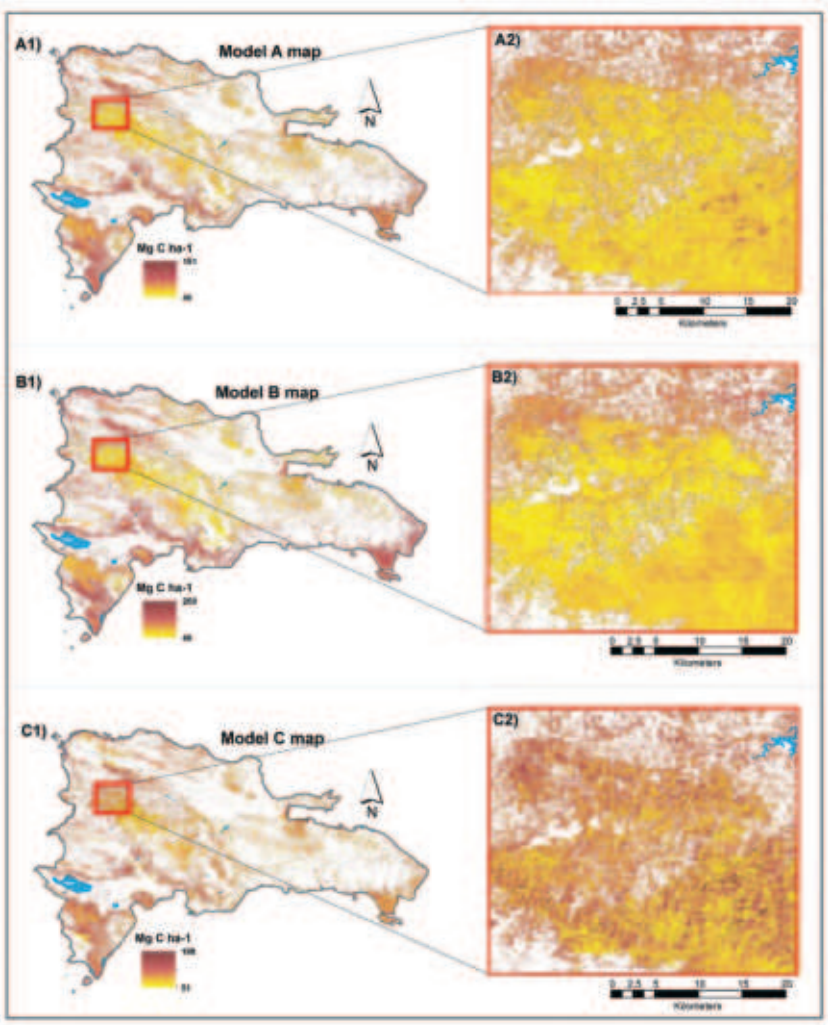


Figure 10. Distribution maps of soil organic carbon (Mg C ha^{-1}) derived with the Random Forest Algorithm. Maps are shown with a masking layer of non-forest land. A1) Model A included all predictive covariates (Multispectral remote sensing variables + topographic variables + climatic variables); A2) zoomed-in image to SOC map derived from Model A. B1). Model B included topographic and climatic variables; B2) zoomed-in image to SOC map derived from Model B. C1) Model C included only multispectral remote sensing variables, C2) zoomed-in image to SOC map derived from Model C.

Soil Information System for Latin America and the Caribbean (SISLAC), or global initiatives such as ISRIC – World Soil Information, legally registered as the International Soil Reference and Information Center, which has the mission to serve the international community as custodian of global soil information. Similarly, the Group of Earth Observation (GEO) has established a Global Soil Information System (GLOSIS) as part of the Global Earth Observation System of Systems (GEOSS). All of these initiatives have encouraged countries to establish national systems for monitoring and measuring SOC. Therefore, there is a need to develop accurate, replicable and low-cost methods to quantify and monitor SOC stock changes.

In the present study, we performed a digital mapping of SOC stocks (in the area under study), using a combination of freely accessible geospatial datasets provided in the GEE cloud computing platform, and field data using the RF algorithm to predict the distribution of SOC in forest soils of the Dominican Republic. To our knowledge, this is the first attempt to map the SOC stocks using this type of technique, in the country and the tropical region of Central America and the Caribbean. The results obtained are encouraging because the three models used had a good performance, even with variables derived only from Landsat 8 OLI images (Model C).

Compared with other methods used to map SOC stocks in tropical regions (Guevara et al., 2018; Ramesh et al., 2015; Rossi et al., 2009; Vasques et al., 2016), our method eliminated excessive soil sampling, which can result in a high cost, particularly in those territories covered by forests where access is very difficult.

Importance of variables in the SOC prediction model

The CLORPT model (CL: Climate; O: Organism, vegetation; R: Relief; P: Parent material; and T: time) (Jenny, 1941) and the SCORPAN model (S: property or soil

Table 7. Descriptive statistics of soil organic carbon (Mg C ha^{-1}) by forest type using different models.

Model A: Multispectral remote sensing variables + topographic variables + climatic variables						
Forest type	Mean	Max	Min	SD	Total (Mg C)	% of SOC
Mangrove forest	131.87	193.09	63.95	25.88	3,451,749	2.4%
Dry forest	120.77	185.90	54.32	19.36	44,909,164	31.2%
Broadleaf forest	100.13	180.39	45.00	22.86	75,065,227	52.1%
Pine forest	89.06	167.88	44.76	19.29	20,625,691	14.3%
Total					*144,051,831*	*100.0%*
Model B: Topographic variables + climatic variables						
Forest type	**Mean**	**Max**	**Min**	**SD**	**Total (Mg C)**	**% of SOC**
Mangrove forest	129.42	201.51	64.75	18.04	3,387,413	2.3%
Dry forest	126.95	193.30	67.07	15.28	47,208,022	32.6%
Broadleaf forest	99.98	191.38	45.61	23.86	74,947,851	51.7%
Pine forest	83.73	171.13	45.63	14.62	19,392,685	13.4%
Total					*144,935,971*	*100.0%*
Model C: Multispectral remote sensing variables						
Forest type	**Mean**	**Max**	**Min**	**SD**	**Total (Mg C)**	**% of SOC**
Mangrove forest	130.86	199.48	48.58	33.62	3,277,858	2.4%
Dry forest	111.82	199.54	44.72	24.68	39,552,568	29.3%
Broadleaf forest	102.26	196.56	41.58	24.84	70,913,089	52.6%
Pine forest	94.72	189.15	41.99	22.87	21,150,587	15.7%
Total					*134,894,102*	*100.0%*

Min: minimum; Max: maximum; SD: standard deviation.

class; C: climate; O: organisms; R: topography; P: parent material; A: age or time factor; and N: space, spatial position) (McBratney et al., 2003) are conceptual models commonly used for digital soil mapping since they relate environmental covariates to soil properties. However, these relationships between covariates and properties differ depending on the geographical area of the soil being analyzed. A review conducted by (McBratney et al., 2003) indicated that the key environmental covariates to infer soil properties were relief (80% of studies), followed by soil class (S) (35%), vegetation (O) and parent material (P) (both 25%), spatial position (N) (20%) and climate (C) (5%).

Several studies have indicated that topographic factors such as elevation and slope have a higher correlation with SOC changes (J.L. Boettinger et al., 2010; Hinge et al., 2018). The NDVI allowed us to understand the importance of the amount of biomass and vegetation cover to predict SOC stocks. Other studies have also reported that SOC can be estimated only by the presence of vegetation (Yang et al., 2008; Zhao & Shi, 2010). Therefore, the NDVI can be used as an approximation to determine SOC. The BSI allowed us to understand the importance of analyzing bare soil, especially in highly fragmented secondary forests such as the forests of the Dominican Republic and many tropical forests. This type of forest structure indicated that vegetation and bare soil combine to generate a forest with a low canopy density (mainly fragmented forests due to human intervention). Our results show that the IBI can significantly enhance the model to predict SOC in fragmented forests with a low canopy density effectively suppressing background noise caused for bare soil. Xu (2008) found that the IBI possesses a positive correlation with land surface temperature and negative correlations with the NDVI.

In terms of the climatic covariates used in our study, temperature was the second most important variable, which explains the SOC changes in Model A. In this sense, previous research has indicated that temperature is a direct predictor of SOC since it has a major influence on determining the type of vegetation, its growth and the microbial decomposition of organic matter (M. Wang et al., 2014).

Comparative analysis of other SOC measurements and mapping initiatives in the region

There have been different local initiatives for the measurement of SOC in the Central American Region in recent years. These have focused on collecting soil C data as part of a multipurpose methodology of local forest inventories, including the 5 pools of carbon defined by the IPCC. However, a wall-to-wall mapping of SOC has not been generated yet. Our study is the first report in which soil C data is used in combination with ML techniques and open-access dataset and available in the GEE cloud computing platform for geographically explicit mapping of the SOC.

ML techniques are widely used in digital SOC mapping as they combine complex and non-linear relationships between different soil attributes and predictive environmental covariates (Drake et al., 2006). Although various prediction algorithms have different capabilities, size of the training sample affects more than the selection of models to improve the prediction accuracy of SOC (Somarathna et al., 2017).

By using the RF algorithm, Model A yielded a SOC mean value (0 – 15 cm) of 110.35 Mg C ha^{-1} (Table 5). This value is higher than the mean value of 81.04 Mg C ha^{-1} reported in the Global Soil Organic Carbon Map (GSOCmap V1.5) prepared by the Food and Agricultural Organization (FAO). The global soil carbon map consists of national SOC maps, developed as 1 km soil grids, at a depth of 0–30 cm (FAO, 2020a); this is the main and most recent SOC mapping initiative existing in the region with which we can compare the results obtained herein. Another comparison with actual SOC mapping, shows that the SOC mean value obtained in the present study is lower than that of 128.80 Mg C ha^{-1} reported in SoilGrids – global gridded soil information; this is a system for digital soil mapping based on a global compilation of soil

profile data (WoSIS) and environmental layers, which uses state-of-the-art ML methods to map SOC contents at a depth of 0–15 cm and 250 m resolution (Hengl et al., 2017). We believe that, due to the source of the data, depth of the soil sample, spatial resolution, scale of the map and the technique used to predict SOC contents, our results are slightly different from the results reported by these two global initiatives for the Dominican Republic. Figure 11 shows a comparative analysis of the three maps described.

Other reports on SOC in Central America are found in El Salvador and Costa Rica, where soil C was measured at 0–20 cm and 0–30 cm depths, respectively. Both measurements were part of the report of the National Multipurpose Forest Inventory, which was built to quantify the SOC stocks of those countries in order to report Forest Reference Emission Levels (FREL/REDD+) to the UNFCCC. However, geographically explicit maps of SOC stocks were not developed. In El Salvador, the mean SOC value in soils covered by forests was 137.45 Mg C ha^{-1} at 0–20 cm depth (García & MARN, 2018), and Costa Rica, the mean value was 108.81 Mg C ha^{-1} at 0–30 cm depth (Emanuelli et al., 2015); both reports are close to those obtained in our study (110.35 Mg C ha^{-1}). As mentioned above, these are the most recent reports for the region, and they provide evidence of the potential that our methodology has as it can be replicated in other countries in the future, and thus contribute to SOC mapping at the regional and global levels.

Conclusions

The present study developed and applied a methodology for SOC mapping in forest lands, using geospatial datasets available in the GEE platform. This approach opens new possibilities for applying ML techniques that will allow countries to develop robust, transparent, consistent, and replicable systems for measuring and monitoring C in soils. In our study,

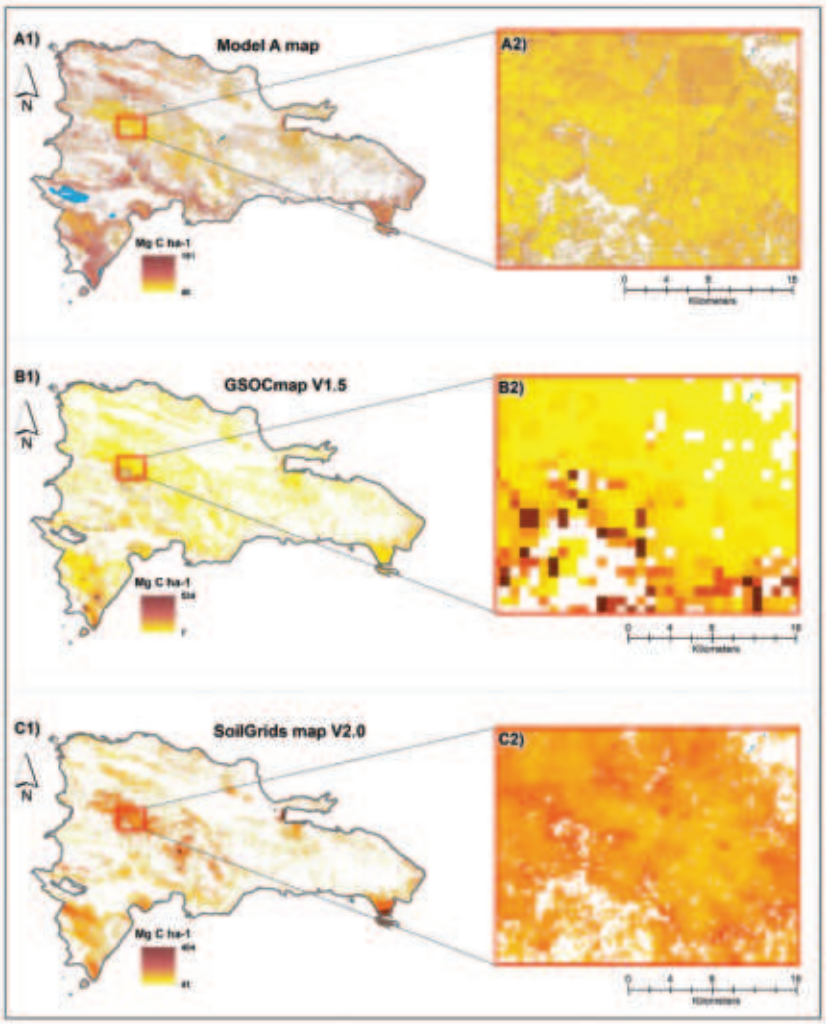


Figure 11. Comparative map of soil organic carbon (Mg C ha^{-1}) obtained with Model A versus Global Soil Organic Carbon Map (GSOCmap V1.5) and SoilGrid map 2.0. A1). Model A included all predictive covariates (Multispectral remote sensing variables + topographic variables + climatic variables); A2) zoomed-in image to SOC map derived from Model A. B1). Global Soil Organic Carbon Map (GSOCmap V1.5); B2) zoomed-in image to GSOCmap V1.5. C1). SoilGrids map V2.0); C2) zoomed-in image to SoilGrids map V2.0.

we found a better coefficient of determination in Model A (topographic, climatic, and Landsat 8 OLI imagery datasets), and determined that the SOC content is mostly related to slope, temperature, and NDVI.

This study shows that freely accessible multispectral optical imagery available in the GEE platform, such as Landsat 8 OLI, can by itself estimate SOC with adequate accuracy in the tropical forests of the Dominican Republic. Adding climatic and topographic covariates improved the model, but not significantly. The results obtained allow indicating that multispectral images are a good tool for SOC digital mapping at 0–15 cm depth.

ML helps simplify model adjustments. This is a great advantage because ML allows mapping SOC content using many predictive variables of climatic, topographic, or vegetation type with minimum human interaction; however, using soil samples distributed and stratified by forest type was crucial to improve model prediction of SOC content with an R^2 of 83%

We found that the GEE platform has excellent potential in "wall-to-wall" SOC mapping in forest lands. Further research is required on the use of these tools for SOC mapping in different land uses (e.g., agricultural and livestock soils), different ecosystems beyond the tropics, and at further depths. We can conclude that the methodology developed may encourage new research that favors the fulfillment of the Pillar 4 Implementation Plan towards a Global Soil Information System within the framework of the Global Soil Alliance, and especially the support of Indicator 15.3.1 of the Objectives of Sustainable Development.

Future research is needed to evaluate; i) new spatial datasets available such as SoilGrid database, Sentinel-1 Synthetic Aperture Radar (SAR) data, and other satellite images like MODIS, Sentinel – 2 MultiSpectral Instrument (MSI), or Planet. ii) other algorithms such as deep learning (neural network), iii) perform hyperparameter estimation and optimization of the RF model to ensure maximum model accuracy, iv) increase the number of SOC stock samples to improve model performance, and v) assessing the land use and land cover change effect on soil organic carbon from landscape to national scale are some outlooks for future studies.

Acknowledgments

We are sincerely grateful to the National Forest Monitoring Unit of the Ministry of Environment and Natural Resources (MARN) of the Dominican Republic, Sud Austral Consulting and the REDD/CCAD-GIZ program for the information and technical support provided. A special thanks to the Doctoral Program in Agronomic Sciences of the University of Concepción for their support of this research.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was funded by the National Agency for Research and Development of Chile (ANID) www.anid.cl

Credit authorship contribution statement

Efraín Duarte: Conceptualization, Methodology, Software, Formal analysis, Writing - original draft, Data curation. Erick Zagal: Supervision, Reviewing and Editing. Juan A. Barrera: Conceptualization. Francis Dube: Reviewing and Editing. Fabio Casco: Conceptualization, Methodology, Software. Alexander J Hernández: Conceptualization.

References

Alba, E., Mello, E. P., Marchesan, J., Silva, E. A., Tramontina, J., & Pereira, R. S. (2017). Spectral characterization of forest plantations with Landsat 8/OLI images for forest planning and management. *Pesquisa AgropecuÃ!`ria Brasileira*, *52*(11), 1072–1079. https://doi.org/10.1590/s0100-204x2017001100013

Baccini, A., Goetz, S. J., Walker, W. S., Laporte, N. T., Sun, M., Sulla-Menashe, D., Hackler, J., Beck, P. S. A., Dubayah, R., Friedl, M. A., Samanta, S., & Houghton, R. A. (2012). Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. *Nature Climate Change*, *2*(3), 182–185. https://doi.org/10.1038/nclimate1354

Badgley, G., Field, C. B., & Berry, J. A. (2017). Canopy near-infrared reflectance and terrestrial photosynthesis. *Science Advances*, *3*(3), e1602244. https://doi.org/10.1126/sciadv.1602244

Boettinger, J. L., Howell, D. W., Moore, A. C., Hartemink, A. E., & Kienast-Brown, S. (2010). *Digital soil mapping*. Springer Netherlands. https://doi.org/10.1007/978-90-481-8863-5

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Cano-Ortiz, A., Musarella, C. M., Piñar, J. C., Spampinato, G., Veloz, A., & Cano, E. (2015). Vegetation of the dry bioclimatic areas in the Dominican Republic. *Plant Biosystems - an International Journal Dealing with All Aspects of Plant Biology*, *149*(3), 451–472. https://doi.org/10.1080/11263504.2015.1040482

Dane Topp, G. C., Campbell Gaylon, S., & H, J. (2002). *Methods of soil analysis. Part 4, Part 4*. Soil Science Society of America.

Drake, J., Randin, C., & Guisan, A. (2006). Modelling ecological niches with support vector machines. *Journal of Applied Ecology*, *43*(3), 424–432. https://doi.org/10.1111/j.1365-2664.2006.01141.x

Duarte, E., Barrera, J. A., Dube, F., Casco, F., Hernández, A. J., & Zagal, E. (2020). Monitoring approach for tropical coniferous forest degradation using remote sensing and field data. *Remote Sensing*, *12*(16), 2531. https://doi.org/10.3390/rs12162531

Edenhofer, O., Pichs-Madruga, R., Sokona, Y., Seyboth, K., Kadner, S., Zwickel, T., Eickemeier, P., Hansen, G., & Schlömer, S. (2014). Climate change 2014: Mitigation of

climate change. In F. E. Edenhofer O, R. Pichs Madruga, Y. Sokona (Ed.), *Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press 162 ISBN 978-92-9169-142-5 Accessed 05 may 2021 https://www.ipcc.ch/site/assets/uploads/2018/03/WGIIIAR5_SPM_TS_Volume-3.pdf .

Emanuelli, P., Milla, F., Duarte, E., Emanuelli, J., Jiménez, A., & Chavarría, M. (2015). *Inventario Nacional Forestal de Costa Rica 2014-2015* (Costa Rica: REDD/CCAD-GIZ).

Ermida, S. L., Soares, P., Mantas, V., Göttsche, F.-M., & Trigo, I. F. (2020). Google Earth Engine open-source code for land surface temperature estimation from the landsat series. *Remote Sensing*, *12*(9), 1471. https://doi.org/10.3390/rs12091471

FAO. (2020). *A protocol for measurement, monitoring, reporting and verification of soil organic carbon in agricultural landscapes – GSOC-MRV Protocol* (FAO). https://doi.org/10.4060/cb0509en

FAO. (2020b). *Forest resources assessment, dominican republic report*. http://www.fao.org/3/cb0101es/cb0101es.pdf

FAO, & ITPS. (2020). *Global Soil Organic Carbon Map (GSOCmap) Version 1.5*. https://doi.org/10.4060/ca7597en

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., & Alsdorf, D. (2007). The shuttle radar topography mission. *Reviews of Geophysics*, *45*(2 Article Number 2005RG000183). https://doi.org/10.1029/2005RG000183

Feliz, K., Rodríguez, L., Galán, M., Ovidio, R., Vargas, O., & de Jong, B. (2019). *Dominican republic reference emissions levels/forest reference levels*. MARN. https://redd.unfccc.int/files/nrfe_-_nrf_rep._dom_revgov2.pdf

Flood, N. (2013). Seasonal composite landsat TM/ETM+ images using the medoid (a Multi-Dimensional Median). *Remote Sensing*, *5*(12), 6481–6500. https://doi.org/10.3390/rs5126481

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., & Michaelsen, J. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data*, *2*(1), 150066. https://doi.org/10.1038/sdata.2015.66

MARN, & García, C. (2018). *Inventario Nacional de Bosques de El Salvador* (San Salvador, El Salvador: MARN) Accessed 10 March 2021 . https://cidoc.marn.gob.sv/documentos/inventario-nacional-de-bosques

Gianelle, D., Oechel, W., Miglietta, F., Rodeghiero, M., Sottocornola, M., & Sills, J. (2010). Cataloguing soil carbon stocks. *Science (New York, N.Y.)*, *330*(6010), 1476–1477. https://doi.org/10.1126/science.330.6010.1476-c

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202* 15 , 18–27 00344257 . https://doi.org/10.1016/j.rse.2017.06.031

Grinand, C., le, M. G., Vieilledent, G., Razakamanarivo, H., Razafimbelo, T., & Bernoux, M. (2017). Estimating temporal changes in soil carbon stocks at ecoregional scale in Madagascar using remote-sensing. *International Journal of Applied Earth Observation and Geoinformation*, *54* 1, 1–14. https://doi.org/10.1016/j.jag.2016.09.002

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., … Vargas, R. (2018). No silver bullet for digital soil mapping: Country-specific soil organic carbon estimates across Latin America. *SOIL*, *4*(3), 173–193. https://doi.org/10.5194/soil-4-173-2018

Harris, N. L., Brown, S., Hagen, S. C., Saatchi, S. S., Petrova, S., Salas, W., Hansen, M. C., Potapov, P. V., & Lotsch, A. (2012). Baseline map of carbon emissions from deforestation in tropical regions. *Science (New York N.Y.)*, *336* (6088), 1573–1576. https://doi.org/10.1126/science.1217962

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*(2), e0169748. https://doi.org/10.1371/journal.pone.0169748

Hinge, G., Surampalli, R. Y., & Goyal, M. K. (2018). Prediction of soil organic carbon stock using digital mapping approach in humid India. *Environmental Earth Sciences*, *77*(5), 172. https://doi.org/10.1007/s12665-018-7374-x

Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, *25*(3), 295–309. https://doi.org/10.1016/0034-4257(88)90106-X

Huete, A. R., Liu, H. Q., Batchily, K., & van Leeuwen, W. (1997). A comparison of vegetation indices over a global set of TM images for EOS-MODIS. *Remote Sensing of Environment*, *59*(3), 440–451. https://doi.org/10.1016/S0034-4257(96)00112-5

Jan, B., & Jeffrey, U. (2018). Achieving the United Nations sustainable development goals: An enabling role for accounting research. *Accounting, Auditing & Accountability Journal*, *31*(1), 2–24. https://doi.org/10.1108/AAAJ-05-2017-2929

Jenny, H. (1941). *Factors of soil formation : a system of quantitative pedology* (Macgraw Hill). https://www.worldcat.org/title/factors-of-soil-formation-a-system-of-quantitative-pedology/oclc/905466

Kennedy, L. M., Horn, S. P., & Orvis, K. H. (2005). Modern pollen spectra from the highlands of the Cordillera Central, Dominican Republic. *Review of Palaeobotany and Palynology*, *137*(1), 51–68. https://doi.org/10.1016/j.revpalbo.2005.08.007

Köchy, M., Hiederer, R., & Freibauer, A. (2015). Global distribution of soil organic carbon – Part 1: Masses and frequency distributions of SOC stocks for the tropics, permafrost regions, wetlands, and the world. *SOIL*, *1*(1), 351–365. https://doi.org/10.5194/soil-1-351-2015

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* 20 , 395–413 0016-7061 . https://doi.org/10.1016/j.geoderma.2019.05.031

Liang, X., Kankare, V., Hyyppä, J., Wang, Y., Kukko, A., Haggrén, H., Yu, X., Kaartinen, H., Jaakkola, A., Guan, F., Holopainen, M., & Vastaranta, M. (2016). Terrestrial laser scanning in forest inventories. *ISPRS Journal of Photogrammetry and Remote Sensing*, *115* 5 , 63–77. https://doi.org/10.1016/j.isprsjprs.2016.01.006

Lin, L. I.-K. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, *45*(1), 255–268. https://doi.org/10.2307/2532051

Mahmoudzadeh, H., Matinfar, H. R., Taghizadeh-Mehrjardi, R., & Kerry, R. (2020). Spatial prediction of soil organic carbon using machine learning techniques in western Iran. *Geoderma Regional*, *21* 13 , e00260 ISSN : 2352-0094 . https://doi.org/10.1016/j.geodrs.2020.e00260

MARN. (2012). *Atlas of the Biodiversity and Natural Resources of the Dominican Republic*. https://biodiversidad-rd.net/atlas-de-la-biodiversidad-y-de-los-recursos-naturales-de-la-republica-dominicana-2012/

MARN. (2019). *Emission Reductions Program Document (ER-PD)*. https://www.forestcarbonpartnership.org/country/dominican-republic

MARN. (2021). *National Forest Inventory of the Dominican Republic*. https://www.sica.int/documentos/inventario-forestal-nacional-de-republica-dominicana_1_126744.html

Martin, P. H., & Fahey, T. J. (2006). Fire history along environmental gradients in the subtropical pine forests of the Cordillera Central, Dominican Republic. *Journal of Tropical Ecology*, *22*(3), 289–302. https://doi.org/10.1017/S0266467406003178

Masek, J. G., Vermote, E. F., Saleous, N., Wolfe, R., Hall, F. G., Huemmrich, K. F., Gao, F., Kutler, J., & Lim, T. K. (2013). *LEDAPS calibration, reflectance, atmospheric correction preprocessing code, version 2*. ORNL Distributed Active Archive Center. https://doi.org/10.3334/ORNLDAAC/1146

McBratney, A. B., Mendonça Santos, M. L., & Minasny, B. (2003). On digital soil mapping. *Geoderma*, *117*(1), 3–52. https://doi.org/10.1016/S0016-7061(03)00223-4

Padarian, J., Minasny, B., & McBratney, A. B. (2019). Using deep learning for digital soil mapping. *SOIL*, *5*(1), 79–89. https://doi.org/10.5194/soil-5-79-2019

Piyoosh, A. K., & Ghosh, S. K. (2018). Development of a modified bare soil and urban index for Landsat 8 satellite data. *Geocarto International*, *33*(4), 423–442. https://doi.org/10.1080/10106049.2016.1273401

Ramesh, T., Manjaiah, K. M., Mohopatra, K. P., Rajasekar, K., & Ngachan, S. V. (2015). Assessment of soil organic carbon stocks and fractions under different agroforestry systems in subtropical hill agroecosystems of north-east India. *Agroforestry Systems*, *89*(4), 677–690. https://doi.org/10.1007/s10457-015-9804-z

Rossi, J., Govaerts, A., de Vos, B., Verbist, B., Vervoort, A., Poesen, J., Muys, B., & Deckers, J. (2009). Spatial structures of soil organic carbon in tropical forests—A case study of Southeastern Tanzania. *CATENA*, *77*(1), 19–27. https://doi.org/10.1016/j.catena.2008.12.003

Ruesch, A., & Gibbs, H. (2008). *New IPCC Tier-1 Global Biomass Carbon Map for the Year 2000* (Oak Ridge, Tennessee.: Oak Ridge National Laboratory). Center for Sustainability and the Global Environment (SAGE), University of Wisconsin https://doi.org/10.15485/1463800

Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T. A., Salas, W., Zutta, B. R., Buermann, W., Lewis, S. L., Hagen, S., Petrova, S., White, L., Silman, M., & Morel, A. (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences*, *108* (24), 9899 LP – 9904. 24. https://doi.org/10.1073/pnas.1019576108 .

Saha, S., Nadiga, S., Thiaw, C., Wang, J., Wang, W., Zhang, Q., van den Dool, H. M., Pan, H.-L., Moorthi, S., Behringer, D., Stokes, D., Pe?a, M., Lord, S., White, G., Ebisuzaki, W., Peng, P., & Xie, P. (n.d.). The NCEP climate forecast system. *Journal of Climate*, *19*(15), 3483–3517. https://doi.org/10.1175/JCLI3812.1

Scharlemann, J. P. W., Tanner, E. V. J., Hiederer, R., & Kapos, V. (2014). Global soil carbon: Understanding and managing the largest terrestrial carbon pool. *Carbon Management*, *5*(1), 81–91. https://doi.org/10.4155/cmt.13.77

Scholten, T., Goebes, P., Kühn, P., Seitz, S., Assmann, T., Bauhus, J., Bruelheide, H., Buscot, F., Erfmeier, A., Fischer, M., Härdtle, W., He, J.-S., Ma, K., Niklaus, P. A., Scherer-Lorenzen, M., Schmid, B., Shi, X., Song, Z., von Oheimb, G., … Schmidt, K. (2017). On the combined effect of soil fertility and topography on tree growth in subtropical forest ecosystems—a study from SE China. *Journal of Plant Ecology*, *10*(1), 111–127. https://doi.org/10.1093/jpe/rtw065

Sobrino, J. A., & Raissouni, N. (2000). Toward remote sensing methods for land cover dynamic monitoring: Application to Morocco. *International Journal of Remote Sensing*, *21*(2), 353–366. https://doi.org/10.1080/014311600210876

Somarathna, P. D. S. N., Minasny, B., & Malone, B. P. (2017). More data or a better model? Figuring out what matters most for the spatial prediction of soil carbon. *Soil Science Society of America Journal*, *81*(6), 1413–1426. https://doi.org/10.2136/sssaj2016.11.0376

Tan, B., Masek, J. G., Wolfe, R., Gao, F., Huang, C., Vermote, E. F., Sexton, J. O., & Ederer, G. (2013). Improved forest change detection with terrain illumination corrected Landsat images. *Remote Sensing of Environment*, *136* 6 , 469–483. https://doi.org/10.1016/j.rse.2013.05.013

Tifafi, M., Guenet, B., & Hatté, C. (2018). Large differences in global and regional total soil carbon stock estimates based on SoilGrids, HWSD, and NCSCD: Intercomparison and evaluation based on field data from USA, England, Wales, and France. *Global Biogeochemical Cycles*, *32*(1), 42–56. https://doi.org/10.1002/2017GB005678

Trumbore, S. (2009). Radiocarbon and soil carbon dynamics. *Annual Review of Earth and Planetary Sciences*, *37*(1), 47–66. https://doi.org/10.1146/annurev.earth.36.031207.124300

Vasques, G. M., Coelho, M. R., Dart, R. O., Oliveira, R. P., & Teixeira, W. G. (2016). Mapping soil carbon, particle-size fractions, and water retention in tropical dry forest in Brazil. *Pesquisa AgropecuÃ!`ria Brasileira*, *51*(9), 1371–1385. https://doi.org/10.1590/s0100-204x2016000900036

Veronesi, F., & Schillaci, C. (2019). Comparison between geostatistical and machine learning models as predictors of topsoil organic carbon with a focus on local uncertainty estimation. *Ecological Indicators*, *101*(7), 1032–1044. https://doi.org/10.1016/j.ecolind.2019.02.026

Viscarra Rossel, R. A., Brus, D. J., Lobsey, C., Shi, Z., & McLachlan, G. (2016). Baseline estimates of soil organic carbon by proximal sensing: Comparing design-based, model-assisted and model-based inference. *Geoderma*, *265*(5), 152–163. https://doi.org/10.1016/j.geoderma.2015.11.016

Walkley, A., & Black, I. A. (1934). An examination of the degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. *Soil Science*, *37*(1), 29–38. https://doi.org/10.1097/00010694-193401000-00003

Wan, Z. (2014). New refinements and validation of the collection-6 MODIS land-surface temperature/emissivity product. *Remote Sensing of Environment*, *140*(1), 36–45. https://doi.org/10.1016/j.rse.2013.08.027

Wang, M., Su, Y., Yang, X., & Bond-Lamberty, B. (2014). Spatial distribution of soil organic carbon and its influencing factors in desert Grasslands of the Hexi Corridor, Northwest China. *PLOS ONE*, *9*(4), e94652. https://doi.org/10.1371/journal.pone.0094652

Wang, B., Waters, C., Orgill, S., Cowie, A., Clark, A., Li Liu, D., Simpson, M., McGowen, I., & Sides, T. (2018). Estimating soil organic carbon stocks using different modelling techniques in the semi-arid rangelands of eastern Australia. *Ecological Indicators*, *88*(5), 425–438. https://doi.org/10.1016/j.ecolind.2018.01.049

WMO. (2020). *About Essential Climate Variables*. https://gcos.wmo.int/en/essential-climate-variables/about

Wulder, M. A., White, J. C., Loveland, T. R., Woodcock, C. E., Belward, A. S., Cohen, W. B., Fosnight, E. A., Shaw, J., Masek, J. G., & Roy, D. P. (2016). The global landsat archive: Status, consolidation, and direction. *Remote Sensing of Environment*, *185*(14), 271–283. https://doi.org/10.1016/j.rse.2015.11.032

Xiao, J., Chevallier, F., Gomez, C., Guanter, L., Hicke, J. A., Huete, A. R., Ichii, K., Ni, W., Pang, Y., Rahman, A. F., Sun, G., Yuan, W., Zhang, L., & Zhang, X. (2019). Remote sensing of the terrestrial carbon cycle: A review of advances over 50 years. *Remote Sensing of Environment*, *233*(14), 111383. https://doi.org/10.1016/j.rse.2019.111383

Xu, H. (2008). A new index for delineating built-up land features in satellite imagery. *International Journal of Remote Sensing*, *29*(14), 4269–4276. https://doi.org/10.1080/01431160802039957

Yang, Y., Fang, J., Tang, Y., Ji, C., Zheng, C., He, J., & Zhu, B. (2008). Storage, patterns and controls of soil organic carbon in the Tibetan grasslands. *Global Change Biology*, *14*(7), 1592–1599. https://doi.org/10.1111/j.1365-2486.2008.01591.x

Zhang, G., Liu, F., & Song, X. (2017). Recent progress and future prospect of digital soil mapping: A review. *Journal of Integrative Agriculture*, *16*(12), 2871–2885. https://doi.org/10.1016/S2095-3119(17)61762-3

Zhao, Y.-C., & Shi, X.-Z. (2010). *Spatial prediction and uncertainty assessment of soil organic carbon in Hebei Province, China BT - Digital soil mapping: Bridging research, environmental application, and operation*. J.L. Boettinger, D. W. Howell, A.C. Moore, A.E. Hartemink, & S. Kienast-Brown, Eds. 227–239. Springer Netherlands. https://doi.org/10.1007/978-90-481-8863-5_19

Zhu, Z., & Woodcock, C. E. (2012). Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sensing of Environment*, *118*(3), 83–94. https://doi.org/10.1016/j.rse.2011.10.028

Appendix A

The codes developed in this study using the GEE cloud-based computing platform are available at the following link: https://github.com/EDuarteCode/SOC_Code.git