

# Introducing a mechanistic model in digital soil mapping to predict soil organic matter stocks in the Cantabrian region (Spain)

Chantal Mechtildis Johanna Hendriks<sup>1</sup>  | Jetse Jacob Stoorvogel<sup>2</sup>  |  
Jose Manuel Álvarez-Martínez<sup>3</sup> | Lieven Claessens<sup>2,4</sup>  | Ignacio Pérez-Silos<sup>3</sup> |  
José Barquín<sup>3</sup>

<sup>1</sup>Sustainable Soil Use Group, Wageningen University & Research, Wageningen, The Netherlands

<sup>2</sup>Soil Geography and Landscape Group, Wageningen University & Research, Wageningen, The Netherlands

<sup>3</sup>Environmental Hydraulics Institute (IH Cantabria), University of Cantabria, C/ Isabel Torres n° 15, Parque Científico y Tecnológico de Cantabria, Santander, Spain

<sup>4</sup>International Institute of Tropical Agriculture (IITA), Arusha, Tanzania

## Correspondence

Chantal Hendriks, Sustainable Soil Use Group, Wageningen University & Research, PO Box 47, 6700 AA Wageningen, The Netherlands.  
Email: chantal.hendriks@wur.nl

## Funding information

CGIAR Research Programme on Climate Change, Agriculture and Food Security (CCAFS); Environmental Hydraulics Institute 'IH Cantabria of Universidad de Cantabria'; Universidad de Cantabria

## Abstract

Digital soil mapping (DSM) is an effective mapping technique that supports the increased need for quantitative soil data. In DSM, soil properties are correlated with environmental characteristics using statistical models such as regression. However, many of these relationships are explicitly described in mechanistic simulation models. Therefore, the mechanistic relationships can, in theory, replace the statistical relationships in DSM. This study aims to develop a mechanistic model to predict soil organic matter (SOM) stocks in Natura2000 areas of the Cantabria region (Spain). The mechanistic model is established in four steps: (a) identify major processes that influence SOM stocks, (b) review existing models describing the major processes and the respective environmental data that they require, (c) establish a database with the required input data, and (d) calibrate the model with field observations. The SOM stocks map resulting from the mechanistic model had a mean error (ME) of  $-2 \text{ t SOM ha}^{-1}$  and a root mean square error (RMSE) of  $66 \text{ t SOM ha}^{-1}$ . The Lin's concordance correlation coefficient was 0.47 and the amount of variance explained (AVE) was 0.21. The results of the mechanistic model were compared to the results of a statistical model. It turned out that the correlation coefficient between the two SOM stock maps was 0.8. This study illustrated that mechanistic soil models can be used for DSM, which brings new opportunities. Mechanistic models for DSM should be considered for mapping soil characteristics that are difficult to predict by statistical models, and for extrapolation purposes.

## Highlights

- Theoretically, mechanistic models can replace the statistical relationships in digital soil mapping.

- Mechanistic soil models were used to develop a mechanistic model for digital soil mapping that predicted SOM stocks.
- The applicability of the mechanistic approach needs to be explored for different soil properties and regions.

**KEYWORDS**

Cordillera cantábrica, Natura2000, organic carbon, soil-forming processes, sustainable development

## 1 | INTRODUCTION

Digital soil mapping (DSM) is a widely adopted and recognized technique to obtain quantitative, spatially exhaustive soil data (Minasny & McBratney, 2016). The technique predicts soil properties spatially by finding a relationship between observed soil properties and selected explanatory variables that are considered to represent the soil-forming factors defined by Jenny (1941): climate, organisms, relief, parent material and time. Often, the relationships between observed soil properties and explanatory variables are described by statistical models. If the residuals (i.e., observed minus predicted values) of these statistical models show spatial autocorrelation, the estimates can be improved by correcting for the interpolated residuals. The methodology for DSM was originally based on regression kriging (Hengl, Heuvelink, & Rossiter, 2007), but is increasingly replaced by linear mixed models (Lark & Cullis, 2004; Lark, Cullis, & Welham, 2006). DSM is seen as an efficient tool for predicting the spatial distribution of soil properties (Ma, Minasny, & Wu, 2017).

Many different statistical models are used for DSM, including multiple regression, regression tree, neural network and fuzzy models (McBratney, Mendonça Santos, & Minasny, 2003). Although statistical models have proven their predictive power, there are certain soil properties that are difficult to predict at a regional scale (e.g., Dorji, Odeh, Field, & Baillie, 2014; Kempen, Brus, & Stoorvogel, 2011; Zhao et al., 2014). They often only incorporate pedological knowledge in the selection of environmental variables. Various authors have explored alternative methods that incorporate pedological knowledge into the conceptual model. For example, Minasny and McBratney (2006) incorporated the within-profile transport of nutrients by using the same relationship as two mechanistic models (Elzein & Balesdent, 1995; Rosenbloom, Harden, Neff, & Schimel, 2006), and Angelini, Heuvelink, Kempen, and Morrás (2016) explored the use of structural equation modelling. Although these approaches incorporated pedological knowledge, they still relied on statistical algorithms for the spatial prediction. Given the many mechanistic soil

models that are available, we want to know whether the relationships of these models can be used for DSM. This study aims to explore whether mechanistic models can be used for DSM of soil organic matter (SOM) stocks.

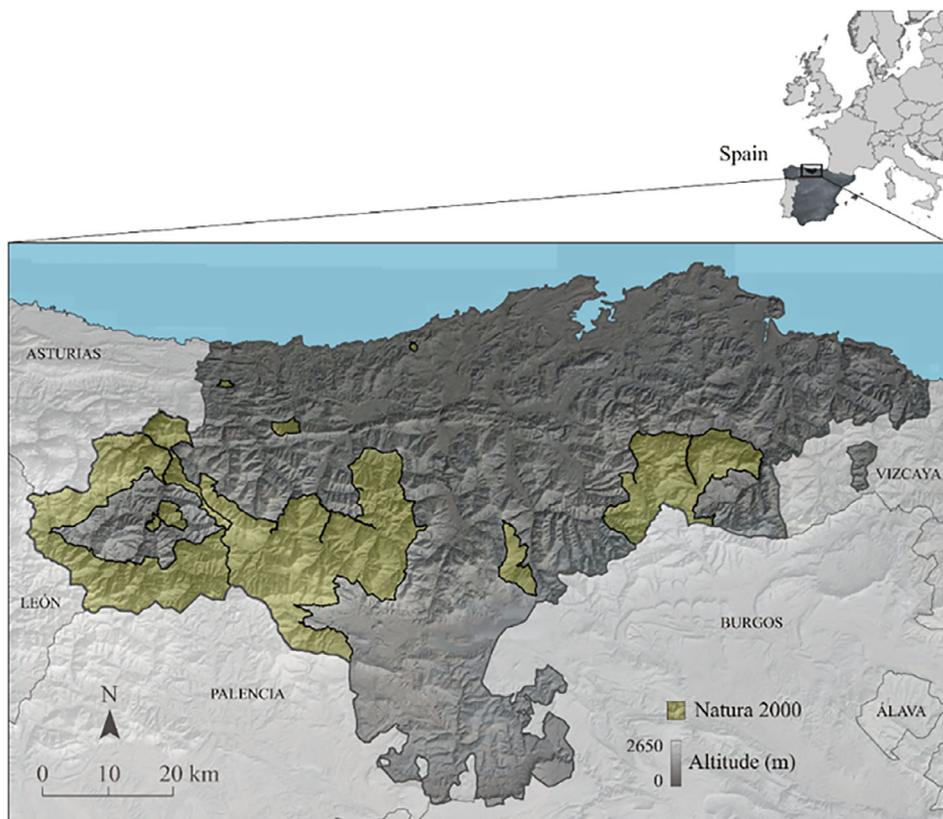
This study focused on the prediction of SOM stocks for three reasons. First, SOM contributes to different soil functions, such as carbon sequestration, nutrient availability and water-holding capacity (Bot & Benites, 2005). These functions are essential in the structure and functioning of ecosystems, which makes SOM an important soil property for a wide range of studies dealing with, for example, climate change, agriculture and biodiversity. Second, processes that influence SOM stocks are complex and non-linear, which can result in poor prediction using, for example, linear regression models. This especially occurs when a limited number of soil observations or explanatory variables is available, or when the explanatory variables are not available at the required scale. Third, processes that influence SOM stocks have been exhaustively studied by many mechanistic soil models (Campbell & Paustian, 2015; Shibu, Leffelaar, Van Keulen, & Aggarwal, 2006).

Member states of the European Union need to monitor the state of Natura2000 areas. Natura2000 is a network of areas across 28 countries of the European Union, where the long-term survival of Europe's most valuable and threatened species and habitats needs to be ensured (Sundseth, Wegefelt, & Mézard, 2008). The network was established in 1992. It is the largest coordinated network of protected areas in the world. Maps of the SOM stocks help monitor land recovery and can be used in vegetation models. This study developed a mechanistic model that spatially predicts SOM stocks for the Natura2000 areas of the Cantabria region (Spain). It is hypothesized that this approach has potential for extrapolation and can easily be updated.

## 2 | MATERIALS AND METHODS

### 2.1 | Study area

The Natura2000 areas of the Cantabrian region (43°20'N, 4°00'W) (Figure 1) have an Atlantic climate along the



**FIGURE 1** The location of the study area in the Natura2000 areas of the Cantabria region (Spain) [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

coast and an Alpine climate in the mountainous area. Annual average temperatures range from 15°C at 23 m to 2°C at 2,531 m above sea level. Long-term (1950–1999) annual precipitation ranges between 526 and 2,155 mm (Ninyerola, Pons, & Roure, 2007). Montane glaciation, periglacial phenomena, alluvial terraces and marine dynamics formed the landscape. This resulted in a hilly to mountainous landscape, with steep slopes and active erosion. The steep slopes and the frequency of dry winds encourage wildfires in autumn and winter, mainly caused by local farmers practising spring grazing. Geomorphological processes also resulted in a rich variation in lithology, including shales, sandstone, limestone, conglomerates and slates. The most dominant soil types are the mollic, haplic, gleyic Solonetz, the albic and haplic Luvisol and the orthic Podzol (IUSS Working Group WRB, 2015). The eutric Cambisol dominates on sloping land (Carballas et al., 2016). The rich lithology and the environmental heterogeneity of the area have resulted in unique ecosystems. The area harbours a mix of temperate deciduous and sclerophyllous vegetation species, including beeches (*Fagus sylvatica*), oaks (*Quercus petraea*, *Q. robur*) and birches (*Betula spp*) in colder, wetter areas, and other oak species (*Q. pyrenaica* and *Q. rotundifolia*) in warmer and dryer areas. Most of the mature forest was deforested during the 20th century for timber production and agriculture. However, nature recovered quickly after

the European Union Directive designated the area a Special Protection Area in 1992. After many years of this protected status, it is assumed that the nature in the area has reached a state of equilibrium. Secondary forests, shrub patches and abandoned and extensive pastures (Álvarez-Martínez, Suárez-Seoane, Stoorvogel, & de Luis Calabuig, 2014) now characterize the area. Brambles (*Rubus spp.*), roses (*Rosa spp.*), hawthorn (*Crataegus monogyna*) and blackthorn (*Prunus spinosa*) dominate the abandoned pastures (Álvarez-Martínez et al., 2018). In flatter alluvial terraces, agriculture still takes place. At high altitudes, well-managed grasslands are used for extensive grazing. It is likely that the variation in climate, geomorphology, lithology and land use results in considerable variation in SOM stocks.

## 2.2 | Data collection

### 2.2.1 | Environmental variables

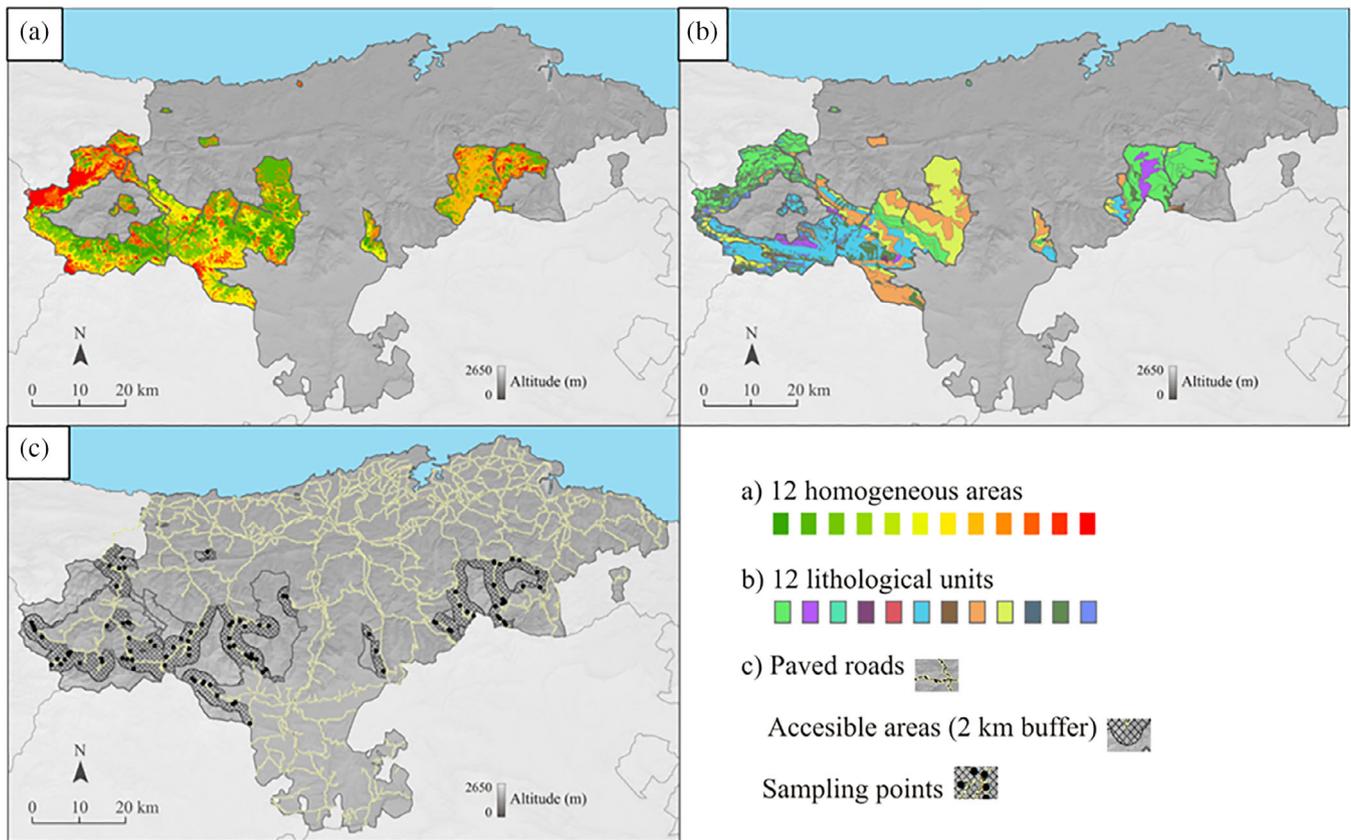
To develop the sampling scheme and to select input data for the models, high-resolution and spatially exhaustive environmental variables were used. The variables were categorized according to the five soil-forming factors (Table 1). Climatic data were obtained from the Iberian Peninsula dataset of Ninyerola et al. (2007). Land-use and land-cover

**TABLE 1** Environmental covariates available for the Cantabria region (Spain)

Soil-forming factor	Description	Variable	Unit	Code	Source
Climate	Mean and coefficient of variation 1981–2010	Precipitation	mm	P	Spanish Climatic Map (Ninyerola et al., 2007)
		Max. temperature	°C	Tmax	Spanish Climatic Map (Ninyerola et al., 2007)
		Mean temperature	°C	Tmean	Spanish Climatic Map (Ninyerola et al., 2007)
		Min. temperature	°C	Tmin	Spanish Climatic Map (Ninyerola et al., 2007)
	Mean solar radiation 2014	Solar radiation	W/m <sup>2</sup> /year	Srad	DEM (CNIG, 2016)
Organisms	Land cover 2014	Deciduous forest	% occupation	DF	Landsat 8 OLI (USGS, 2016)
		Pine forest	% occupation	PF	Landsat 8 OLI (USGS, 2016)
		Shrub land	% occupation	SL	Landsat 8 OLI (USGS, 2016)
		Agricultural land	% occupation	AL	Landsat 8 OLI (USGS, 2016)
		Grassland	% occupation	GL	Landsat 8 OLI (USGS, 2016)
		Rock outcrops	% occupation	Rock	Landsat 8 OLI (USGS, 2016)
		Urban	% occupation	Urban	Landsat 8 OLI (USGS, 2016)
	Average normalized difference vegetation index (NDVI) 2014	NDVI	-	NDVI	Landsat 8 OLI (USGS, 2016)
	Average normalized difference water index (NDWI) 2014	NDWI	-	NDWI	Landsat 8 OLI (USGS, 2016)
	Brightness vegetation	Tasseled cap component 1, - brightness	-	Bn	Landsat 8 OLI (USGS, 2016)
Greenness vegetation	Tasseled cap component 2, - greenness	-	Gn	Landsat 8 OLI (USGS, 2016)	
Wetness vegetation	Tasseled cap component 3, - wetness	-	Wn	Landsat 8 OLI (USGS, 2016)	
Average vegetation height 2014	Vegetation height	m	Vegheight	LiDAR PNOA (CNIG, 2016)	
Relief	Digital elevation	Altitude	m	Alt	DEM (CNIG, 2016)
	Model	Slope	Degrees	Sl	DEM (CNIG, 2016)
		Southness	°	South	DEM (CNIG, 2016)
		Eastness	°	East	DEM (CNIG, 2016)
		Topographic wetness index	-	TWI	DEM (CNIG, 2016)
Parent material	Geological map	Geological unit	Categorical	GU	GEODE, (CNIG, 2016)
Time		Lithology class 1	Categorical	Lit1	GEODE, (CNIG, 2016)
		Lithology class 2	Categorical	Lit2	GEODE, (CNIG, 2016)
		Age 1	Categorical	Age1	GEODE, (CNIG, 2016)
		Age 2	Categorical	Age2	GEODE, (CNIG, 2016)

data represent the soil-forming factor “organisms”. A composite, cloud-free land-use map was created using Landsat 8 Operational Land Imager (OLI) scene mosaic data from

2013 to 2016 (Path 202, Row 30; Álvarez-Martínez et al., 2018). This satellite image was also used to derive the Normalized Difference Vegetation Index (NDVI; Rouse,



**FIGURE 2** Soil sampling procedure. The area was divided into 12 strata (a). Twelve lithology classes were reclassified from the continuous geological map (b). If present, three land-cover classes, including forest, shrubs and grassland, were sampled in each occurring combination of strata and lithology. The accessibility of the study area is limited and therefore the samples were taken within a 2-km buffer around the paved roads (c) [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

Haas, Schell, & Deering, 1973), the Normalized Difference Water Index (NDWI) and the Tasselled Cap (TC) transformation (Crist & Cicone, 1984). Data on vegetation height were derived from LiDAR data (Álvarez-Martínez et al., 2018). Relief is represented by a digital elevation model at 5-m resolution, obtained from the National Geographical Information Centre of Spain (CNIG, 2016), from which various topographic parameters were calculated (e.g., slope and topographic wetness index). Parent material and time were represented by the Continuous Geological Map of Spain from the Geological and Mining Institute of Spain. All variables were resampled to 30-m resolution before they were used for the analysis. For continuous variables the nearest neighbour assignment was used to resample the original datasets, whereas the majority algorithm was used for categorical variables.

## 2.2.2 | Soil sampling and laboratory analysis

For the calibration of the mechanistic model, soil data were collected using a sampling protocol that covers most

of the spatial soil variation within the study area. The area was stratified using a two-step approach. Firstly, the environmental variables were analysed by a principal component analysis (PCA). Secondly, a cluster analysis (CA) was carried out on the results of the PCA using the Iso cluster algorithm (Ball & Hall, 1967). The CA sets a threshold for the maximum number of classes at 20. The algorithm performs an iterative, self-organizing clustering. Clusters will merge with neighbours when their statistical values are similar after the clusters become stable. Initial strata were evaluated in terms of expert knowledge related to vegetation composition (see Álvarez-Martínez et al., 2018; Rodríguez-Arango et al., 2003) and freely available land-cover and vegetation maps such as The National Forestry Inventory of Spain, and Corine Land Cover maps. This process divided the study area into 12 homogeneous strata (Figure 2a). The geological map and land-use map were not included in the CA because these two maps were already used to subdivide the area into 12 homogeneous strata. We are aware of the potential loss of information caused by the removal of these two maps. The lithology classes of the geological map were reclassified into 12 major classes (Figure 2b). If present,

three land-cover classes (forest, shrubs and pasture) were sampled in each occurring combination of strata and lithology. This resulted in 100 sampling locations. The accessibility of the study area is limited and therefore the samples were taken within a 2-km buffer around the paved roads (Figure 2c). A representative sample per stratum was taken within the accessible area at a location where the slope, altitude, vegetation height and NDVI were most similar to the average of the unit. The sampling protocol was specifically designed for the calibration of the mechanistic model by maximizing the variation in sampling conditions. In the second step of DSM (i.e., kriging of the residuals), the spatial autocorrelation of the residuals is considered (Lark et al., 2006; Lark & Cullis, 2004). The limitation of the sampling design is that it can result in an insufficient number of samples at different intervals apart to model the spatial autocorrelation.

A composite sample was taken at each sampling location. The design draws a square and takes one sample in the centre of the square and four in each corner. The distance from the centre to the corner was about 5 m. Each composite sample was assumed to represent a support of 50m<sup>2</sup>. At each point, the depth of the soil profile (i.e., depth to saprolite or bedrock) was annotated and a mixed sample from the entire soil profile was taken. The depth of the soil profile was averaged over the five points and the samples were thoroughly mixed. After removing coarse fragments (> 2000µm) and roots, the soil samples were analysed for SOM content using the Walkley and Black method (Walkley, 1947; Walkley & Black, 1934). The SOM stocks were estimated by multiplying the SOM content by the soil profile depth and the bulk density. Soil texture was determined using the pipette method (Gee & Or, 2002). The bulk density was estimated on the basis of soil texture and soil organic matter according to the pedotransfer function of Balland, Pollacco, and Arp (2008). The input parameters for this pedotransfer function were soil depth and soil organic matter content. This sampling protocol did not account for changes in soil properties (e.g., bulk density) within the soil profile and did not account for stoniness within the soil profile to estimate the SOM stock of the soil profile.

### 2.3 | Mechanistic model

The mechanistic model was developed in four steps.

Step 1: Selection of major processes that influence SOM stocks and that dominate at the regional scale. The selection was based on a comparison of soil models that describe associated carbon (C) and nitrogen (N) (Grace & Malhi, 2002; Shibu et al., 2006). Processes that could be

modelled at a regional scale and that occurred most frequently in soil models were selected.

Step 2: Evaluation of processes in various soil models. Much research has been carried out on the understanding of the carbon flows in the soil system (Shibu et al., 2006). Soil models often describe a process using similar parameters and relationships. For example, the amount of SOM that turns into CO<sub>2</sub> and nutrients is considered to be exponentially related to the potential mineralizable carbon, the mineralization rate and time (Andr n & Paustian, 1987; Rey & Jarvis, 2006; Stanford & Smith, 1972). To describe the major processes in our model, we make use of these known and exhaustively studied relationships.

Step 3: Collection of the input data for the model. It is likely that not all variables that are used for mechanistic soil models are available as maps and therefore the model potentially needs to make use of proxies or default values. The available environmental variables listed in Table 1 can be selected for the model, either directly or as a proxy.

Step 4: It can be assumed that natural systems, or systems that did not undergo major changes over last decades, have reached a certain level of equilibrium (Jenkinson, Adams, & Wild, 1991; Pimm, 1982). Therefore, the model included various fixed parameters. These parameters were estimated using the generalized reduced gradient non-linear algorithm. As such, the Lin's concordance correlation coefficient (Lin, 1989) is maximized, whereas the root mean square error (RMSE) of the residuals is minimized. The model sensitivity is tested by re-estimating the fixed parameters with various initializations of the fixed parameters.

### 2.4 | Statistical model

To analyse the potential of a mechanistic approach for DSM, we also estimated a statistical model. For a better comparison of the statistical model and the mechanistic model, regression kriging was selected in which the linear regression can be replaced directly by the mechanistic model. Subsequently, the residuals of both models can be analysed in a similar way. It should be noted that regression kriging is one method out of a wide array of different statistical models (e.g. linear mixed models, neural networks and structural equation modelling).

In comparison to the mechanistic model, all the environmental variables listed in Table 1 enter the regression model. Variables that show most statistically insignificant deterioration of the model fit are one by one deleted from the regression equation using backward elimination. The regression is completed when no other variable can be

deleted without a statistically significant loss of fit. There are many other, and perhaps even better, statistical models available. We chose regression kriging because it is easy to understand by the wider audience. Lamichhane, Kumar, and Wilson (2019) gave a detailed review of other digital DSM techniques and their suitability for predicting SOC, but the study also confirmed the strengths of regression kriging.

## 2.5 | Evaluation of the SOM stock models and maps

The SOM stock models and maps that result from the mechanistic and statistical model were evaluated by estimating:

1. The goodness of fit for each model by calculating the bias (Equation (1)), error (Equation (2)), the Lin's concordance correlation coefficient ( $\rho_c$ ) (Equation (3)) and the variation explained by the model (Equation (4)):

$$ME = \frac{1}{n} \sum_{i=1}^n (SOM_{obs,i} - SOM_{pred,i}), \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (SOM_{obs,i} - SOM_{pred,i})^2}, \quad (2)$$

in which  $ME$  is the mean error,  $RMSE$  is the root mean square error,  $SOM_{obs,i}$  is the observed SOM stocks at location  $i$ ,  $SOM_{pred,i}$  is the predicted SOM stocks at location  $i$ , and  $n$  is the total number of samples.

$$\rho_c = \frac{2\rho\sigma_{obs}\sigma_{pred}}{\sigma_{obs}^2 + \sigma_{pred}^2 + (\mu_{obs} - \mu_{pred})^2}, \quad (3)$$

in which  $\mu_{obs}$  and  $\mu_{pred}$  are the means of the observed and predicted SOM stocks,  $\sigma_{obs}$  and  $\sigma_{pred}$  are the corresponding variances and  $\rho$  is the correlation coefficient between the observed and predicted SOM stocks.

$$AVE = \frac{\sum_{i=1}^n (\hat{SOM}_i - \bar{SOM})^2}{\sum_{i=1}^n (SOM_i - \bar{SOM})^2}, \quad (4)$$

in which  $AVE$  is the amount of variance explained,  $\hat{SOM}_i$  the predicted SOM stocks at location  $i$  and  $\bar{SOM}$  is the overall SOM stocks mean.

2. The  $ME$ ,  $RMSE$ , Lin's concordance correlation coefficient and the  $AVE$  of the SOM stock maps using the soil observations.

3. The Pearson correlation coefficient between the two SOM stock maps to compare both models and maps.

## 3 | RESULTS AND DISCUSSION

### 3.1 | Data collection

#### 3.1.1 | Environmental variables

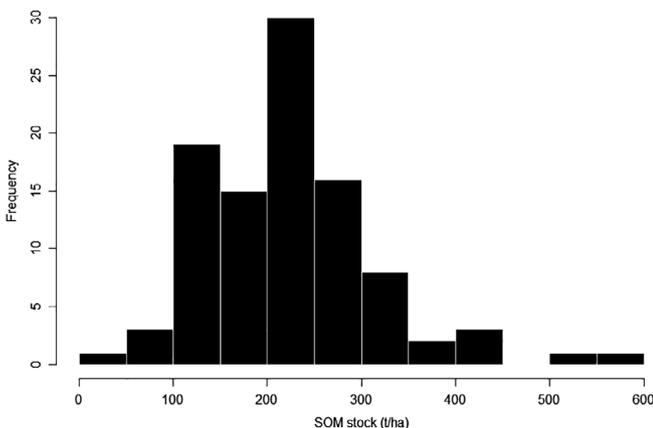
The area is dominated by deciduous forest (38%) and shrubland (43%), followed by grassland (12%) and rock outcrops (6%). The average and interquartile range (IQR) of the environmental variables are given in Table 2. Vegetation height, maximum temperature and slope have the highest IQR, indicating largest variation within the study area. Some pairs of environmental variables are highly correlated with each other. For statistical models it is preferred that environmental variables are weakly correlated with each other, because it increases the potential for fitting a combination of environmental variables to explain the variation in SOM stocks. High correlation coefficients existed, for example, between vegetation height and NDVI or NDWI, altitude and minimum temperature, and solar radiation and southness of slope.

#### 3.1.2 | Summary statistics of soil observations

Soil profile depths ranged between 10 and 27 cm. Bedrock or saprolite was reached immediately below the A-horizon. The bulk density was predicted to be between 0.84 and 1.21 g cm<sup>-3</sup>. The soil profile depth and bulk density were used to estimate SOM stocks. The SOM stocks ranged between 164 and 482 t ha<sup>-1</sup>. The SOM stocks of the samples were positively skewed (Figure 3), with Q1, Q2 and Q3 being 193, 222 and 248 t ha<sup>-1</sup>, respectively. Although vegetation was assumed to be a good predictor for SOM stocks, a weak correlation coefficient was found between SOM stocks and vegetation height and between SOM stocks and NDVI. This can be caused by the low mineralization rate in the sparsely vegetated, wet and cold mountainous areas in the southern part of the study area. A stronger correlation coefficient was found between SOM stocks and altitude ( $r = 0.44$ ), probably also caused by the low mineralization rate in colder climates (Manzoni & Porporato, 2009). There is also a negative correlation between SOM stocks and slope ( $r = -0.22$ ), which indicates that erosional processes might influence the SOM stocks in the study area. In general, the Pearson correlation coefficient between SOM stocks and environmental variables is weak (Table 3).

**TABLE 2** Average and interquartile range (IQR) of environmental variables for the Natura2000 areas in the Cantabria region (Spain)

	Average	IQR
Precipitation (mm)	1,260	289
Max. temperature (°C)	11.4	7.6
Mean temperature (°C)	5.6	2.2
Min. temperature (°C)	3.9	1
Solar radiation (W/m <sup>2</sup> /year)	1,107,690	318,692
Normalized difference vegetation index (-)	0.64	0.19
Normalized difference water index (-)	0.49	0.15
Brightness	0.85	0.15
Greenness	0.30	0.18
Wetness	0.19	0.08
Vegetation height (m)	4.40	7.30
Altitude (m)	1,067	587
Slope (°)	24	14
Southness	-0.16	-
Eastness	0.08	-
Topographic wetness index	9.8	1.5

**FIGURE 3** The frequency distribution of the soil organic matter stocks resulting from 100 soil samples that were collected in the Natura2000 areas of the Cantabria region (Spain)

In our study, soil samples with an exceptional high organic matter content of 17% or higher were tested twice in the laboratory and resulted in nearly the same values. Nine out of 13 samples with organic matter content above 17% were taken in grasslands, which could have been influenced by human activities. The other samples corresponded to the SOM contents that were measured in the Cantabria region by Rodríguez Martín et al. (2016). These grasslands are dominantly located in the highlands

**TABLE 3** Pearson correlation coefficients between soil organic matter (SOM) stocks and available environmental variables

Environmental variable	SOM stocks
NDVI <sup>a</sup>	-0.01
TWI <sup>b</sup>	-0.02
NDWI <sup>c</sup>	-0.01
Elevation	0.44
Slope	-0.22
Radiation	0.16
Southness	-0.10
Mean precipitation	-0.02
Mean temperature	-0.42
CV <sup>d</sup> of precipitation	-0.06
CV of temperature	0.36

<sup>a</sup>Normalized Difference Vegetation Index.

<sup>b</sup>Topographic wetness index.

<sup>c</sup>Normalized Difference Wetness Index.

<sup>d</sup>Coefficient of Variation.

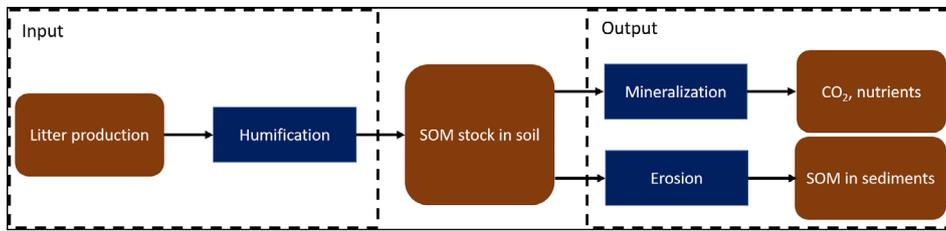
where the mineralization rates are lower due to the low temperatures or in poorly drained valleys. In total, 48 samples were collected in shrubland, 43 in forest and nine in grasslands.

## 3.2 | Mechanistic model

### 3.2.1 | Developing the mechanistic model

The mechanistic model was developed by applying the four steps described in Section 2.3.

Step 1. Available soil models describe carbon processes in different levels of detail and complexity (Grace & Malhi, 2002), but the processes each carbon model includes can be framed in the following generic way: (a) part of the organic matter inputs, including roots, wood and leaves, decomposes, (b) resistant plant material breaks down into CO<sub>2</sub>, microbial biomass and humified organic matter, (c) organic matter humifies and enters the organic matter pool(s), and (d) carbon mineralizes, releasing CO<sub>2</sub> and nutrients. Besides mineralization, erosion can also cause a severe depletion of the organic carbon pool (Lal, 2003). Based on the characteristics of our study area, we decided to select erosion as a major process as well. This resulted in the selection of three major processes: humification, mineralization and erosion. A conceptual framework of the mechanistic model is shown in Figure 4. Start and end products are indicated by the



**FIGURE 4** Conceptual framework of the mechanistic model that is used for predicting soil organic matter (SOM) stocks using digital soil mapping. Start and end products of the flow diagram are indicated by the rounded boxes and processes are indicated by the squared boxes [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

rounded boxes and the major processes by the squared boxes.

Steps 2 and 3. Compared to dynamic soil models, the processes in our mechanistic model had to be simplified. Processes had to be described at regional scale and by environmental variables that are spatially available. The mechanistic model also considered only one organic carbon pool to reduce complexity. The relatively simple carbon model Roth-C (Coleman & Jenkinson, 2014) was used as a starting point for building the mechanistic SOC model.

The SOM input was estimated by multiplying the litter production (LP) by the humification rate (HR). The LP depends on vegetation type and cover. Temperate deciduous forests produce between 8,800 and 14,100 kg ha<sup>-1</sup> per year (Tateno, Hishi, & Takeda, 2004) and the Carpathian grasslands produce about 1,470 to 2,870 kg ha<sup>-1</sup> litter per year (Galvánek & Lepš, 2012). A linear relationship was considered between litter production and NDVI. Therefore, NDVI is taken as a proxy for the LP (Equation (5)):

$$LP = c_1 + c_2 NDVI, \quad (5)$$

in which  $c_1$  and  $c_2$  are fixed parameters.

The humification rate (HR) of humified organic matter to soil organic matter depends on the clay content (Coleman & Jenkinson, 2014). Organic matter binds to clay particles and therefore clay affects the way organic matter decomposes (Coleman & Jenkinson, 2014). The lithology class can be used to represent the clay content. However, the course scale of the lithology map made us decide to fix HR at 0.32. The clay content of the soil samples was used as input data for the Coleman and Jenkinson (2014) relationship (Equation (6)) to estimate the HR content. The value 0.32 is the average of the resulting HR contents.

$$HR = \frac{1}{(3.09 + 2.7e^{(CL)})}, \quad (6)$$

in which  $CL$  is the observed clay content (%).

The soil releases nutrients and CO<sub>2</sub> through mineralization. The mineralization rate (MR) is exponentially related to the temperature. In the study area, mean annual temperature and precipitation were correlated ( $r = 0.28$ ). The temperature showed the strongest correlation with the observed organic matter content ( $r = -0.42$ ) and therefore the MR was estimated using Equation (7).

$$MR = c_3 + e^{c_4 T}, \quad (7)$$

in which  $c_3$  and  $c_4$  are fixed parameters and  $T$  is temperature (°C).

The Universal Soil Loss Equation (USLE) is a commonly used equation to estimate the erosion rate (ER) (Wischmeier & Smith, 1965; Wischmeier & Smith, 1978). In this equation, the erosion rate depends on slope, precipitation and vegetation cover. Except for some rocky outcrops, the entire area has a permanent vegetation cover, which made us decide to use a fixed value for vegetation cover. Erosion is estimated using slope and rainfall erosivity (Equation (8)).

$$ER = c_5 * ((c_6 * S) + (c_7 * R)), \quad (8)$$

in which  $c_5$ ,  $c_6$  and  $c_7$  are fixed parameters,  $S$  is the slope (°) and  $R$  is the rainfall erosivity (MJ · mm ha<sup>-1</sup> h<sup>-1</sup> yr<sup>-1</sup>) obtained from Panagos et al. (2017).

Step 4. To predict the SOM stocks, the balance  $SOM_{in} = SOM_{out}$  needs to be optimized (Equation (10)). This results in Equation (9) and finally in Equation (10).

$$LP * HR = (MR * SOM_{stock} + ER * (SOM)), \quad (9)$$

$$SOM_{stock} = \frac{LP * HR}{MR + \frac{ER}{BD * SD * 10,000}}, \quad (10)$$

in which  $BD$  is the soil bulk density (g cm<sup>-3</sup>), which is fixed at 1.2 g cm<sup>-3</sup> and  $SD$  is the soil depth (cm), which is fixed at the average soil depth of the collected samples (20 cm).

### 3.2.2 | Calibration of the mechanistic model

The model was calibrated using 99 soil samples, because one sample with an organic matter content of 34% was assumed to be an outlier. To make every process act within realistic boundaries, litter production ranged between 5 and 20 t ha<sup>-1</sup> and mineralization rate between 0.01 and 0.05. Optimizing the fixed parameters based on the objectives (minimizing the RMSE and maximizing the Lin's concordance correlation coefficient), resulted in the calibration results listed in Table 4. The mechanistic model predicted SOM stocks best by having  $c_3$  and  $c_7$  set to zero, meaning that no fixed parameter was required to predict mineralization rate and the erosion rate was predicted without an estimate of rainfall erosivity. A Lin's concordance correlation coefficient of 0.24 was found between SOM<sub>pred</sub> and SOM<sub>obs</sub>. The predicted SOM stocks show much less variation compared to the observed SOM stocks. This indicates that the explanatory variables cannot explain all spatial variation. After calibration of the model, SOM stocks ranged between 175 and 285 t SOM ha<sup>-1</sup>. The litter production was estimated at 19.3 t SOM ha<sup>-1</sup>, with a standard deviation of 0.17 t SOM ha<sup>-1</sup>. The mineralization rate varied between 0.01 and 0.03, and the erosion rate between 1.3 and 13.2 t ha<sup>-1</sup>. The modelled results had a mean error (ME) of -1 t ha<sup>-1</sup>, an RMSE of 68 t ha<sup>-1</sup> and an amount of variance explained (AVE) of 17%. The average SOM stocks balance (SOM<sub>in</sub> - SOM<sub>out</sub>) is negative (-0.17 t SOM ha<sup>-1</sup>).

Changing the fixed parameters affected the RMSE (Figure 5a) and the Lin's concordance correlation coefficient (Figure 5b) of the model unequally and non-systematically. Model performance did not respond symmetrically to lower or higher values of the fixed parameters. For example, increasing the value of  $c_1$  gives an exponential increase in the RMSE, whereas decreasing the value of  $c_1$  changes the RMSE to a more gradual, linear relation. The parameters  $c_3$  and  $c_7$ , which were used

for estimating the *MR* and *ER* respectively, did not show any sensitivity because the fixed parameters were estimated to be zero. The model was not sensitive to changes in  $c_5$  and  $c_6$  and most sensitive to changes in  $c_1$ ,  $c_2$  and  $c_4$ . From the sensitivity analysis it can be concluded that the fixed parameters that were used to estimate *LP* were most sensitive. Different parameters did not show minimum RMSE and maximum correlation coefficient at the value calibrated by the model, because the model aimed to optimize two objectives.

### 3.3 | Mechanistic SOM stocks map

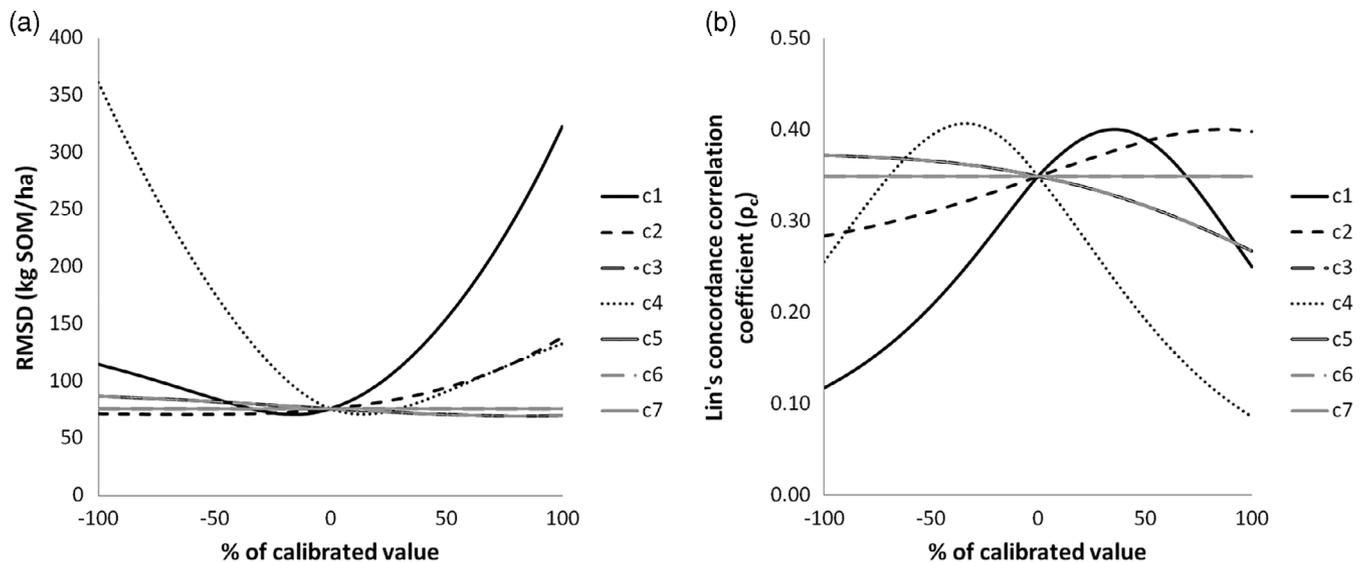
Mapping the SOM stocks for the Cantabria region resulted in values between 136 and 1,039 t SOM ha<sup>-1</sup> (Figure 6). The model did not show any spatial dependence between the residuals. The mapped SOM stocks had at the sampling locations an ME of -2 t SOM ha<sup>-1</sup> and an RMSE of 66 t SOM ha<sup>-1</sup>. The Lin's concordance correlation coefficient between the observed and predicted values was 0.47 and the AVE of the map was 21%.

### 3.4 | Statistical SOM stocks map

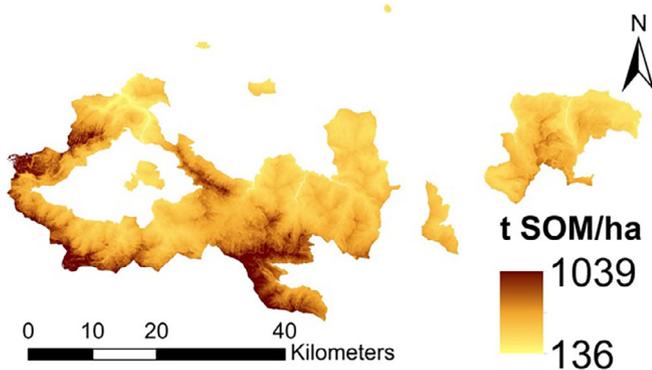
The stepwise regression resulted in the selection of five variables: altitude, slope, annual precipitation coefficient of variation (CV), mean annual temperature CV and topographic wetness index (TWI). The model predicted SOM stocks at the sampling locations between 123 and 282 t SOM ha<sup>-1</sup>. Mapping the SOM stocks for the Cantabria region resulted in values between 1 and 347 t SOM ha<sup>-1</sup> (Figure 7), reflecting the wide range in ecological conditions in the area. At the sampling locations, the difference between the predicted and observed values had an ME of 1 t SOM ha<sup>-1</sup> and an RMSE of 64 t SOM ha<sup>-1</sup>. The Lin's concordance correlation coefficient between the observed and predicted values was 0.42. Also, the statistical model did not show any spatial dependence between the residuals. The SOM stocks map of the Cantabria region had an ME of -0.4 t SOM ha<sup>-1</sup>, an RMSE of 62 t SOM ha<sup>-1</sup> and a Lin's concordance correlation coefficient of 0.49, and the AVE was 31%. The statistical model served as a comparison with the mechanistic approach, and therefore we used a straightforward regression kriging approach. It is expected that predictions can be improved by using a linear mixed model that uses the empirical best linear unbiased predictor to estimate random effects (Lark et al., 2006). However, if the residuals exhibit no spatial autocorrelation, ordinary least squares regression will provide equivalent results.

**TABLE 4** Fixed parameters used to fit the model

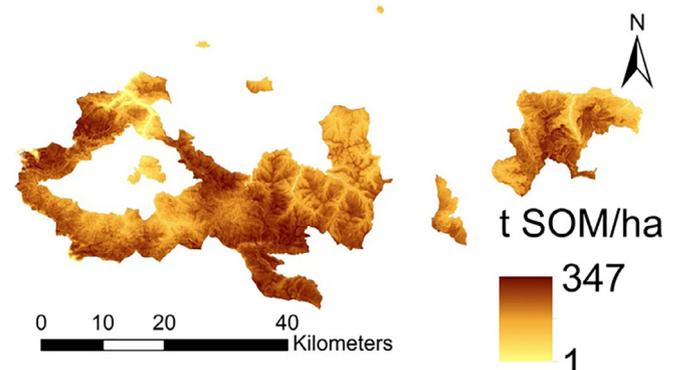
Fixed parameters	Value
$c_1$	14,618
$c_2$	1923
$c_3$	0
$c_4$	0.0024
$c_5$	0.6728
$c_6$	519.6
$c_7$	0



**FIGURE 5** Sensitivity analysis of the mechanistic model. Deviating model parameters (c1 to c7) individually in steps of 10% from the calibration point (0%) up to -100% and +100% results in asymmetrical changes of the root mean square error (RMSE) and the Lin's concordance correlation coefficient



**FIGURE 6** Predicted soil organic matter (SOM) stocks of the mechanistic model [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



**FIGURE 7** Predicted soil organic matter (SOM) stocks of the statistical model [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

### 3.5 | Evaluation of the SOM stock maps

Table 5 summarizes the statistical evaluation of the models and the SOM stock maps. The statistical model performs slightly better compared to the mechanistic model. The complexity of the system and the limited data we had to support and to validate the model makes it difficult to draw conclusions from the model predictions. At the sampling locations, the correlation coefficient between the two SOM stock maps was 0.8. The similarities between the two maps became especially visible when taking a snapshot of the central part of the study area (Figure 8). The highest organic matter contents were predicted in mountain ridges and stream valleys. The low mineralization rate in mountain ridges and SOM

accumulation in stream valleys increase SOM stocks. The lowest SOM stocks were found at the steepest slopes. This indicates that erosion processes have influenced SOM stocks in the study area.

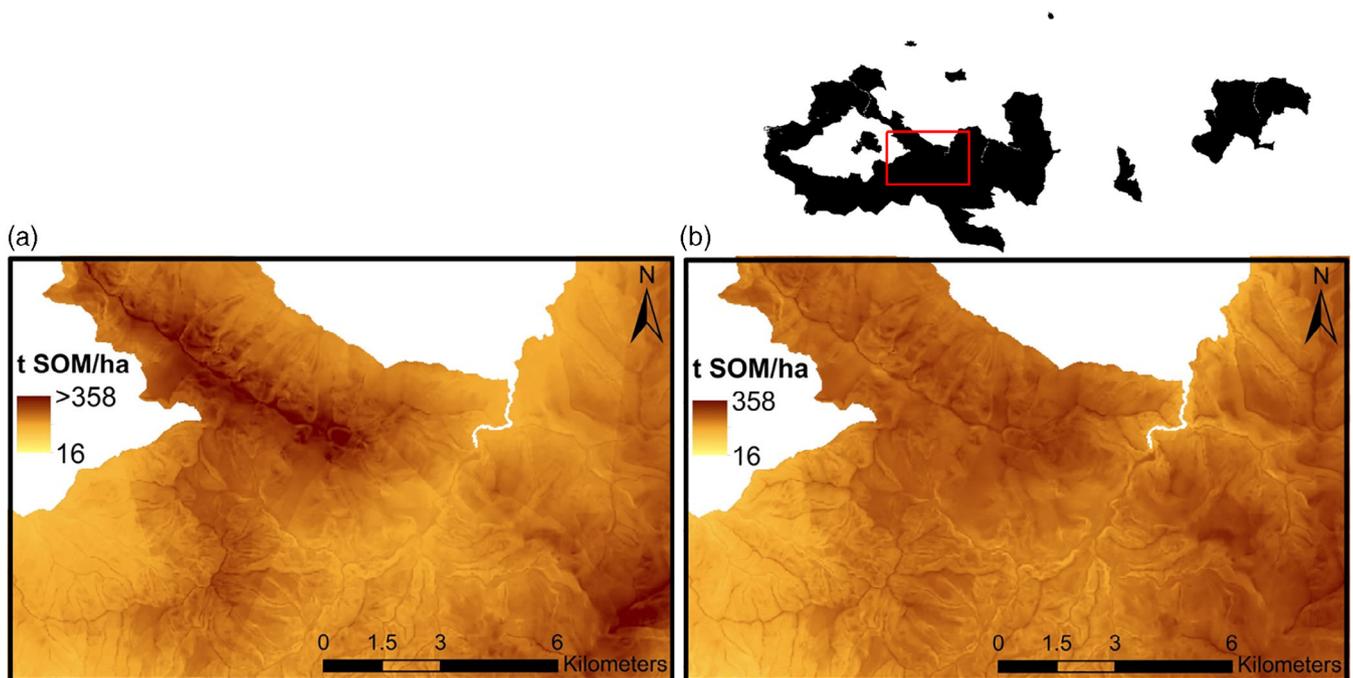
## 4 | GENERAL DISCUSSION

### 4.1 | Potential for combining mechanistic and statistical models

The study illustrated the potential of developing a mechanistic soil model as an alternative for statistical DSM. Although we do not have a convincing account of a mechanistic approach, this alternative brings new

**TABLE 5** Statistical analysis of the mechanistic soil organic matter (SOM) stocks model and map

	Mechanistic SOM stocks		Statistical SOM stocks	
	Model	Map	Model	Map
SOM stocks (t SOM/ha)	175–285	136–1,039	123–282	1–347
ME <sup>a</sup>	–1	–2	1	–0.4
RMSE <sup>b</sup>	68	66	64	62
Lin's concordance CC <sup>c</sup>	0.24	0.47	0.42	0.49
AVE <sup>d</sup>	0.17	0.21	0.26	0.31

<sup>a</sup>Mean error.<sup>b</sup>Root mean square error.<sup>c</sup>Correlation coefficient.<sup>d</sup>Amount of variance explained.**FIGURE 8** Taking a snapshot (red rectangular area) illustrating differences/similarities between the mechanistic soil organic matter (SOM) stocks map (a) and the statistical SOM stocks map (b) [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

opportunities for DSM. Instead of increasing the complexity of the algorithms of statistical models in order to improve SOM stock predictions (e.g., Hengl et al., 2017; Keskin, Grunwald, & Harris, 2019), the potential of combining mechanistic and statistical models should be explored. For example, one can use the fitted parameters of the mechanistic approach as input data for the statistical approach. The mechanistic DSM approach selected the environmental variables beforehand using pedological models, data and knowledge, whereas in a statistical DSM approach all environmental variables have the chance of being selected for the model. This resulted in the selection of different environmental variables

between the mechanistic and statistical models. Combining mechanistic and statistical models has a key advantage of the ability to build in causal effects based on pedological processes, which makes the prediction of soil properties underpinned and less of a black box.

## 4.2 | Potential for different soil properties and extrapolation

Cause–effect relationships of the complex soil characteristics need to be well known before a mechanistic model for DSM can be developed. The approach faces

limitations for soil properties with less well-understood cause–effect relationships (e.g., cation exchange capacity (CEC) and soil pH). In those cases, statistical models are probably still the best. Another reason for developing mechanistic models was the potential of extrapolation. There is evidence that the extrapolation of a mechanistic model to areas with similar characteristics and processes has more potential than the extrapolation of a statistical model, because the mechanistic model depends on cause–effect relationships, whereas the statistical model depends on associations. Several studies that used a statistical model for DSM explored the potential of extrapolation (Afshar, Ayoubi, & Jafari, 2018; Grinand, Arrouays, Laroche, & Martin, 2008; Wolski et al., 2017). However, the soil property predictions in these extrapolated areas were poor. Statistical models tend to exaggerate the error associated with the interpolation (Robinson & Metternicht, 2006). This can result in strong differences in accuracy between the interpolation and extrapolation areas (e.g., Grinand et al., 2008).

### 4.3 | Potential for other areas

Mechanistic DSM was applied to a nature area, because the area was assumed to be in a state of equilibrium between  $SOM_{in}$  and  $SOM_{out}$ . It is possible to incorporate a factor that accounts for systematic depletion or addition of SOM, but the potential to apply the approach to a non-balanced system is limited. The processes that are included in the mechanistic model need to be evaluated before the model is applied to a different nature area. This is because (a) processes can differ, (b) the relationships between the environmental variables and the process can differ, and (c) the availability of environmental variables can differ. The potential of applying mechanistic DSM to a non-balanced system, such as an agricultural landscape, should be explored.

### 4.4 | Model performance

The observed SOM stocks showed a wider range compared to the predicted SOM stocks. This means that the variation in SOM stocks could not be covered completely by the selected environmental variables. Lack of fit can have different causes: (a) the resolution of some environmental variables was not detailed enough to detect the causes of spatial variation in SOM stocks, (b) the spatial variation needs to be explained by different environmental variables (for example, the effect soil biota have on SOM stocks could not be included in this study), (c) the

number of observations was not sufficient to find relations between the observed and explanatory variables, or (d) the sampling protocol did not detect all variation, because different processes operate at different spatial scales.

The model relies on several fixed parameters, which can result in equifinality (i.e., the same outcome could be produced by several different combinations of parameter values). We tried to reduce the potential of equifinality by making every process act between realistic boundaries. Another disadvantage of the model is that the optimal sampling protocol for the calibration of the mechanistic model opposes the second step in DSM, namely the analysis of spatial autocorrelation between the residuals. Taking multiple randomly chosen soil samples within a stratum could have solved the issue, but increases the chance of missing some of the spatial soil variation.

## 5 | CONCLUSIONS

The study illustrated that mechanistic soil models can be used for DSM, which means that DSM no longer needs to rely on statistical models only. The model could be developed because the predicted soil property (SOM stocks) has well-known cause–effect relationships, the study area could be assumed to be in a state of equilibrium, and spatially exhaustive environmental variables were available at fine resolution. It is difficult to draw conclusions from the comparison because of the complexity of the system and the limited data we had to support and to validate the model. The conclusion we can draw from the comparison is that the two maps were highly correlated with each other. Mechanistic models for DSM are an interesting alternative for mapping soil characteristics that are difficult to predict by statistical models and for extrapolation purposes. The potential of combining mechanistic and statistical models for DSM should be explored, because both models have advantages and disadvantages that can complement each other.

### AUTHOR CONTRIBUTIONS

C.M.J.H., J.J.S., J.M.Á.-M. designed the research; C.M.J.H. conducted the research with supervision from J.J.S.; J.M.Á.-M. collected data, facilitated the fieldwork, and contributed to the interpretation of the results; C.M.J.H. and I.P.-S collaborated in the collection and analyses of soil samples. C.M.J.H. wrote the manuscript, but got substantial contributions and feedback from J.J.S., J.M.Á.-M., L.C., I.P.-S, and J.B. The final draft of the manuscript was approved by all authors.

## ACKNOWLEDGEMENTS

This research was financially supported by the Environmental Hydraulics Institute 'IH Cantabria of Universidad de Cantabria' and the CGIAR Research Programme on Climate Change, Agriculture and Food Security (CCAFS). The CCAFS project is carried out with support from CGIAR Fund Donors and through bilateral funding agreements. Besides the financial support, we would like to thank Sara Alcalde Aparicio for collaboration in the collection and analyses of soil samples.

## DATA AVAILABILITY STATEMENT

Data available on request from the authors.

## ORCID

Chantal Mechtildis Johanna Hendriks  <https://orcid.org/0000-0001-6749-7232>

Jetse Jacob Stoorvogel  <https://orcid.org/0000-0003-4297-122X>

Lieven Claessens  <https://orcid.org/0000-0003-2961-8990>

## REFERENCES

- Afshar, F. A., Ayoubi, S., & Jafari, A. (2018). The extrapolation of soil great groups using multinomial logistic regression at regional scale in arid regions of Iran. *Geoderma*, 315, 36–48. <https://doi.org/10.1016/j.geoderma.2017.11.030>
- Álvarez-Martínez, J. M., Jiménez-Alfaro, B., Barquín, J., Ondiviela, B., Recio, M., Silió-Calzada, A., & Juanes, J. A. (2018). Modelling the area of occupancy of habitat types with remote sensing. *Methods in Ecology and Evolution*, 9, 580–593. <https://doi.org/10.1111/2041-210X.12925>
- Álvarez-Martínez, J. M., Suárez-Seoane, S., Stoorvogel, J. J., & de Luis Calabuig, E. (2014). Influence of land use and climate on recent forest expansion: A case study in the Eurosiberian-Mediterranean limit of north-west Spain. *Journal of Ecology*, 102, 905–919. <https://doi.org/10.1111/1365-2745.12257>
- Andrén, O., & Paustian, K. (1987). Barley straw decomposition in the field: A comparison of models. *Ecology*, 68, 1190–1200. <https://doi.org/10.2307/1939203>
- Angelini, M. E., Heuvelink, G. B. M., Kempen, B., & Morrás, H. J. M. (2016). Mapping the soils of an Argentine Pampas region using structural equation modelling. *Geoderma*, 281, 102–118. <https://doi.org/10.1016/j.geoderma.2016.06.031>
- Ball, G., & Hall, F. (1967). A clustering technique for summarizing multivariate data. *Behavioral Science*, 12, 153–155. <https://doi.org/10.1002/bs.3830120210>
- Balland, V., Pollacco, A. P., & Arp, P. (2008). Modeling soil hydraulic properties for a wide range of soil conditions. *Ecological Modelling*, 219, 300–316. <https://doi.org/10.1016/j.ecolmodel.2008.07.009>
- Bot, A., & Benites, J. (2005). *The importance of soil organic matter: Key to drought-resistant soil and sustained food production*. FAO Soils Bulletin 80. Rome, Italy: Food and Agriculture Organization of the United Nations.
- Campbell, E. E., & Paustian, K. (2015). Current developments in soil organic matter modeling and the expansion of model applications: A review. *Environmental Research Letters*, 10, 123004. <https://doi.org/10.1088/1748-9326/10/12/123004>
- CNIG (2016). Digital Elevation Models: altimetric information of landforms and their elements. Centro Nacional de Información Geográfica (CNIG). Retrieved from [http://centrodedescargas.cnig.es/CentroDescargas/locale?request\\_locale=en](http://centrodedescargas.cnig.es/CentroDescargas/locale?request_locale=en)
- Coleman, K., & Jenkinson, D. S. (2014). *RothC – A model for the turnover of carbon in soil: Model description and users guide*. Harpenden, UK: Rothamsted Research.
- Crist, E. P., & Cicone, R. C. (1984). A physically-based transformation of thematic mapper data – The TM tasseled cap. *IEEE Transactions on Geoscience and Remote Sensing*, GE-22, 256–263. <https://doi.org/10.1109/TGRS.1984.350619>
- Dorji, T., Odeh, I. O. A., Field, D. J., & Baillie, I. C. (2014). Digital soil mapping of soil organic carbon stocks under different land use and land cover types in montane ecosystems, Eastern Himalayas. *Forest Ecology and Management*, 318(Supplement C), 91–102. <https://doi.org/10.1016/j.foreco.2014.01.003>
- Elzein, A., & Balesdent, J. (1995). Mechanistic simulation of vertical distribution of carbon concentrations and residence times in soils. *Soil Science Society of America*, 59, 1328–1335. <https://doi.org/10.2136/sssaj1995.03615995005900050019x>
- Carballas, T., Rodríguez-Rastrero, M., Artieda, O., Gumuzzio, J., Díaz-Ravina, M., Martín, Á (2016). Soils of the temperate humid zone. In: Gallardo, J. F. (eds.) The soils of Spain. Cham, Switzerland: Springer International Publishing. [https://doi.org/10.1007/978-3-319-20541-0\\_1](https://doi.org/10.1007/978-3-319-20541-0_1)
- Galvánek, D., & Lepš, J. (2012). The effect of management on productivity, litter accumulation and seedling recruitment in a Carpathian mountain grassland. *Plant Ecology*, 213, 523–533. <https://doi.org/10.1007/s11258-011-9999-7>
- Gee, W. G., & Or, D. (2002). Particle-size analysis. In J. Dane & G. C. Topp (Eds.), *Methods of soil analysis. Book series: 5. Part 4* (pp. 255–293). Madison, WI: Soil Science Society of America.
- Grace, J., & Malhi, Y. (2002). Global change: Carbon dioxide goes with the flow. *Nature*, 594, 594–595. <https://doi.org/10.1038/416594b>
- Grinand, C., Arrouays, D., Laroche, B., & Martin, M. P. (2008). Extrapolating regional soil landscapes from an existing soil map: Sampling intensity, validation procedures, and integration of spatial context. *Geoderma*, 143, 180–190. <https://doi.org/10.1016/j.geoderma.2007.11.004>
- Rodríguez-Arango, B. F., Arriola, E. F., Díaz, G.S., López, C.C., Gómez, B. A. M. (2003). Los pastos en Cantabria y su aprovechamiento memoria. Centro de Investigación y formación agrarias (CIFA), Muriendas, Spain.
- Hengl, T., Heuvelink, G. B. M., & Rossiter, D. G. (2007). About regression-kriging: From equations to case studies. *Computers & Geosciences*, 33, 1301–1315. <https://doi.org/10.1016/j.cageo.2007.05.001>
- Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., ... Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLoS One*, 12, e0169748.
- IUSS Working Group WRB (2015). World reference base for soil resources 2014, update 2015. International soil classification system for naming soils and creating legends for soil maps. World Soil Resources Reports No. 106. FAO, Rome.
- Jenkinson, D. S., Adams, D. E., & Wild, A. (1991). Model estimates of CO<sub>2</sub> emissions from soil in response to global warming. *Nature*, 351, 304–306. <https://doi.org/10.1038/351304a0>

- Jenny, H. (1941). *Factors of soil formation: A system of quantitative pedology*. New York, NY: McGraw-Hill.
- Kempen, B., Brus, D. J., & Stoorvogel, J. J. (2011). Three-dimensional mapping of soil organic matter content using soil type-specific depth functions. *Geoderma*, *162*, 107–123. <https://doi.org/10.1016/j.geoderma.2011.01.010>
- Keskin, H., Grunwald, S., & Harris, W. G. (2019). Digital mapping of soil carbon fractions with machine learning. *Geoderma*, *339*, 40–58. <https://doi.org/10.1016/j.geoderma.2018.12.037>
- Lal, R. (2003). Soil erosion and the global carbon budget. *Environment International*, *29*, 437–450. [https://doi.org/10.1016/S0160-4120\(02\)00192-7](https://doi.org/10.1016/S0160-4120(02)00192-7)
- 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>
- Lark, R. M., & Cullis, B. R. (2004). Model-based analysis using REML for inference from systematically sampled data on soil. *European Journal of Soil Science*, *55*, 799–813. <https://doi.org/10.1111/j.1365-2389.2004.00637.x>
- Lark, R. M., Cullis, B. R., & Welham, S. J. (2006). On spatial prediction of soil properties in the presence of a spatial trend: The empirical best linear unbiased predictor (E-BLUP) with REML. *European Journal of Soil Science*, *57*, 787–799. <https://doi.org/10.1111/j.1365-2389.2005.00768.x>
- Lin, L. I.-K. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, *45*, 255–268. <https://doi.org/10.2307/2532051>
- Ma, Y., Minasny, B., & Wu, C. (2017). Mapping key soil properties to support agricultural production in eastern China. *Geoderma Regional*, *10*, 144–153. <https://doi.org/10.1016/j.geodrs.2017.06.002>
- Manzoni, S., & Porporato, A. (2009). Soil carbon and nitrogen mineralization: Theory and models across scales. *Soil Biology and Biochemistry*, *41*, 1355–1379. <https://doi.org/10.1016/j.soilbio.2009.02.031>
- 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](https://doi.org/10.1016/S0016-7061(03)00223-4)
- Minasny, B., & McBratney, A. B. (2006). Mechanistic soil–landscape modelling as an approach to developing pedogenetic classifications. *Geoderma*, *133*, 138–149. <https://doi.org/10.1016/j.geoderma.2006.03.042>
- Minasny, B., & McBratney, A. B. (2016). Digital soil mapping: A brief history and some lessons. *Geoderma*, *264*, 301–311. <https://doi.org/10.1016/j.geoderma.2015.07.017>
- Ninyerola, M., Pons, X., & Roure, J. M. (2007). Objective air temperature mapping for the Iberian Peninsula using spatial interpolation and GIS. *International Journal of Climatology*, *27*, 1231–1242. <https://doi.org/10.1002/joc.1462>
- Panagos, P., Borelli, P., Meusburger, K., Yu, B., Klik, A., & Lim, K. J., et al. (2017). Global rainfall erosivity assessment based on high-temporal resolution rainfall records. *Scientific Reports*, *7*, 4175. <https://doi.org/10.1038/s41598-017-04282-8>
- Pimm, S. L. (1982). *Food webs*. London, UK: Chapman and Hall. [https://doi.org/10.1007/978-94-009-5925-5\\_1](https://doi.org/10.1007/978-94-009-5925-5_1)
- Rey, A., & Jarvis, P. (2006). Modelling the effect of temperature on carbon mineralization rates across a network of European forest sites (FORCAST). *Global Change Biology*, *12*, 1894–1908. <https://doi.org/10.1111/j.1365-2486.2006.01230.x>
- Robinson, T. P., & Metternicht, G. (2006). Testing the performance of spatial interpolation techniques for mapping soil properties. *Computers and Electronics in Agriculture*, *50*, 97–108. <https://doi.org/10.1016/j.compag.2005.07.003>
- 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*(Part A), 117–125. <https://doi.org/10.1016/j.geoderma.2015.10.010>
- Rosenbloom, N. A., Harden, J. W., Neff, J. C., & Schimel, D. S. (2006). Geomorphic control of landscape carbon accumulation. *Journal of Geophysical Research*, *111*, G01004. <https://doi.org/10.1029/2005JG000077>
- Rouse, J. W., Haas, R. H., Schell, J. A., Deering, D. W. (1973). Monitoring vegetation systems in the Great Plains with ERTS. In 3rd ERTS Symposium, NASA SP-351 I: 309–317. DOI: 19740022614.
- Shibu, M. E., Leffelaar, P. A., Van Keulen, H., & Aggarwal, P. K. (2006). Quantitative description of soil organic matter dynamics—A review of approaches with reference to rice-based cropping systems. *Geoderma*, *137*, 1–18. <https://doi.org/10.1016/j.geoderma.2006.08.008>
- Stanford, G., & Smith, S. J. (1972). Nitrogen mineralization potentials of soils 1. *Soil Science Society of America Journal*, *36*, 465–472. <https://doi.org/10.2136/sssaj1972.03615995003600030029x>
- Sundseth, K., Wegefelt, S., & Mézard, N. (2008). *NATURA 2000: Protecting Europe's biodiversity*. Oxford, UK: European Commission Information Press.
- Tateno, R., Hishi, T., & Takeda, H. (2004). Above- and belowground biomass and net primary production in a cool-temperate deciduous forest in relation to topographical changes in soil nitrogen. *Forest Ecology and Management*, *193*, 297–306. <https://doi.org/10.1016/j.foreco.2003.11.011>
- USGS (2016). Landsat—Earth observation satellites (ver. 1.1, August 2016). U.S. Geological Survey Fact Sheet 2015–3081: DOI: <https://doi.org/10.3133/fs20153081>.
- Walkley, A. (1947). A critical examination of a rapid method for determining organic carbon in soils - effect of variations in digestion conditions and of inorganic soil constituents. *Soil Science*, *63*, 251–264.
- Walkley, A., & Black, I. A. (1934). An examination of Degtjareff method for determining soil organic matter and a proposed modification of the chromic acid titration method. *Soil Science*, *37*, 29–37.
- Wischmeier, W. H., & Smith, D. D. (1978). Predicting rainfall erosion losses. In *Agriculture Handbook No. 537*. Washington, DC: U.S. Department of Agriculture.
- Wischmeier, W. H., & Smith, D. D. (1965). Predicting rainfall-erosion losses from cropland east of the Rocky Mountains—Guide for selection of practices for soil and water conservation. In *Agricultural Handbook No. 282*. Washington, DC: U.S. Government Printing Office.
- Wolski, M. S., Dalmolin, R. S. D., Flores, C. A., Moura-Bueno, J. M., Ten Caten, A., & Kaiser, D. R. (2017). Digital soil mapping and its implications in the extrapolation of

soil-landscape relationships in detailed scale. *Pesquisa Agropecuária Brasileira*, 52, 633–642. <https://doi.org/10.1590/s0100-204x2017000800009>

Zhao, M.-S., Rossiter, D. G., Li, D.-C., Zhao, Y.-G., Liu, F., & Zhang, G.-L. (2014). Mapping soil organic matter in low-relief areas based on land surface diurnal temperature difference and a vegetation index. *Ecological Indicators*, 39(Supplement C), 120–133. <https://doi.org/10.1016/j.ecolind.2013.12.015>

**How to cite this article:** Hendriks CMJ, Stoorvogel JJ, Álvarez-Martínez JM, Claessens L, Pérez-Silos I, Barquín J. Introducing a mechanistic model in digital soil mapping to predict soil organic matter stocks in the Cantabrian region (Spain). *Eur J Soil Sci.* 2021;72:704–719. <https://doi.org/10.1111/ejss.13011>