



High resolution 3D mapping of soil organic carbon in a heterogeneous agricultural landscape



Marine Lacoste^{a,*}, Budiman Minasny^b, Alex McBratney^b, Didier Michot^{a,c},
Valérie Viaud^a, Christian Walter^{a,c}

^a INRA, AGROCAMPUS OUEST, UMR1069, Sol Agro et hydrosystème Spatialisation, F-35000 Rennes, France

^b Faculty of Agriculture, Food & Natural Resources, The University of Sydney, NSW 2006, Australia

^c Université Européenne de Bretagne, France

ARTICLE INFO

Article history:

Received 18 January 2012

Received in revised form 2 July 2013

Accepted 2 July 2013

Available online 15 September 2013

Keywords:

Soil organic carbon

Soil bulk density

Depth functions

Data mining

LiDAR DEM

3D mapping

ABSTRACT

Soil organic carbon (SOC) is a key element of agroecosystems functioning and has a crucial impact on global carbon storage. At the landscape scale, SOC spatial variability is strongly affected by natural and anthropogenic processes and linear anthropogenic elements (such as hedges or ditches). This study aims at mapping SOC stocks distribution in the A-horizons for a depth up to 105 cm, at a high spatial resolution, for an area of 10 km² in a heterogeneous agricultural landscape (North-Western France). We used a data mining tool, Cubist, to build rule-based predictive models and predict SOC content and soil bulk density (BD) from a calibration dataset at 8 standard layers (0 to 7.5 cm, 7.5 to 15 cm, 15 to 30 cm, 30 to 45 cm, 45 to 60 cm, 60 to 75 cm, 75 to 90 cm and 90 to 105 cm). For the models calibration, 70 sampling locations were selected within the whole study area using the conditioned Latin hypercube sampling method. Two independent validation datasets were used to assess the performance of the predictive models: (i) at landscape scale, 49 sampling locations were selected using stratified random sampling based on a 300-m square grid; (ii) at hedge vicinity, 112 sampling locations were selected along transects perpendicular to 14 purposively chosen hedges. Undisturbed samples were collected at fixed depths and analysed for BD and SOC content at each sampling location and continuous soil profiles were reconstructed using equal-area splines. Predictive environmental data consisted in attributes derived from a light detection and ranging digital elevation model (LiDAR DEM), geological variables, land use data and a predictive map of A-horizon thickness. Considering the two validation datasets (at landscape scale and hedge vicinity), root mean square errors (RMSE) of the predictions, computed for all the standard soil layers (up to a depth of 105 cm), were respectively 7.74 and 5.02 g kg⁻¹ for SOC content, and 0.15 and 0.21 g cm⁻³ for BD. Best predictions were obtained for layers between 15 and 60 cm of depth. The SOC stocks were calculated over a depth of 105 cm by combining the prediction of SOC content and BD. The final maps show that the carbon stocks in the soil below 30 cm accounted for 33% of the total SOC stocks on average. The whole method produced consistent results between the two predicted soil properties. The final SOC stocks maps provide continuous data along soil profile up to 105 cm, which may be critical information for supporting carbon policy and management decisions.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Soil organic carbon (SOC), identified as an important pool in the global carbon cycle (Grace, 2004), is a key variable in influencing many agronomical, environmental and political issues. SOC directly controls soil fertility and soil structure, and small changes in SOC can have a large impact on the atmospheric carbon dioxide concentrations (Johnston et al., 2004). SOC assessment and dynamics have been studied at different scales, from field to global scale. The choice of the scale depends on the problem to deal with (impact of cropping practices and spreading, impact of land management and land use, setting up of

agricultural policies, etc.), but it has been argued that the most relevant scale to address the issue for the environment and agroecosystem issues was the landscape scale (Viaud et al., 2010). This scale enables to take into account the interaction between SOC dynamics, natural and anthropogenic processes, for instance the relation between SOC storage and soil erosion. In most studies on SOC assessment and modeling, only the topsoil is considered, i.e. the soil layer up to 30 cm and more rarely to 50 cm (e.g. Mishra et al., 2010; Ungaro et al., 2010; Zhang et al., 2010). However, it has been shown that an important amount of carbon is also stored at greater depth (Meersmans et al., 2009). It appears therefore necessary to produce SOC data not only for the topsoil, but also for the underlying layers. Previous studies dealing with SOC prediction in depth (up to 1 m or more) applied depth functions (Meersmans et al., 2009), mainly combined with geostatistical methods (mostly kriging; Kempen et al., 2011; Mishra et al., 2009; Vasques et al., 2010)

* Corresponding author. Tel.: +33 223487047; fax: +33 223 48 54 30.
E-mail address: Marine.Lacoste@rennes.inra.fr (M. Lacoste).

or with machine learning methods such as neural networks (Malone et al., 2009; Minasny and McBratney, 2006). In these studies, SOC stocks are always calculated by multiplication of SOC content and soil bulk density (BD), and the BD is always estimated by pedotransfer functions. Statistical methods have several advantages, such as simplicity and ease of interpretation. Moreover, geostatistical methods unequivocally take into account spatial correlation between soil observations. However, they also present disadvantages: they require fairly dense soil datasets and the effects of the different drivers of SOC dynamics are hardly included in statistical models. Machine learning methods allow using either limited or large datasets. Machine learning methods offer useful tools to predict soil properties from spatially exhaustive environmental predictors and limited or large soil datasets by describing without preconceived ideas the complex relationship between soil properties and covariates known as *scorpan* factors. These *scorpan* factors were first formalized by Jenny (1941), latter by McBratney et al. (2003) and comprise climatic, biological, topographic, geological, temporal and spatial factors. However, depending on the tool used for machine learning, these relationships may be quite difficult to interpret. Moreover, a major limitation of the learning methods is the computation of an error map. However, different methods have been proposed to produce such a map (e.g. Lemerrier et al., 2011; Malone et al., 2011; Solomatine and Shrestha, 2009).

In this study, we aim to map SOC stocks in a heterogeneous agricultural landscape, with a high resolution (grid spacing of 2 m), and to quantify the uncertainties of the final 3D map. This fine resolution is required to take into account the linear elements of the landscape (e.g. hedges, mainly on bank, and ditches). To achieve this goal, we propose to map both SOC content (g kg^{-1}) and soil BD (g cm^{-3}) in 3D (i.e. from soil surface to a depth up to 105 cm). These maps are then combined to produce a map of SOC stocks. The following steps were achieved: (i) selection of soil sampling locations, (ii) fitting of equal-area spline functions to SOC content and soil BD data to obtain continuous soil profiles, and extraction of soil properties for eight standard soil layers, to be used as calibration and validation datasets, (iii) creation of 2D rule-based models based on spatially exhaustive environmental covariates, to predict SOC content and BD for the eight standard soil layers from the surface to a depth of 105 cm, (iv) reconstruction of continuous soil profiles of SOC content and soil BD using a spline function, (v) production of prediction interval established by fuzzy clustering for each predictive map of SOC content and soil BD (one map per standard layer), (vi) mapping of SOC stocks (tC ha^{-1}) and the associate uncertainties from the soil surface and up to 105 cm by combining SOC content and soil BD maps.

2. Material and methods

2.1. Study area

The study area (NW France, $48^{\circ} 36' \text{ N}$, $1^{\circ} 32' \text{ W}$), was part of the European Long-Term Ecosystem Research Network, and covers an area of 10 km^2 (Fig. 1). Landscape scale was considered here as an intermediate integration level between the field and the physiographic region (Burel and Baudry, 2003; Turner, 1989). Its exact extent, however, depends on the ecological process under study and on the spatial range of the biophysical and anthropogenic processes driving it (Forman, 1995). In Western Europe, considering the spatial scale for the biophysical and pedological processes controlling soil organic matter dynamics and given the average size of farms, the extent of the landscape can be considered to range from 1 to 100 km^2 and would correspond to small watersheds (Viaud et al., 2010). Our study area was characterized by high soil heterogeneity at short distances. The soil types encountered are mainly Cambisols and Luvisols, but also Leptosols and Fluvisols (IUSS Working Group WRB, 2007). These soils present varying redoximorphic conditions, soil and A-horizon thickness (for 17 to 550 cm and from 3 to 250 cm, respectively), and soil parent material (granite, hard and soft schist with a heterogenic cover of

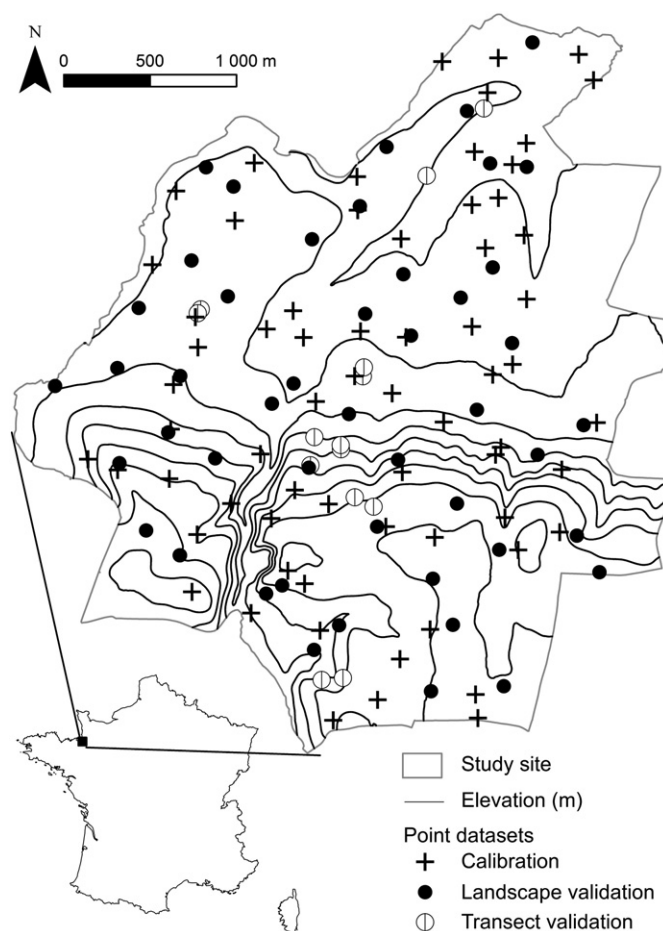


Fig. 1. Location of the study area in France and the soil sampling locations within the study area.

superficial deposits such as Aeolian loam, alluvium and colluvium). The general topography is highly correlated with the geological formations: plateau on granite (south of the site), plain on soft schist (north of the site), hillside on hard schist at the transition between granite and soft schist. However, the study site presents a marked micro-topography, mainly due to the presence of hedges (often on banks) and ditches at the field boundaries. The hedge density differs from less than 90 to 200 m ha^{-1} according to past land consolidation initiatives. Land use comprises mainly annual crops (corn, wheat, barley) and temporary or permanent grasslands, but the study area also includes wood plots and natural areas.

2.2. Soil data

We used three independent datasets based on the following sampling strategy:

- (i) Calibration dataset: designed to calibrate the predictive models. We selected 70 sampling locations using conditioned Latin hypercube sampling (cLHS), a sampling method described as efficient for calibration purposes (Lagacherie et al., 2007; Minasny and McBratney, 2006). This method is a stratified random procedure that provides an efficient way of sampling variables from their multivariate distributions (Minasny and McBratney, 2006). cLHS allows selecting locations with a good representation of the study area according to environmental covariates. As covariates we used elevation, a topographic wetness index (Eq. (1); Beven and Kirkby, 1979;

Merot et al., 2003), natural gamma emission of potassium and grassland frequency over 15 years (from 1993 to 2007).

$$wi = \ln\left(\frac{A_{sm}}{DG}\right) \quad (1)$$

where wi is the topographic wetness index, A_{sm} the specific multidirectional upslope contributing area and DG represents the downslope gradient.

Fig. 2 shows that the distribution of the variables resulting from cLHS sampling was similar to their distribution over the whole study area.

- (ii) Landscape validation dataset: designed to assess the models performance at the landscape scale. We selected 49 sampling locations by using stratified random sampling, a sampling method efficient for validation issues (Brus et al., 2011). The strata were defined by a 300-m square grid, so all strata have an equal area of 9 ha. In each stratum, one sampling location was selected by random sampling.
- (iii) Transects validation dataset: designed to assess the models' performance at the hedge vicinity. We selected 14 hedges by expert knowledge to represent the whole study area by taking into account soil parent material, A-horizon and soil thickness, field size and local slope up to hedges. All selected hedges were perpendicular to the slope and 8 sampling locations were defined along each transect: 1 site under the hedge, 4 sites upslope the hedge (2, 5, 10, and 20 m from the hedge) and 3 sites downslope the hedge (2, 5 and 10 m from the hedge). In total, 112 sampling locations were selected.

Sampling locations were placed either in agricultural or in natural areas. For the calibration and the landscape validation dataset, sampling was done using a manual root auger with a diameter of 8 cm; for the transect validation dataset, it was done with a 7.5 cm diameter mechanical auger. Both methods allow sampling soil core (continuous soil profiles) up to 105 cm. Each core was cut to provide undisturbed soil increment of known volume. Two soil increments were consistently sampled (0 to 7.5 cm and 7.5 to 15 cm), and then the soil cores have been cut into 15 cm soil increments until the bottom limit of the A horizons. When an A-horizon limit was found inside a 15 cm soil increment, this increment was cut again on this limit to provide soil samples belonging to only one A-horizon. The A-horizons were distinguished in the field from other soil horizons by color observation (Munsell value and chroma < 5 on wet samples). For the calibration and landscape validation dataset, SOC content (g kg^{-1}) was determined by dry combustion (Thermo scientific Finnigan EA 1112 Flash elemental analyzer). For the transect validation dataset, SOC content was estimated using diffuse reflectance spectroscopy within the visible and near-infrared spectral ranges, combined with a partial least squares regression model (Aïchi et al., 2009). Initially, this model has been calibrated and validated on a regional database (SOC content analysed by dry combustion), for a spatial area where our study area where included. For this study, this model has been re-calibrated by adding the calibration dataset that we used in this study (SOC content from dry combustion) to the regional database. BD of the fine earth was measured (g cm^{-3}) for all the soil datasets (calibration and validation datasets): soil samples with a known volume were dried at 105°C and sieved at 2 mm; the resulting fine earth and gravels were weighed, and the volume of

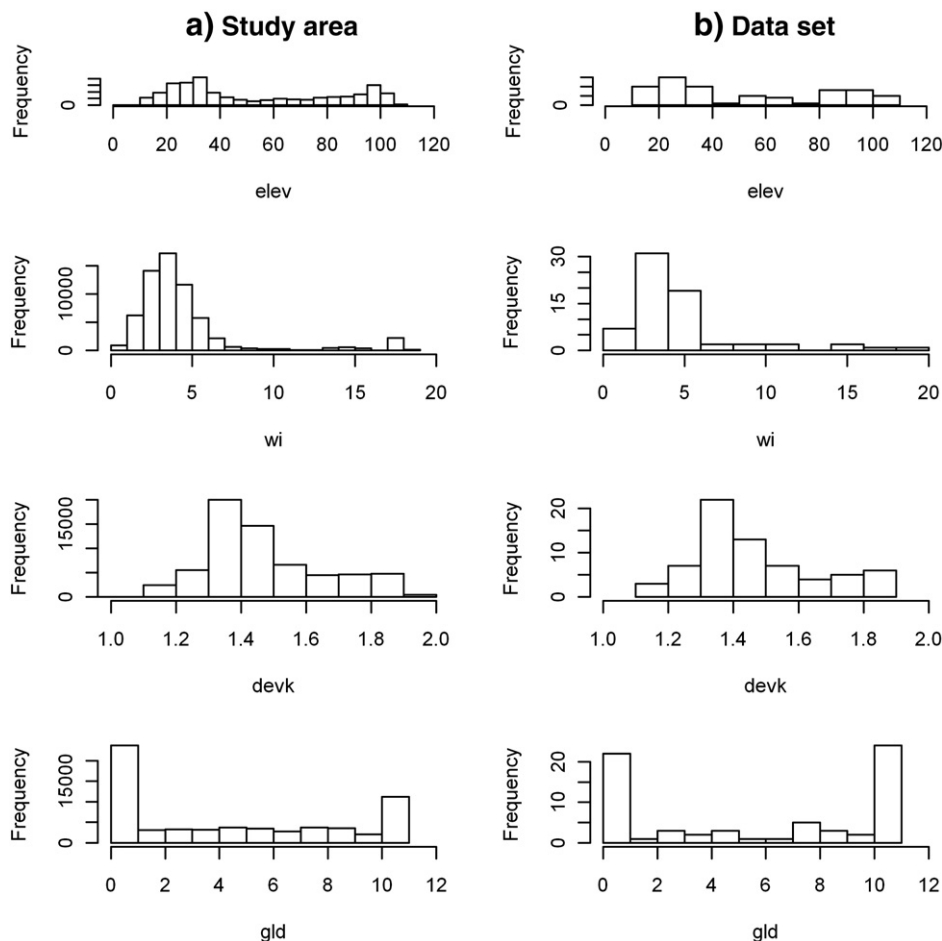


Fig. 2. Comparison of the distribution of the environmental data used for the cLHS (a) entire study area, b) calibration dataset, see Table 1 for the covariates names and units).

the gravels was measured. Soil sampling and measurements were done between May 2009 and October 2010.

2.3. Environmental covariates

Nineteen environmental variables were selected to be used as covariates in the SOC content and BD models (Table 1). Eleven covariates were derived from a 2 m LiDAR DEM (light detection and ranging digital elevation model): elevation, difference of elevation between a 20 m and a 5 m DEM (the 5 m and 20 m DEM were resampled from the 2 m LiDAR DEM), local hillslope gradient, profile and tangential curvatures, elevation above the nearest stream, topographic wetness index (Eq. (1), Beven and Kirkby, 1979; Merot et al., 2003), horizontal and vertical overland flow distances (Freeman, 1991; O'Callaghan and Mark, 1984), multiresolution indexes of ridge top and valley bottom flatness (Gallant and Dowling, 2003).

These attributes have been computed using the open source GIS (Geographical Information System) SAGA v.2.0.5 (System for Automated Geoscientific Analyses, <http://www.saga-gis.org/en/index.html>). All terrain attributes were originally derived from a 2 m LiDAR DEM, but such DEM shows locally noisy variations, which could lead to irrelevant terrain attributes. To deal with this issue, the terrain attributes were derived from DEM resampled at 5, 10 and 20 m from the 2 m LiDAR DEM and then used as covariates. The difference of elevation between a 20 m and a 5 m DEM was considered to identify local soil deposit or erosion. The geology of the study area was described by two maps: a 1:50,000 map of lithological units (Chantraine et al., 2002) and a predictive map of soil parent material (50 m grid spacing; Lacoste et al., 2011). This predictive map includes superficial deposits, i.e. Aeolian loam, colluvial and alluvial deposits. Gamma-ray spectrometry was based on existing airborne geophysical surveys interpolated by ordinary kriging at 250 m resolution (Bonijoly et al., 1999). Deviation from mean K emission of the related lithological unit, defined using the exhaustive lithological map produced by Chantraine et al. (2002), was considered in order to detect thick covers of superficial deposits, especially Aeolian loam. Land use and landscape were described by four attributes: grassland

and forest frequencies over 15 years (from 1993 to 2007), distances to the nearest hedge (uphill and downhill from the hedge). Finally, we used one attribute to describe the soil: the A-horizon thickness, provided by a predictive map produced by Lacoste (2012) with data collected at the same time of the one used to predict SOC stocks and soil BD, using the Cubist algorithm Quinlan (1992, 1993, 1994) to create a predictive model based on the same covariates as described in the Table 1.

2.4. Prediction and mapping methods

The method used to predict and map the total SOC stocks up to 105 cm involved six main steps (Fig. 3):

i) Preparation of the calibration and validation datasets

The first step achieved depth standardization of the soil profile observations provided by soil core sampling. Continuous SOC content and BD profiles (from the soil surface to a maximum depth of 105 cm) were produced by fitting equal-area spline functions to the soil observations. The equal-area spline function that we used was a generalization of the quadratic equal-area spline model of Bishop et al. (1999) and was described by Malone et al. (2009). We fitted the spline function to the measured values for each soil property (SOC content and BD). The spline function depends on a smoothing parameter λ . For this parameter λ , we tested the following values: 10, 1, 0.1, 0.01, 0.001, 0.0001 and 0.00001 (the higher the value λ , the smoother is the fitted profile). For each profile, we recorded the value λ that produced the fitted profile with the highest fit quality, according to the root mean square error (RMSE). For each variable (SOC content and BD), we chose the λ value that occurred most frequently among profiles. Eight standard soil layers were defined: 0 to 7.5 cm, 7.5 to 15 cm, 15 to 30 cm, 30 to 45 cm, 45 to 60 cm, 60 to 75 cm, 75 to 90 cm, and 90 to 105 cm. The two first layers were described to consider rapid SOC content and BD changes with depth in grasslands or in fields with minimum tillage

Table 1
Environmental covariates used to predict SOC and BD.

Name	Description (units)	Soil forming factor	Type	Mean	(Minimum, maximum)
Terrain attributes derived from the DEM					
<i>elev</i>	Elevation (m)	r	Q	52.5	(11, 110)
<i>dif.elev</i>	Difference between the elevation from the 20 m DEM and the 5 m DEM (m)	r	Q	0	(−14, 14)
<i>slope</i>	Local hillslope gradient (%)	r	Q	3.6	(0, 71)
<i>vcurv</i>	Profile (vertical) curvature ($m\ 100\ m^{-1}$)	r	Q	0	(−61, 61)
<i>hcurv</i>	Tangential (horizontal) curvature ($m\ 100\ m^{-1}$)	r	Q	0	(−44, 39)
<i>zc</i>	Altitude above the channel network (m)	r	Q	3.1	(0, 35)
<i>wi</i>	Modified topographic wetness index	r	Q	10.2	(3, 15)
<i>hof</i>	Horizontal overland flow distance (m)	r	Q	138.8	(1, 890)
<i>vof</i>	Vertical overland flow distance (m)	r	Q	7	(0, 56)
<i>mrrtf</i>	Multiresolution index of ridge top flatness	r	Q	1.2	(0, 6)
<i>mrxbf</i>	Multiresolution index of valley bottom flatness	r	Q	1.5	(0, 6)
Geological data					
<i>spm</i>	Predictive map of soil parent material (50 m grid)	p	C	14 classes	
<i>geol</i>	Bedrock lithological units (1/50 000)	p	C	4 classes	
Gamma-ray spectrometry (airborne geophysical survey)					
<i>devk</i>	Deviation from mean K emissions of The corresponding lithogeological unit (Bonijoly et al., 1999)	p	Q	0	(−0.5, 0.5)
Land use/landscape					
<i>gld</i>	Grassland frequency on 15 years	o	Q	6	(0, 15)
<i>frt</i>	Forest frequency on 15 years	o	Q	1	(0, 15)
<i>lowh</i>	Distance to the lowest hedge (m)	o	Q	24.9	(0, 265)
<i>highh</i>	Distance to the highest hedge (m)	o	Q	29.6	(0, 307)
Soil data					
<i>hza</i>	Predictive map of the A horizon (cm)	s	Q	48.4	(3, 241)

r = relief, p = parent material, s = soil properties, o = organisms, Q = quantitative, C = categorical.

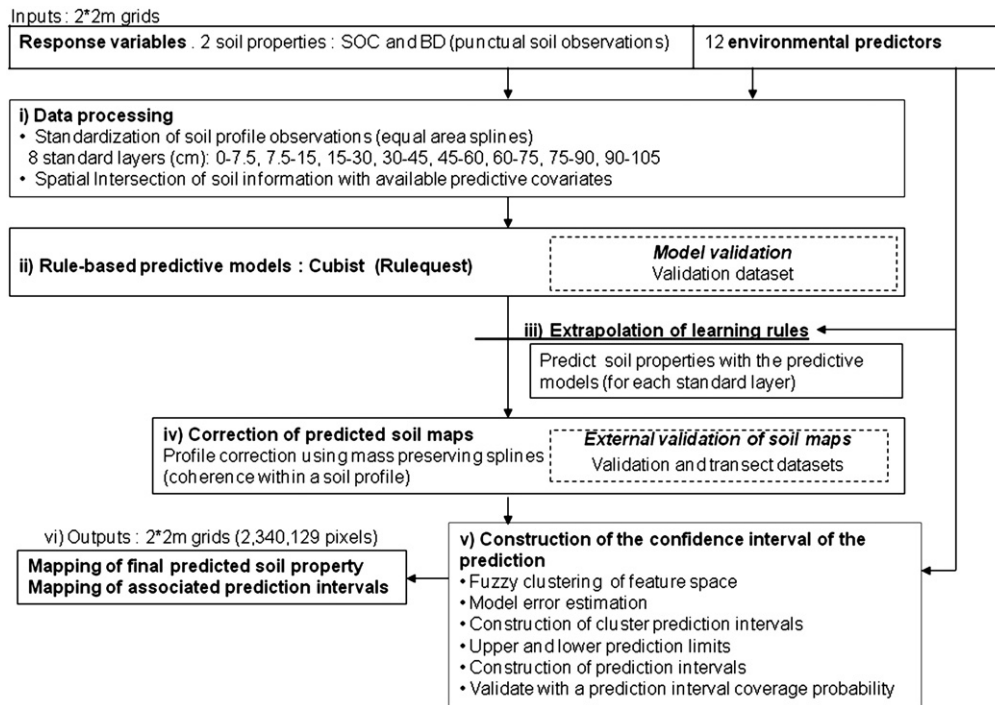


Fig. 3. Overview diagram of the method implemented to produce 3D continuous predictive maps of SOC content and BD at the landscape scale.

(Follain et al., 2007; Viaud et al., 2010). From the fitted spline, we derived the mean value of SOC content and BD within the 8 defined depth increments (i.e. the eight standard soil layer). These fitted values were used as calibration or validation data, and spatially intersected with the available spatially exhaustive environmental covariates.

- ii) Creation of predictive models based on the calibration dataset and the spatially exhaustive environmental covariates. Models were created separately both for SOC content and BD, and for each standard soil layers. We used Cubist v.2.04, a commercial software tool distributed by Rulequest© (<http://www.rulequest.com/>) and based on the tree model algorithm M5 developed by Quinlan (1992, 1993, 1994). Cubist is a rule-based regression algorithm, allowing the creation of a predictive model based on conditions rules (partition of the calibration space according to the associate environmental covariates) and linear predictive rules (prediction of the variable of interest by combining the environmental covariates values). Cubist has been used in various fields with good results and is recognized as an efficient tool in term of computation time (Walton, 2008). Cubist creates an explicit model and gives the relative importance of the predictors, allowing an easy model interpretation.

iii) Models performance

First, the models constructed by Cubist were evaluated on the calibration dataset, from which they were generated (internal evaluation). Second, the models were evaluated on the two independent datasets not used to create the models, taken separately (external evaluation on the landscape and transect validation datasets) and together. Four different error indexes were calculated both for the internal and external evaluations: (i) the Mean Absolute Error (MAE, Eq. (2)), (ii) the Root Mean Square Error, which is a measurement of the average error (or accuracy) of the prediction (RMSE, Eq. (3)), (iii) the coefficient of determination, which measures the agreement between measured and predicted data (R^2 , Eq. (4)), and (iv) Lins's Concordance Correlation Coefficient (ρ_c , Eq. (4)) which evaluates the degree to which pairs of observations fall on the 45° line through the origin

(Eq. (5); Lin, 1989, 2000). The value of this index is scaled between -1 and 1, with a value of 1 representing complete agreement between all paired sites.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - x_i| \quad (2)$$

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

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

$$\rho_c = \frac{2S_{xy}}{S_x^2 + S_y^2 + (\bar{x} - \bar{y})^2} \quad (5)$$

where $S_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$, $S_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$, $S_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$.

For all these four equations, $\bar{x}$ and $\bar{y}$ are the mean value for populations x (measured) and y (predicted) of size n , x_i and y_i are paired i th values from the population x and y .

- iv) Production of spatially exhaustive maps of SOC content and BD. Predictive maps of SOC content and BD have been produced separately for every standard soil layer by applying the learning rules created by Cubist to the entire study area (2,340,129 pixels), where spatially exhaustive covariates were available. For each of the standard soil layer (from the soil surface to a depth of 105 cm), the predictive models have been applied for every pixels of the study area.
- v) Reconstruction of continuous soil profiles from the predictive SOC content and BD values and depth correction. Continuous SOC content and BD profiles (from the soil surface to a depth of 105 cm) were produced for each pixel of the study area, by fitting equal-area spline functions to the SOC content and BD values predicted by the Cubist learning rules. This profile reconstruction step allowed obtaining smoothed continuous soil

profiles, used to create continuous 3D maps. The influence of the soil profile reconstruction on soil properties values was assessed by comparing the values of the four error indexes previously described calculated from (i) the predicted values for the standard soil layers (without soil profile reconstruction) and all the measured values from the two validation datasets considered together, and (ii) the new predicted values extracted from the continuous soil profile for the standard soil layers and the two validation datasets considered together. The coherence between SOC content profiles and A-horizon thickness was controlled using a predictive map of A-horizon thickness available for the study area (Lacoste, 2012). It allowed correcting the continuous soil profiles by removing SOC content and BD values where SOC contents values were predicted under the A-horizon limits.

vi) Mapping of SOC stocks

SOC stocks were computed by combining predictions of SOC content, soil BD and soil gravel content. Predictions of soil gravel content were produced with the same field datasets and covariates which we used to produce the predictive maps of SOC contents and soil BD (Lacoste, 2012). For a soil layer i , the SOC stock are given by Eq. (6):

$$SOC_{stock,i} = (SOC_{content,i} \times BD_i \times (1 - Gr_i/100) \times t_i)/10 \quad (6)$$

where $SOC_{stock,i}$ is the SOC stock in $tC\ ha^{-1}$, $SOC_{content,i}$ the SOC content in $g\ kg^{-1}$, BD_i the fine earth BD in $g\ cm^{-3}$, Gr_i the volume percentage of gravel (%) and t_i the thickness of the depth increment i (cm).

For every pixel in the study area, total SOC stocks (0 to 105 cm) were estimated by summation of the stocks for all the intermediate soil increments.

vii) Construction of the prediction interval of predictive maps

The prediction intervals were calculated for each standard soil layer, using the method fully described by Malone et al. (2011). This method presented the major advantage to not require the identification and the determination of the contribution of each source of uncertainty to the overall prediction uncertainty. It was based on the assumption that pixels in the same feature space (determined by the covariates values) will share similar distribution of errors. This method was implemented as follows for each standard soil layer: (i) clustering of the feature space of the environmental covariates associated to the two validation datasets, using a fuzzy k-means routine with extragrades; (ii) determination of the error distribution model for each cluster, based on the residuals determined by comparing the values of predicted and measured soil properties; (iii) for each cluster, calculation of prediction intervals with a 95% confidence level (95% of the real soil properties values should fall within these bounds); (iv) partition of the whole study area according to the defined clusters; and (v) computation of the prediction interval for each pixel using the error distribution models defined for each cluster. The prediction interval obtained on each standard soil layers were combined to produce a final prediction interval until a depth of 105 cm.

3. Results

3.1. Models description

3.1.1. Example of one model used to predict SOC content: the case of the 30-to-45 cm standard soil layer

We obtained an explicit model for each soil property and for each standard soil layer. Fig. 4 gives an example of the rules obtained to predict SOC content for the 30-to-45 cm standard soil layer. This particular model included four linear predictive rules (introduced by “then”), applied to different partitions of the study area determined by condition rules (introduced by “if”).

```

Rule 1: [112 cases, mean 8.112288, range 1.116017 to 20.19062, est err 2.3599148]
  If
    Gld in {1, 2, 3, 4, 5, 6, 8, 10, 11, 12, 13, 14, 15}
  Then
    SOC = 8.1774597 - 0.78 mrvbf + 0.03 hza + 1.5 dif.elev
Rule 2: [118 cases, mean 8.9629173, range 1.116017 to 16.52321, est err 2.1300650]
  If
    Smp in {1, 2, 3}
  Then
    SOC = 3.8902728 + 0.0084 hza
Rule 3: [7 cases, mean 12.0846968, range 6.084407 to 20.19062, est err 6.4643254]
  If
    Smp in {6,14}
    High <= 6.929646
  Then
    SOC = 9.8467357
Rule 4: [12 cases, mean 20.6772995, range 6.774977 to 31.96915, est err 2.8409903]
  If
    Smp in {6, 14}
    Gld in {0, 9}
    High > 6.929646
  Then
    SOC = 27.0501001 - 0.212 highh - 3.49 vcurv - 0.018 lowh

```

Fig. 4. Example of the Cubist model used for the prediction of SOC content in the 30-to-45-cm soil layer.

3.1.2. Fitting performance evaluation: internal and external validations

Tables 2 and 3 show the results of the fitting performance for the models used to predict SOC content and BD, respectively. The four errors indexes described in Section 2.4 are given for each standard soil layer (allowing an estimation of the model performance according to the considered depth) and for the soil standard layers all together (allowing an overall assessment of the model performance). Considering the evaluation on the calibration dataset (internal validation), the SOC content models showed an overall MAE of $2.31\ g\ kg^{-1}$, RMSE of $4.05\ g\ kg^{-1}$, R^2 of 0.80 and ρ_c of 0.88 (Table 2). For the evaluation on the landscape validation dataset (external validation), these models showed an overall MAE of $4.08\ g\ kg^{-1}$, RMSE of $7.74\ g\ kg^{-1}$, R^2 of 0.31 and ρ_c of 0.52. Finally, for the evaluation on the transect validation dataset (external validation), these models showed an overall MAE of $3.07\ g\ kg^{-1}$, RMSE of $5.02\ g\ kg^{-1}$, R^2 of 0.71 and ρ_c of 0.82. When considering together the two datasets used for external validation (landscape and transect validation datasets), these SOC content models showed an overall MAE of $3.35\ g\ kg^{-1}$, RMSE of $5.92\ g\ kg^{-1}$, R^2 of 0.60 and ρ_c of 0.73. Considering the four error indexes and the three evaluation datasets, the best predictions of SOC content were obtained for the middle of the soil profiles (depth between 15 and 75 cm), and the worst prediction in the deeper horizons (depth of 75 cm and below).

Considering the evaluation on the calibration dataset (internal validation), the BD models showed an overall MAE of $0.07\ g\ cm^{-3}$, RMSE of $0.09\ g\ cm^{-3}$, R^2 of 0.77 and ρ_c of 0.86 (Table 3). For the evaluation on the landscape validation dataset (external validation), these models showed an overall MAE of $0.11\ g\ cm^{-3}$, RMSE of $0.15\ g\ cm^{-3}$, R^2 of 0.26 and ρ_c of 0.49. Finally, for the evaluation on the transect validation dataset (external validation), these models showed an overall MAE of $0.15\ g\ cm^{-3}$, RMSE of $0.21\ g\ cm^{-3}$, R^2 of 0.23 and ρ_c of 0.49. When considering together the two datasets used for external validation (landscape and transect validation datasets), these BD models showed an overall MAE of $0.14\ g\ cm^{-3}$, RMSE of $0.19\ g\ cm^{-3}$, R^2 of 0.23 and ρ_c of 0.38. Considering the influence of the depth on the BD models performance, errors indexes showed in Table 3 didn't allowed determining a clear distribution pattern, except for the evaluation on the transect validation dataset which has higher values of MAE and RMSE for the very topsoil layer (0 to 7.5 cm) and for the deeper soil layers (60 to 95 cm).

Table 4 shows the SOC content and BD models performances assessed using the transect validation dataset, according to the distance

Table 2
Fitting performance for the SOC content predictive models.

Standard soil layers (cm)		0–7.5	7.5–15	15–30	30–45	45–60	60–75	75–90	90–105	All layers
Evaluation the calibration dataset	MAE (g kg ⁻¹)	3.02	4.06	2.12	1.29	0.73	1.01	7.49	0.94	2.31
	RMSE (g kg ⁻¹)	5.35	6.11	3.06	1.63	0.90	1.60	7.97	0.94	4.05
	R ²	0.79	0.38	0.55	0.73	0.85	0.46	0.71	0.11	0.80
	ρ _c	0.87	0.47	0.72	0.80	0.92	0.65	0.26	0.10	0.88
Evaluation the validation dataset	MAE (g kg ⁻¹)	6.85	5.07	2.97	2.33	2.77	1.50	13.62	7.64	4.08
	RMSE (g kg ⁻¹)	12.64	9.17	4.29	3.55	4.56	2.22	13.89	7.84	7.74
	R ²	0.12	0.01	0.03	0.05	0.08	0.51	0.29	0.25	0.31
	ρ _c	0.28	0.05	-0.15	0.21	0.14	0.55	0.03	0.16	0.52
Evaluation the transect dataset	MAE (g kg ⁻¹)	3.26	3.72	2.42	1.54	1.50	2.04	12.48	2.42	30.7
	RMSE (g kg ⁻¹)	4.60	6.06	3.98	2.16	1.93	2.97	13.19	3.23	5.02
	R ²	0.77	0.48	0.72	0.84	0.85	0.60	0.25	0.51	0.71
	ρ _c	0.84	0.55	0.82	0.91	0.92	0.72	0.09	0.58	0.82

Table 3
Fitting performance for the BD predictive models.

Standard soil layers (cm)		0–7.5	7.5–15	15–30	30–45	45–60	60–75	75–90	90–105	All layers
Evaluation on the calibration dataset	MAE (g kg ⁻¹)	0.08	0.10	0.05	0.08	0.05	0.06	0.05	0.06	0.07
	RMSE (g kg ⁻¹)	0.11	0.12	0.06	0.11	0.06	0.08	0.08	0.07	0.09
	R ²	0.79	0.56	0.85	0.32	0.74	0.66	0.80	0.66	0.77
	ρ _c	0.84	0.69	0.89	0.49	0.81	0.63	0.64	0.73	0.86
Evaluation on the validation dataset	MAE (g kg ⁻¹)	0.13	0.13	0.12	0.10	0.07	0.08	0.04	0.12	0.11
	RMSE (g kg ⁻¹)	0.17	0.18	0.16	0.13	0.10	0.13	0.05	0.17	0.15
	R ²	0.08	0.03	0.01	0.09	0.47	0.05	0.38	0.27	0.26
	ρ _c	0.21	-0.15	0.00	0.25	0.56	-0.03	0.47	-0.27	0.49
Evaluation on the transect dataset	MAE (g kg ⁻¹)	0.27	0.13	0.11	0.11	0.11	0.20	0.18	0.17	0.15
	RMSE (g kg ⁻¹)	0.31	0.17	0.15	0.16	0.17	0.26	0.22	0.21	0.21
	R ²	0.08	0.09	0.15	0.03	0.08	0.30	0.10	0.02	0.23
	ρ _c	0.02	0.05	-0.12	-0.03	0.08	0.16	0.11	-0.01	0.49

to the hedge and the depth. Considering all the standard soil layers (without depth distinction), both predictions of SOC content and BD had low RMSE variation according to the distance to the hedge: RMSE for SOC content models range from 3.11 to 4.36 g kg⁻¹, and RMSE for BD models range from 0.12 to 0.21 g cm⁻³. Highest RMSE values were found for locations 2 m downhill from the hedges for the SOC content models (4.36 g kg⁻¹), and for locations just under the hedge for the BD models (0.21 g cm⁻³). The evolution with depth of the performance of SOC content models and BD models slightly varied according to the distance to the hedge: the SOC content models showed a better performance in the middle of the soil profile (between 7.5 cm and 60 cm), with poorer performance for topsoil and deep layers, and the BD models showed a higher RMSE values under the hedges (2 m downhill and uphill from the hedge, and just under the hedges), especially for the very topsoil layer (0 to 7.5 cm).

3.1.3. Relative importance of the covariates in the predictive models rules

Tables 5 and 6 give the relatives importance of the covariates used to create the SOC content and BD models, respectively. They also show the resolution of the DEM and its derived terrain attributes used as covariates in the models. Considering SOC content models

(Table 5), grassland frequency (*gld*) was the most used covariate in the condition rules. This attribute was considered as a stratifier for all the standard soil layers, except the two deeper ones. Other covariates also appeared important in the condition rules, namely vertical overland flow distance (*vof*), soil parent material (*spm*) or vertical curvature (*vcurv*), but mostly for one or two standard soil layers. A larger panel of covariates was integrated in the linear predictive rules. Among the covariates exerting a high influence in the predictions were the horizontal overland flow distance (*hof*), the *slope* and the difference between the elevation from the 20 m and the 5 m DEM (*dif.elev*). BD models involved less covariate in the construction of both the condition and linear predictive rules (Table 6). The condition rules used only the vertical overland flow distance (*vof*) and the grassland frequency (*gld*) in the models for the first 3 standard soil layers (from 0 cm to 30 cm depth). In the linear predictive rules, important covariates were the elevation (*elev*), the deviation from the mean K emissions (*devk*), the topographic wetness index (*wi*, Eq. (1)), the multiresolution index of valley bottom flatness (*mrvalf*) and the vertical overland flow distance (*vof*).

For the OC content models, the terrain attributes with a 5 m resolution were mostly used in the predictive rules, except for two standard

Table 4
RMSE for SOC content and BD prediction, according to the distance to the hedge, calculated on transect validation dataset (SOC content/BD).

Distance to the hedge (m, -: downhill, +: uphill)	Standard soil layers (cm)									
	0–7.5	7.5–15	15–30	30–45	45–60	60–75	75–90	90–105	All layers	
-10	5.5/0.24	4.99/0.15	2.32/0.13	3.49/0.21	3.14/0.18	5.15/0.42	7.78/0.22	4.21/0.19	3.30/0.15	
-5	5.46/0.32	5.68/0.15	7.01/0.11	2.35/0.09	2.05/0.18	3.21/0.01	7.23/-	5/-	3.50/0.14	
-2	11.24/0.37	9.06/0.2	7.7/0.13	2.87/0.13	3.16/0.11	4.76/0.17	0.79/-	2.24/-	4.36/0.15	
0	6.29/0.39	5.18/0.2	2.92/0.25	2.53/0.21	2.51/0.22	1.28/0.3	2.82/0.34	2.47/0.29	2.77/0.21	
2	4.47/0.38	4.6/0.17	4.23/0.12	1.53/0.09	2.39/0.14	3.74/0.21	7.64/0.22	5.72/0.21	3.24/0.15	
5	5.18/0.32	4.09/0.09	4.74/0.06	1.34/0.08	1.59/0.1	3.61/0.18	6.83/0.23	4.42/0.19	3.00/0.12	
10	4.13/0.26	3.02/0.13	2.64/0.1	2.72/0.16	2.21/0.18	4.44/0.27	9.34/0.12	4.57/0.12	2.91/0.14	
20	3.43/0.29	2.89/0.12	2.38/0.1	2.47/0.15	2.98/0.18	6.12/0.22	9.33/0.15	4.37/0.18	3.11/0.14	

Table 5

Relative importance (%) of the environmental covariates used in the SOC content predictive models' rules and resolution of the DEM and its derived attributes used as covariates (see Table 1 for the covariates' names and units).

Model properties	Standard soil layers								All layers
	0–7.5	7.5–15	15–30	30–45	45–60	60–75	75–90	90–105	
Relative importance of covariates used in condition rules (%)									
<i>gld</i>	100	27	100	50	100	23	–	–	50
<i>vof</i>	–	18	–	–	–	–	100	–	15
<i>spm</i>	–	–	–	55	–	–	–	–	7
<i>hof</i>	–	–	–	–	47	–	–	–	6
<i>vcurv</i>	–	–	–	–	41	4	–	–	6
<i>highh</i>	–	–	–	8	31	–	–	–	5
<i>slope</i>	–	–	30	–	–	–	–	–	4
<i>hcurv</i>	–	–	–	–	–	4	–	–	1
Relative importance of covariates used in linear predictive rules (%)									
<i>hof</i>	100	91	84	–	60	–	95	20	56
<i>slope</i>	–	16	86	–	31	19	95	20	33
<i>dif.elev</i>	59	54	100	45	–	19	5	–	35
<i>hza</i>	41	–	30	92	–	58	–	–	28
<i>vof</i>	41	55	–	–	–	11	100	–	26
<i>mrrtf</i>	59	–	–	–	–	–	95	40	24
<i>elev</i>	59	–	–	–	59	38	–	20	22
<i>hcurv</i>	–	54	70	–	–	49	–	–	22
<i>Wi</i>	–	19	100	–	19	–	–	–	17
<i>vcurv</i>	–	–	–	5	59	2	–	60	16
<i>zc</i>	–	53	–	–	–	–	–	60	14
<i>highh</i>	–	20	30	5	18	–	–	20	12
<i>lowh</i>	–	19	5	59	3	–	–	11	–
<i>mrvmf</i>	41	–	–	45	–	–	–	–	11
<i>devk</i>	–	–	–	–	–	30	–	–	4
Resolution of the LIDAR DEM and its derived attributes (m)									
	5	20	20	5	5	5	5	5	

soil layers (7.5 to 15 cm and 15 to 30 cm) where a 20 m resolution was used. For the BD models, the choice of the optimal terrain attribute resolution was less uniform: terrain attributes were used with a resolution of 10 m for the two surface layers (0 to 7.5 cm and 7.5 to 15 cm), 5 m for three standard layers (30 to 45 cm, 60 to 75 cm and 75 to 90 cm), 20 m for the three remaining standard layers (15 to 30 cm, 45 to 60 cm and 90 to 105 cm).

3.2. Soil profile reconstruction

The following values of error indexes compare the soil data not used to create the predictive model (landscape plus transect validation datasets, taking into account all the soil layers) with the SOC content and BD predicted values, calculated before and after the spline reconstruction process. For SOC content prediction, MAE changed with the

soil profile reconstruction from 3.35 g kg^{-1} to 3.43 g kg^{-1} , RMSE decreased from 5.92 to 5.66 g kg^{-1} , R^2 improved from 0.60 to 0.66 , and ρ_c improved from 0.73 to 0.75 . For BD prediction, MAE stayed at 0.14 g cm^{-3} with the soil profile reconstruction, RMSE stayed at 0.19 g cm^{-3} , R^2 improved from 0.23 to 0.24 , and ρ_c decreased from 0.38 to 0.34 . Fig. 5 shows that this profile reconstruction step allowed smoothing the soil profiles and removing extreme predictions, without compromising prediction accuracy.

3.3. Description of the SOC content and BD predictions

Table 7 gives the descriptive statistics of the measured SOC content and fine earth BD for the calibration, landscape and transect validation datasets, and the descriptive statistics of the predicted SOC content and fine earth BD. Measured values of SOC content ranged from 1.7 to

Table 6

Relative importance (%) of the environmental covariates used in the BD predictive models' rules and resolution of the DEM and its derived attributes used as covariates (see Table 1 for the covariates' names and units).

Model properties	0–7.5	7.5–15	15–30	30–45	45–60	60–75	75–90	90–105	All layers
Relative importance of covariates used in condition rules (%)									
<i>vof</i>	60	60	–	–	–	–	–	–	15
<i>gld</i>	–	–	20	–	–	–	–	–	3
Relative importance of covariates used in linear predictive rules (%)									
<i>elev</i>	51	–	–	60	–	40	40	–	24
<i>devk</i>	–	–	60	60	20	–	10	–	19
<i>wi</i>	51	48	–	–	40	–	–	–	17
<i>mrvmf</i>	–	–	–	60	–	–	10	60	16
<i>vof</i>	4	44	–	40	20	–	–	–	14
<i>mrrtf</i>	2	48	20	–	–	–	–	–	9
<i>zc</i>	36	–	–	–	–	20	–	–	7
<i>vcurv</i>	–	–	54	–	–	–	–	–	7
<i>hof</i>	–	32	–	20	–	–	–	–	7
<i>slope</i>	–	48	–	–	–	–	–	–	6
Resolution of the LIDAR DEM and its derived attributes (m)									
	10	10	20	5	20	5	5	20	

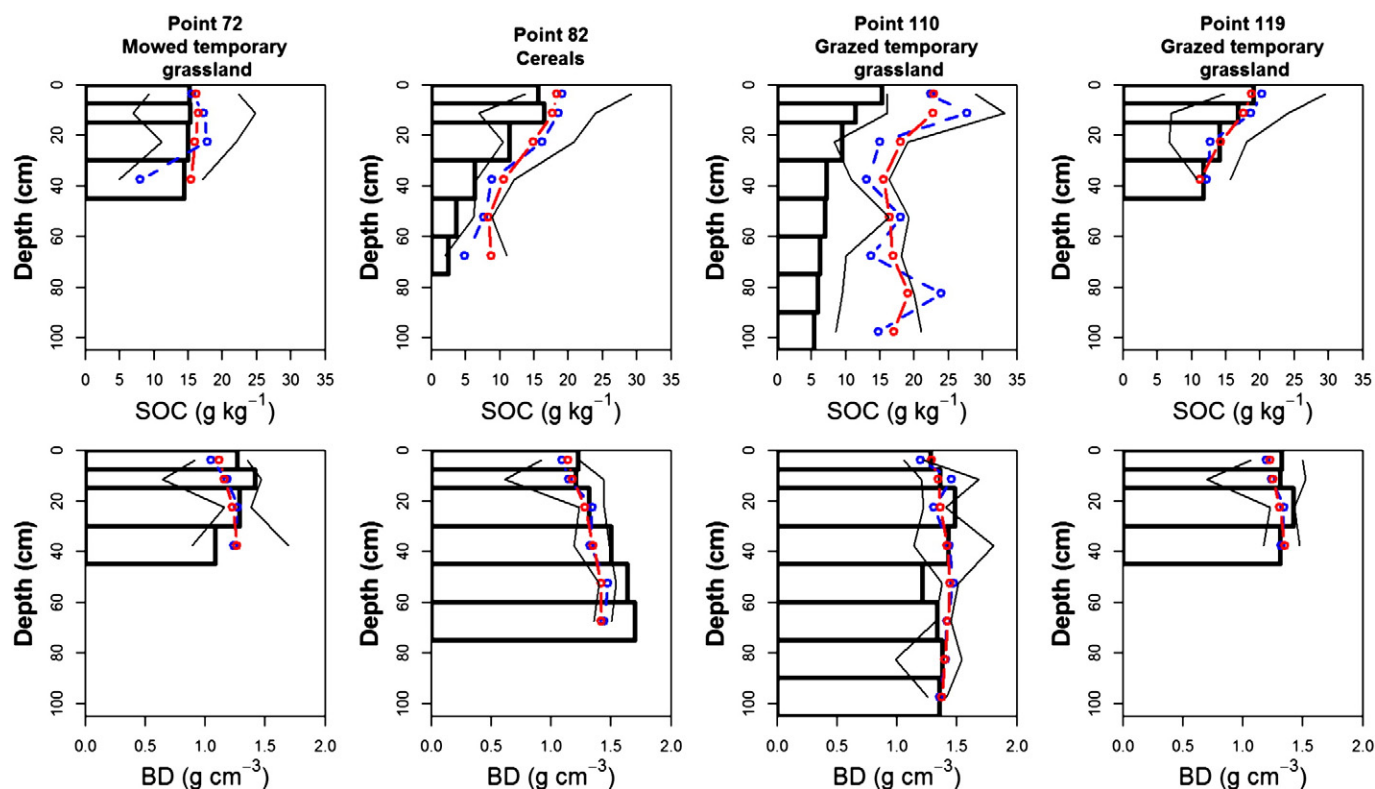


Fig. 5. Examples of soil profile reconstruction considering four randomly selected sampled soil profiles (histogram: validation data derived from measurements for the standard layers, dashed blue line: predictions from Cubist (raw values), solid red line: predictions after the reconstruction of soil profiles using the mass preserving splines, solid black lines: upper and lower prediction limits of the prediction interval).

63 g kg⁻¹ (mean value = 13 g kg⁻¹) for the calibration dataset, 0.1 to 66 g kg⁻¹ (mean value = 14 g kg⁻¹) for the landscape validation dataset, and 1.5 to 75 g kg⁻¹ (mean value = 11 g kg⁻¹) for the transect validation dataset. SOC content predictions ranged from 1 to 61 g kg⁻¹ (mean value = 11 g kg⁻¹). We observed the same trend in SOC content distribution with depth in the calibration dataset, landscape validation dataset, transect validation dataset and the predicted values of SOC content (Fig. 6). Surface horizon (0 to 7.5 cm) was proved to be on average the most concentrated horizon in SOC content. However, the deeper horizons also presented significant SOC content concentration.

Measured values of BD ranged from 0.6 to 1.7 g cm⁻³ (mean value = 1.4 g cm⁻³) for the calibration dataset, from 0.4 to 2.3 g cm⁻³ (mean value = 1.3 g cm⁻³) for the landscape validation dataset, and from 0.7 to 1.7 g cm⁻³ (mean value = 1.4 g cm⁻³) for the transect validation dataset. BD predictions ranged from 0.8 to 1.6 g cm⁻³ (mean value = 1.4 g cm⁻³). These four different datasets revealed the same general trend in BD distribution with depth (Fig. 7): BD increased

with depth until 60 cm before slightly decreasing again. The transect validation dataset showed a more dispersed distribution for the deeper layers (Fig. 7c).

Fig. 8 shows the relation between SOC content and BD, for the measured values (calibration, landscape and transect validation datasets, Fig. 8a) and for the predicted values (Fig. 8b, 1000 randomly selected points). The pattern observed in the measured data (BD decrease with an increasing SOC content) also was found in the predicted data and followed the same trend ($R^2 = 0.51$ for the measured data and $R^2 = 0.54$ for the predicted data).

3.4. Example of spatial variation for topsoil predictions

Fig. 9 shows the SOC content and BD predictions and prediction interval maps for the surface soil layer (0 to 7.5 cm). SOC content prediction for this layer ranged from 5 to 60 g kg⁻¹ (mean value = 18 g kg⁻¹), whereas prediction of BD ranged from 0.6 to 1.5 g cm⁻³ (mean

Table 7

Descriptive statistics of the SOC content and BD values for the calibration dataset, landscape and transect validation datasets, and for the predictions.

	Minimum	1st quartile	Median	Mean	3rd quartile	Maximum	Standard deviation
SOC content (g kg ⁻¹)							
Calibration dataset	1.71	6.74	11.26	12.86	15.84	63.26	9.03
Landscape validation dataset	0.04	7.96	13.05	14.30	18.93	65.53	9.22
Transect validation dataset	1.51	5.61	10.41	11.22	13.82	74.89	8.97
Predicted values	1.08	7.42	10.27	11.41	14.45	60.47	5.23
Soil BD (g cm ⁻³)							
Calibration dataset	0.60	1.25	1.39	1.35	1.48	1.70	0.19
Landscape validation dataset	0.42	1.19	1.32	1.31	1.46	2.29	0.23
Transect validation dataset	0.70	1.28	1.39	1.37	1.49	1.71	0.17
Predicted values	0.80	1.32	1.40	1.38	1.46	1.57	0.11

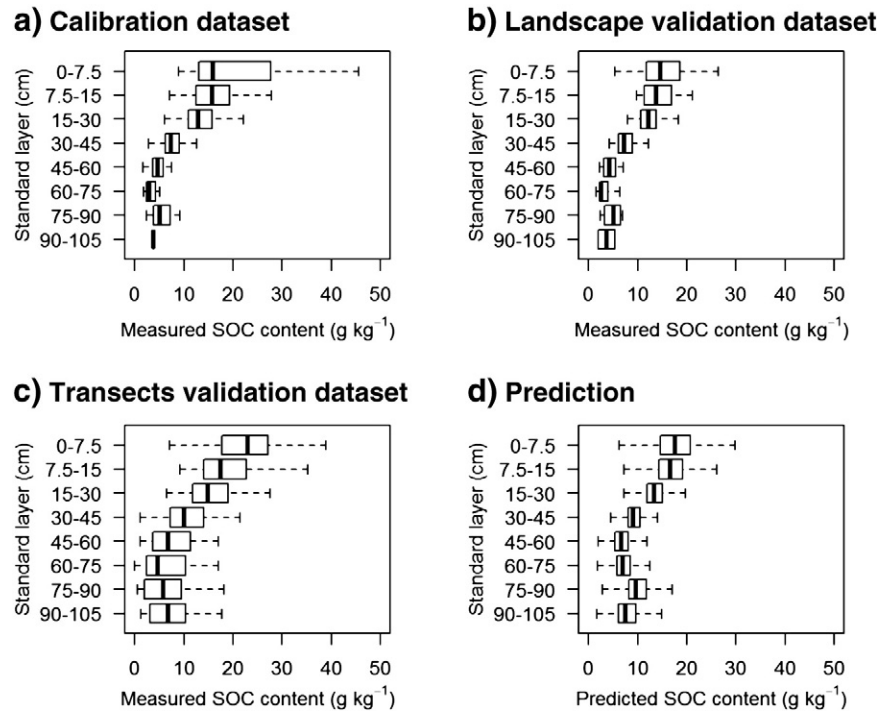


Fig. 6. Distribution of the SOC content with depth according to standard soil layers for calibration dataset (a), landscape validation dataset (b), transect validation dataset (c) and the prediction of SOC content (d).

value = 1.2 g cm⁻³). The spatial variability of SOC content prediction was mainly driven by land use and more precisely by the grassland frequency (*gld*, Table 5): higher SOC content was predicted for fields with a high frequency of grassland. Other covariates were also used to predict SOC content for this topsoil layer: the horizontal overland flow distance (*hof*), the difference between the elevation from the 20 m DEM and the 5 m DEM (*dif.elev*), the multiresolution index of ridge top flatness (*mrrtf*) and elevation (*elev*). The spatial variation of predicted BD was also driven

by the vertical overland flow distance (*vof*), the elevation (*elev*) and the wetness index (*wi*, Table 6).

3.5. Estimation of SOC stocks

Fig. 10 gives example of the spatial distribution of SOC stocks at the landscape scale. The total predicted SOC stocks (0 to 105 cm) ranged from 5 to 360 tC ha⁻¹ (mean value = 89 tC ha⁻¹, Fig. 10a). In the

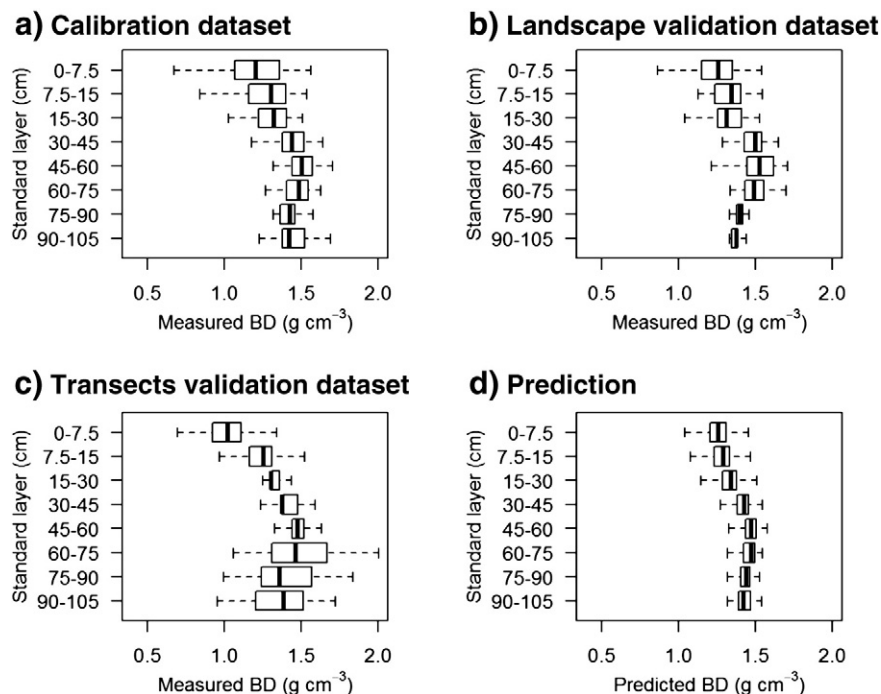


Fig. 7. Distribution of the BD with depth according to standard soil layers for calibration dataset (a), landscape validation dataset (b), transect validation dataset (c) and the prediction of BD (d).

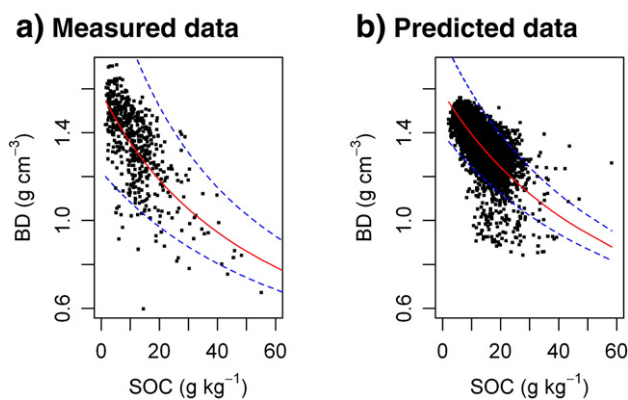


Fig. 8. Relation between BD and SOC content in measured (calibration and validation datasets) and predicted data. Solid red line = trend line $BD = 1 / (0.01 \text{ SOC} + 0.63)$, blue dashed lines = 95% confidence interval.

tillage layer (0 to 30 cm), SOC stocks ranged from 3 to 175 tC ha⁻¹ (mean value = 58 tC ha⁻¹, Fig. 10b). We estimated the mean SOC stocks for the 0-to-30 cm layer at 44 t ha⁻¹ under cultivated fields, 67 t ha⁻¹ under grasslands and 30 t ha⁻¹ under forest. SOC stocks were significant in the deeper horizons (Fig. 10c, 30–105 cm), taking values up to 85% of total SOC stocks (mean value = 32 t ha⁻¹, counting for 33% of the total SOC stocks), mostly in situation of soil accumulation uphill from hedges. Detailed maps (Fig. 10) show that SOC stocks also vary within fields, mostly according to distance to the hedges and local topography.

4. Discussion

4.1. The predictive process

4.1.1. Use of a machine learning tool

First, we choose to use a machine learning tool, Cubist, to produce a high resolution predictive map of SOC stocks by combining SOC contents and soil BD predictive maps. Machine learning was a useful method here to define the existing relationships between soil properties from a limited soil dataset and spatially exhaustive covariates. Cubist is a powerful tool where considering the computation time (model creation in less than 1 s, extrapolation to 2.3 million pixels in less than 5 min using a computer Intel® Core™2 CPU, 1.83 GHz, 2.00 GB RAM). Cubist also allows an easy interpretation of the predictive models, by the creation of useful outputs such as the detailed learning rules, or the importance of the covariates in the models. This makes more explicit the distribution of soil properties at the landscape scale.

4.1.2. Use of 2 m LiDAR DEM

The use of the DEM from LiDAR allowed taking into account the variations at short distance of the two soil properties that we predicted, which was relevant for the typical landscape of our study area. However, this 2 m DEM was proved to be too noisy, and its derived terrain attributes became irrelevant. To deal with this problem, we used as covariates the terrain attributes derived from a DEM resampled at 5 m, 10 m and 20 m. For the SOC content prediction, the best model for each standard layer was created using mostly the terrain attributes at the 5 m resolution, except for two horizons (7.5 to 15 cm and 15 to 30 cm, Table 5). This result could be explained

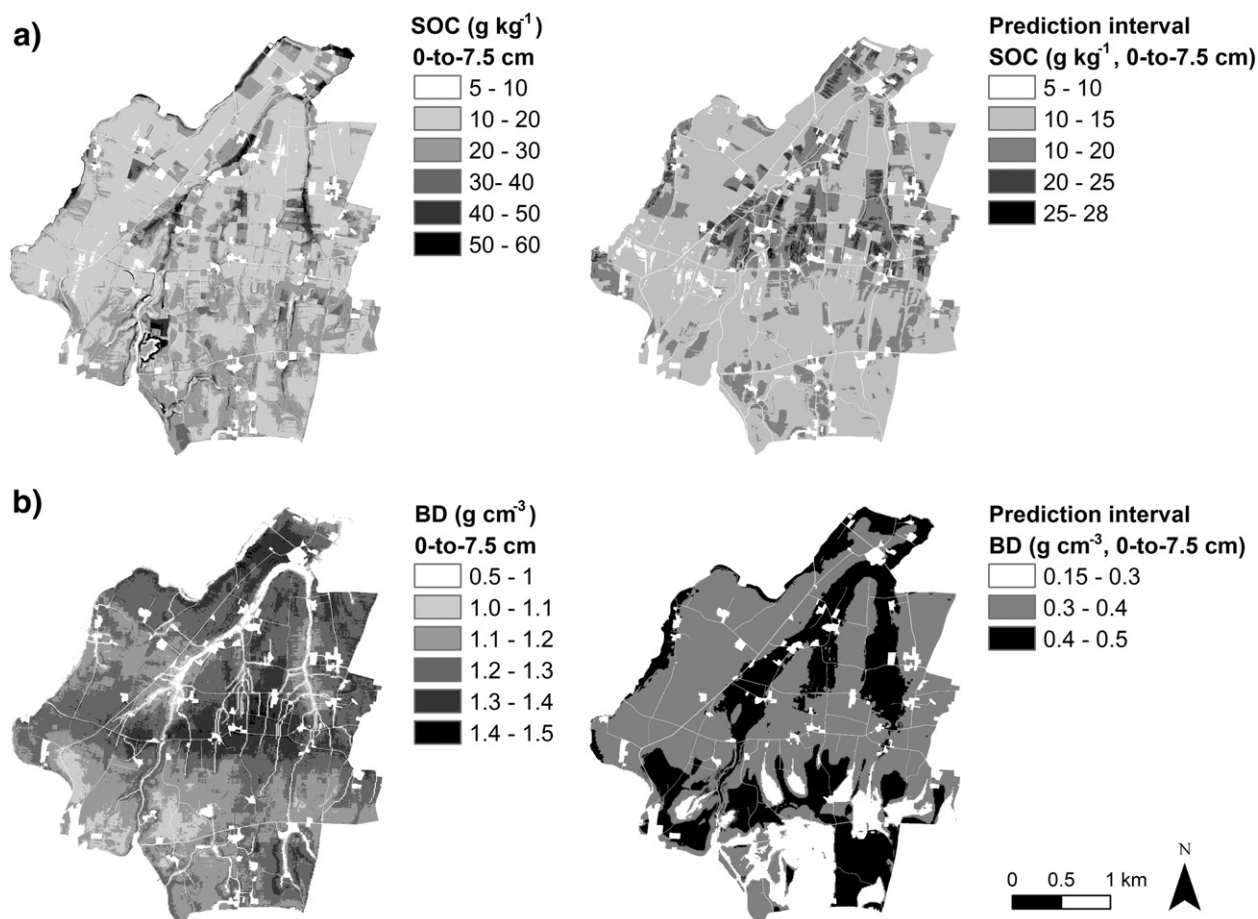


Fig. 9. Predictive maps of SOC content and BD for the 0-to-7.5 cm soil layer. (a) SOC content and its prediction interval, (b) BD and its prediction interval.

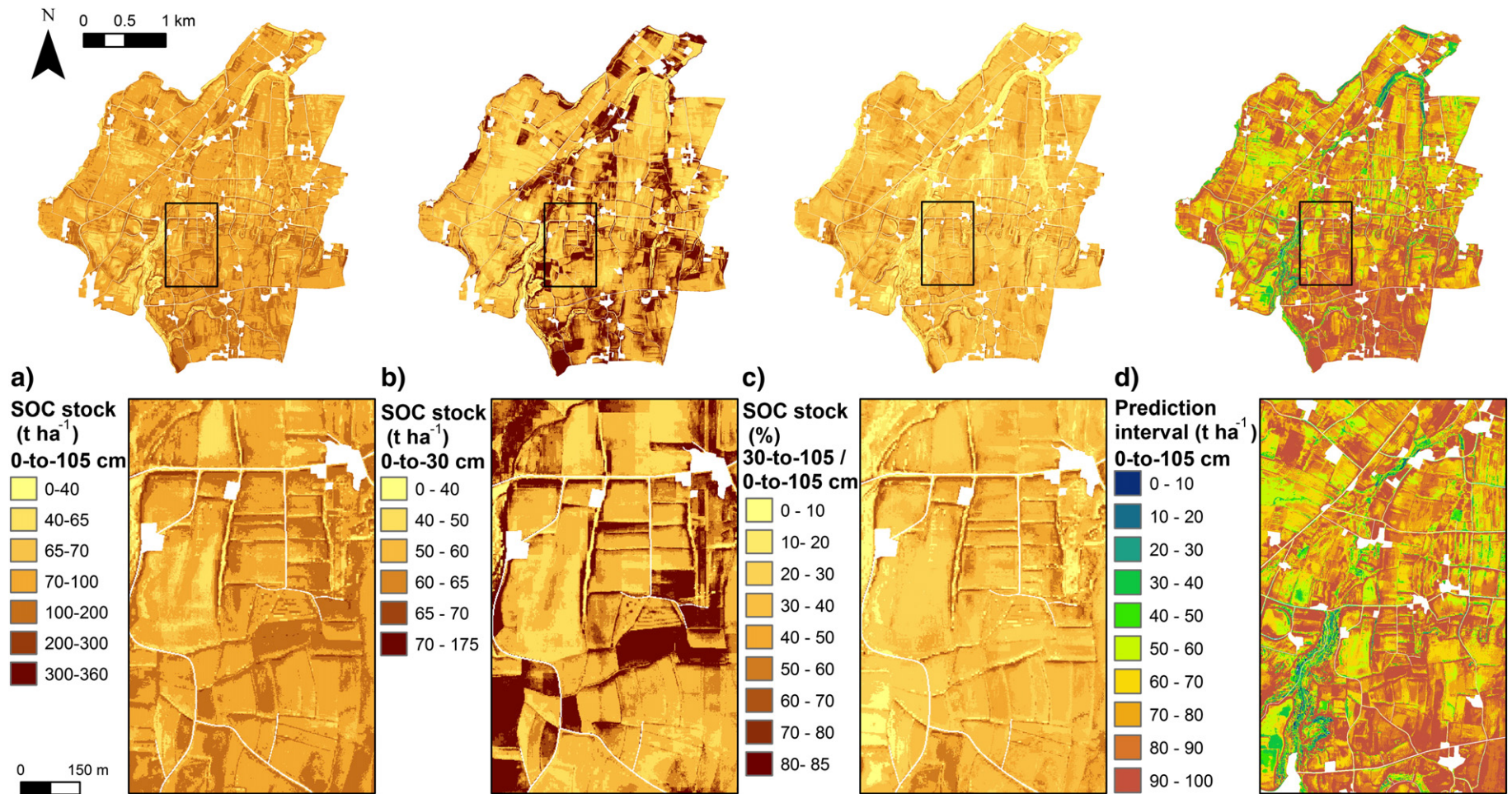


Fig. 10. Predictive maps of SOC stocks over the 10 km² study area with a local zoom on a 60 ha area (square maps are details of the study area): a) total SOC stocks (0–105 cm), b) SOC stocks within the tillage horizon (0–30 cm), c) SOC stocks in deep horizons (30–105 cm), in proportion of the total SOC stock, d) prediction interval of the prediction of the total SOC stocks (0–105 cm soil layer, tC ha^{-1}).

by tillage mixing of SOC content in these two topsoil layers, so there may be no need of very fine auxiliary information (i.e. with resolution <20 m here) for the prediction of SOC content in these layers. In our landscape, SOC content showed a high variability in the very topsoil layer (0 to 7.5 cm), mostly because of fresh organic matter supply in areas close to hedges. In deep soil layers and for our study area, high SOC content could be mostly the result of soil erosion and accumulation processes, in strong link with linear structures within the landscape (e.g. the ditches or the hedges; Follain et al., 2009). Actually, the impact of soil redistribution (soil erosion and deposit) on the SOC storage is subject to discussion. The conclusion depends on which scale and processes are taken into account. For example, Lal and Pimentel (2008) argued that soil erosion leads to SOC loss, but Van Oost et al. (2007, 2008) maintained that it leads to SOC storage. The impact of soil erosion depends on (i) the fate of the SOC in the eroded soil (exportation or deposition out of the area of interest) and (ii) the evolution of the SOC mineralization (at the eroded sites, during soil transportation by the erosion processes and at the deposit sites, in soil horizon at different depth). Van Oost et al. (2006) modeled the impact of soil erosion on SOC stocks by testing various hypothesis on SOC dynamic during soil transportation and on SOC dynamic in the deep soil horizons (after burying). They concluded that soil erosion could conduce both to SOC loss and storage. Doetterl et al. (2012) estimated that soil erosion processes could lead to carbon stabilization in areas where of soil deposit, but also in areas of erosion. The net balance would contribute to net carbon storage. However, they didn't take into account the priming effect observed in other studies (e.g. Fontaine et al., 2007). Moreover, the carbon dynamic in the deep soil layers seems to be highly dependent on the spatial distribution of the fresh organic matter, the stable carbon and the microorganisms responsible for SOC mineralization (Rumpel and Koegel-Knabner, 2011). Berhe (2012) also underlined that the SOC mineralization rate could vary according to the location in landscape. Concerning the impact of linear landscape features (hedges) on SOC storage, Follain et al. (2006) demonstrated that (i) the combination of diffusive transport and water erosion could significantly modify the topography and soil redistribution over a few centuries and (ii) hedges modify soil distribution and landforms by favoring deposition in the uphill position and soil erosion in the downhill position in agreement with field observations. Actually, such soil accumulation and A-horizon thickening uphill from hedges have been observed in the type of landscape that we studied (Follain, 2006; Follain et al., 2009; Lacoste, 2012). Moreover, SOC storage due to hedges has been described in previous studies, due to (i) a local effect under the hedge due to tree activity and (ii) an anti-erosive effect at the hill-slope scale (Follain et al., 2007; Walter et al., 2003b). Then, SOC content attested high spatial variability in the very topsoil layer and in deep soil layer, and this could explain why finer predictors (terrain attributes with a 5 m resolution) were used for SOC content predictions in this soil layers. According to the BD predictions, the choice of the terrain attributes' resolution for each standard layer was not so consistent and more difficult to explain (Table 6). Terrain attributes at coarser resolution (10 and 20 m) were mostly preferred to those at fine resolution (5 m) to be used as predictors.

4.1.3. Creation of independent predictive models (for each standard soil layer and for each soil properties)

The creation of the predictive models was made independently for each of the standard soil layer: there was no initial guarantee of consistency for the predictions within soil profiles. The reconstruction of the soil profiles by fitting the spline function to the predicted values of SOC content and BD was a relevant solution to deal with this problem. This method was shown to be able to remove part of the inconsistencies in the predictions, without altering the accuracy of the prediction (Fig. 5). Moreover, the predicting process we used was proved to maintain the existing relation between the soil properties

(SOC content and BD, Fig. 8). Finally, SOC content and BD were both predicted until the depth of 105 cm, for every pixels of the study area. But A-horizon was characterized by highly variable thickness (from 3 to 250 cm), and at some locations the soil was shallower than 105 cm, and deeper at others. To deal with this depth problem, we used a predictive map of A-horizons' thickness, also produced with Cubist (Lacoste, 2012). This predictive map of A-horizons' thickness was used to correct the SOC content and BD maps, removing values in the soil layers not predicted as A-horizon. However, it remained a problem concerning the A-horizons deeper than 105 cm, mostly predicted under or close to some hedges (from 0 to 5 m uphill from hedges), where soil accumulation was possible, or where Aeolian loam was the soil parent material. SOC content and BD were not predicted in soils at depths over 105 cm and then, we possibly underestimated SOC stocks in soils deeper than 105 cm.

4.2. Importance of environmental predictors

According to SOC content predictions, we found a strong influence of land use (frequency of grassland), hydrologic terrain attributes (horizontal and vertical flow distances) and some topographic attributes (slope, difference between the 20 m and the 5 m DEM; Table 5). Other studies have also demonstrated a strong influence of topography on the distribution of SOC content (Florinsky et al., 2002; Grimm et al., 2008; Kheir et al., 2010; Kravchenko et al., 2006a, 2006b; Malone et al., 2009; Vasques et al., 2010). This could be explained by dependences between the spatial differentiation of SOC content, soil redistribution (soil erosion and deposit) and soil waterlogging, highly linked to landscape morphology (Arrouays et al., 1998; Florinsky et al., 2002; Moore et al., 1993; Walter et al., 2003a). Regarding BD, the best predictors were topographic terrain attributes (elevation and multiresolution index of ridge top flatness), the deviation from the mean K emissions and a hydrographic terrain attribute (the modified topographic wetness index). In our landscape, these covariates allowed identifying the global position in the landscape (plateau, hillslope, plain) and the position from local topographical object such as hedges on bank or ditches, but also the soil parent material and the waterlogging and drainage conditions which were highly linked to the landscape geomorphology.

4.3. Predictive models performance

4.3.1. Overall performance

Regarding SOC content and without distinction of depth (overall performance on soil profiles), the predictive models showed a MAE of 2.31 g kg^{-1} , a RMSE of 4.05 g kg^{-1} , a R^2 of 0.80, and a ρ_c of 0.88 for evaluation on the calibration dataset. For evaluation on the two datasets used for external evaluation and taken together, these models showed a MAE of 3.36 g kg^{-1} , a RMSE of 5.92 g kg^{-1} , a R^2 of 0.60, and a ρ_c of 0.73. These values proved that the models reached to explain SOC content values from the calibration dataset, according to the environmental covariates that we used. The predictive models were also quite efficient to predict SOC content values for new locations, not used in the model creation step. This attested that: (i) our calibration dataset, even limited (70 sampling location for a study area of 10 km^2), was representative of the SOC content value at the scale of our study area (this idea was also supported by the results in Table 7), and then, the sampling method that we used, i.e. cLHS, was proved to be efficient to design a sampling population able to capture the SOC content variation at the landscape scale; (ii) the environmental covariates that we used were efficient to explain the SOC content values at landscape scale, and at a fine resolution; and (iii) the machine learning tool that we used was able to capture the relationship between the SOC content values and the environmental covariates.

In comparison to previous study, our predictive models present quite better performance. For example, Florinsky et al. (2002) used

a regression equation with a R^2 of 0.37 to predict SOC content in A-horizons in an agricultural area in the middle part of Canada. Malone et al. (2009) produced SOC content maps for soil depth up to 1 m in an agricultural area of Western Australia. Their models fit the measured values with an R^2 of 0.44, but cross-validation on the training datasets results in R^2 value of 0.10. Moreover, Malone et al. (2009) argues that, for quantitative soil spatial models, R^2 over 0.50 is unusual and values of 0.50 or less are common.

Regarding BD and without distinction of depth (overall performance on soil profiles), the predictive models showed a MAE of 0.07 cm^{-3} , a RMSE of 0.09 cm^{-3} , a R^2 of 0.77, and a ρ_c of 0.86 for evaluation on the calibration dataset. For evaluation on the two datasets used for external evaluation and taken together, these models showed a MAE of 0.14 g cm^{-3} , a RMSE of 0.19 g cm^{-3} , a R^2 of 0.23, and a ρ_c of 0.38. The good models performance showed in internal validation proved that these models reached to explain SOC content values from the calibration dataset, according to the environmental covariates that we used. Nevertheless, the predictive models showed poorer performance according to external evaluation. Results given by Table 7 indicate that distribution of BD for the calibration dataset was in accordance with those of the other datasets (landscape and transect validation datasets), which suggest that the calibration dataset was quite well representative to the BD values for the study area. This could imply that the poor models performance in external evaluation could be due to a lack of covariates to explain the BD variations. In fact, some variables that are important to explain BD variation were hard to take into account, e.g. the type of tillage (depth and tool used) or the history of farming practices (e.g. the time elapsed since the last tillage, the use of tool which could lead to soil compaction). Other studies mainly predict BD using pedotransfer functions based on SOC content, depth and clay, silt and sand content. Martin et al. (2009) found that the best predictor was the soil concentration in organic carbon. As we see that there was a strong relation between SOC content and BD in our data (Fig. 8). A two steps method may be an interesting way to improve the BD prediction: a predictive map of SOC content could be produced in a first step, and then, in a second, used among the covariates step to predict BD.

However, our results were consistent with some previous studies. Martin et al. (2009) compared seven predictions of BD at the French national scale in agricultural or natural areas, five produced with PTFs with a R^2 ranging from 0.29 to 0.37 and two produced with a boosted regression tree with a R^2 ranging from 0.83 to 0.94. Tranter et al. (2007) compared BD predictions across Australia using neural networks, multiple linear regression and regression trees (Cubist), and obtained R^2 ranging from 0.39 to 0.48, 0.41 to 0.50 and 0.20 to 0.43, respectively. These same models had an RMSE ranging from 0.13 to 0.16, 0.15 to 0.18 and 0.16 to 0.19 g cm^{-3} , respectively.

4.3.2. Variation of models performance with depth and distance to the hedges

We used two validation datasets for the evaluation of the SOC content and BD predictive models: the landscape validation dataset and the transect validation dataset, which were two independent datasets. The landscape validation dataset incorporated sampling locations distributed in the whole study area, whereas the transect dataset was concentrated on locations close to hedges. In fact, we assessed that the SOC content and BD distributions in space and depth differed regarding the locations inside the study area, and especially according to the hedges proximity. For the SOC content, the predictive models showed a better performance for the evaluation on the transect validation dataset than for the evaluation on the landscape validation dataset (Table 2). This could attest that the prediction of SOC content at hedges proximity (until 20 m uphill from hedges and 10 m downhill from hedges) was better than the general SOC content prediction at the Landscape scale. This could be explained by the fact that SOC distribution close to hedges may mostly be due to fresh organic matter supply from the hedge and soil redistribution (Walter et al., 2003b). Yet, no information about

fresh organic matter supply was contained in the covariates used in the predictive models creation (Table 1). Moreover, these covariates used in the predictive models were mostly linked to soil redistribution (slope, overland flow distance, difference in elevation for the 5 m and 20 m DEM, etc., Table 5). The SOC content prediction at the scale of the study area could be improved by using covariates including data on soil organic improver. For the BD, the predictive models showed a slightly lower performance for the evaluation on the transect validation dataset than for the evaluation on the landscape validation dataset (Table 3). This involved that the performances of the BD predictive model were equivalent for locations close to hedges or at the scale of the study area.

In our study, we found better predictions of SOC content for layers in the middle of the soil profile (between 15 and 60 cm, Table 2). These results differ from other studies: for Malone et al. (2009), the accuracy of carbon prediction decreased for each depth interval; Kempen et al. (2011), predicting SOC stocks up to a depth of 90 cm for agricultural soils in the Netherlands, found highest prediction accuracy for the topsoil layer (0–30 cm), with higher errors in the 30–60 cm soil layer. Our results could be explain in three points: (i) poorer models performance for the very topsoil layer (0 to 7.5 cm) and in the deeper soil layers could be due to a larger SOC content range for this layer (Fig. 6), which make harder to explain the SOC form the covariates; (ii) better performance in the bottom part of the tilled soil layer (15 to 30 cm) could be due to a lower range of the SOC content value because of the soil mixing during soil tillage; and finally (iii), poorer performance in deeper soil horizon could be due to a decrease of data for these layers both in calibration or in landscape and transect validation datasets (deep A-horizons were not observed for every soil profiles), and so on, to a greater difficulty to create predictive soil models and to estimate their performance. Concerning the prediction of SOC content, our results indicated that the predictions were not influenced by the distance to hedges, except at 2 m downhill from the hedge where the RMSE was higher, especially for the 0-to-30 cm soil layers. This could be explained by a larger range of SOC values and a higher spatial variation due to the hedge and the topography influence on SOC content close to hedges (Follain et al., 2007; Walter et al., 2003b). Finally, the models showed a good predictive performance in areas close to hedges (Table 4).

Concerning the BD predictions, the highest prediction error was found for the evaluation on the transect validation dataset, and especially for the topsoil horizons and the deeper soil horizons. In comparison with the other datasets, the transect validation dataset was characterized by the larger range of BD values. This can explain the poor predictive performance we obtained. Similar results were observed in De Vos et al. (2005) when predicting BD for forest soil. Regarding the BD predictions, our results attested that the distance to the hedges didn't have strong influence on the predictive performance, except for the topsoil layer (0 to 7.5 cm) and for location close the hedges (2 m downhill and uphill from the hedge and just under the hedge) where the RMSE values were slightly higher (Table 4). This result could be explained by the fact that (i) hedges have a specific impact on soil properties and moreover on soil BD (Walter et al., 2003b), so they induce a larger variability of BD values which make harder the BD prediction close to hedges, and (ii) there is no soil tillage at hedges vicinity, so there is no soil homogenisation and the BD values for the 0-to-7.5 cm soil layer were very different from the BD values for the below soil layers (Fig. 7).

4.4. SOC distribution at the landscape scale

Our results are consistent with previous estimations of SOC stocks in French soils (Arrouays et al., 2002), except for the forest. For Arrouays et al. (2002), the mean SOC stocks for the 0-to-30 cm layer was estimated at 40 t ha^{-1} under cultivated fields, 65 t ha^{-1} under grassland and 70 t ha^{-1} under forest. From our results, we estimated the mean SOC stocks for the 0-to-30 cm soil layer to 44 t ha^{-1} under cultivated fields,

67 t ha⁻¹ under grasslands and 30 t ha⁻¹ under forest. The lower SOC stocks under forest could be explained in our study area by the fact that forest was encountered mostly on shallower soils from hard schist or in hard slopes, whereas deep soils were used for grasslands and cultivated fields. Consequently, in our study area, the soil carbon storage capacity was lower under forest than under the cultivated fields or grassland.

5. Conclusion

The method that we implemented to map the 3-dimensional spatial distribution of SOC stocks at the landscape scale is innovative in three main aspects. First, we chose to model and map the SOC content and the BD from soil surface up to a depth of 105 cm, with a resolution of 2 m to deal with our heterogeneous study site. According to our knowledge, there is no previous digital soil mapping study at such a high resolution. Second, we used the cLHS to select our calibration dataset: this sampling method has been proposed for data mining and digital soil mapping (Minasny and McBratney, 2006), but there is no example in the literature of its use in a real mapping study. Finally, the combination of depth functions (equal area splines) and machine learning method turn out to be efficient to produce continuous 3D maps of SOC content and BD, which are essential data to understand the soil functions and even more to prospect soil evolution.

Acknowledgments

This research was financially supported by the French National Research Agency, “Vulnerability program: environments, climates and societies (VMCS)”, Project “LandSoil” n. ANR-08-VMCS-006-01. The authors thank the Zone Atelier Armorique for LiDAR and land-use data provision, and are grateful to G. Dutin for technical assistance.

References

- Aïchi, H., Fouad, Y., Walter, C., Viscarra Rossel, R.A., Chabaane, Z.L., Sanaa, M., 2009. Regional predictions of soil organic carbon content from spectral reflectance measurements. *Biosystems Engineering* 104, 442–446.
- Arrouays, D., Daroussin, J., Kicin, J.L., Hassika, P., 1998. Improving topsoil carbon storage prediction using a digital elevation model in temperate forest soils of France. *Soil Science* 163, 103–108.
- Arrouays, D., Balesdent, J., Germon, J.C., Jayet, P.A., Soussana, J.F., Stengel, P., 2002. Contribution à la lutte contre l'effet de serre: Stocker du carbone dans les sols agricoles de France? Expertise Scientifique Collective. INRA, Paris (332 pp.).
- Berhe, A.A., 2012. Decomposition of organic substrates at eroding vs. depositional land-form positions. *Plant and Soil* 350, 261–280.
- Beven, K.J., Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Journal* 24, 43–69.
- Bishop, T.F.A., McBratney, A.B., Laslett, G.M., 1999. Modelling soil attribute depth functions with equal-area quadratic smoothing splines. *Geoderma* 91, 27–45.
- Bonijoly, D., Perrin, J., Truffert, C., Asfirne, F., 1999. Couverture géophysique aéroportée du Massif armoricain (Rap. BRGM R 40471. 75 pp.).
- Brus, D.J., Kempen, B., Heuvelink, G.B.M., 2011. Sampling for validation of digital soil maps. *European Journal of Soil Science* 62, 394–407.
- Burel, F., Baudry, J., 2003. *Landscape Ecology: Concepts, Methods and Applications*. Science Publ., Enfield, NH.
- Chantraine, J., Rabu, D., Béchenne, F., 2002. Carte géologique numérique à 1:250 000 du Massif armoricain. Version 1.1., BRGM (Bureau de Recherches Géologiques et Minières).
- De Vos, B., Van Meirvenne, M., Quataert, P., Deckers, J., Muys, B., 2005. Predictive quality of pedotransfer functions for estimating SBD of forest soils. *Soil Science Society of America Journal* 69, 500–510.
- Doetterl, S., Six, J., Van Wesemael, B., Van Oost, K., 2012. Carbon cycling in eroding landscapes: geomorphic controls on soil organic C pool composition and C stabilization. *Global Change Biology* 18, 2218–2232.
- Florinsky, I.V., Eilers, R.G., Manning, G.R., Fuller, L.G., 2002. Prediction of soil properties by digital terrain modelling. *Environmental Modelling & Software* 17, 295–311.
- Follain, S., 2006. Effet du réseau bocage sur l'organisation des sols. Redistribution des sols et stockage en carbone organique. (Thesis) E.N.S.A.R., Rennes, France (241 pp.).
- Follain, S., Minasny, B., McBratney, A.B., Walter, C., 2006. Simulation of soil thickness evolution in a complex agricultural landscape at fine spatial and temporal scales. *Geoderma* 133, 71–86.
- Follain, S., Walter, C., Legout, A., Lemerrier, B., Dutin, G., 2007. Induced effects of hedgerow networks on soil organic carbon storage within an agricultural landscape. *Geoderma* 142, 80–95.
- Follain, S., Walter, C., Bonte, P., Marguerie, D., Lefevre, I., 2009. A-horizon dynamics in a historical hedged landscape. *Geoderma* 150, 334–343.
- Fontaine, S., Barot, S., Barre, P., Bdioui, N., Mary, B., Rumpel, C., 2007. Stability of organic carbon in deep soil layers controlled by fresh carbon supply. *Nature* 450, 277–U10.
- Forman, R.T.T., 1995. Some general principles of landscape and regional ecology. *Landscape Ecology* 10, 133–142.
- Freeman, T.G., 1991. Calculating catchment-area with divergent flow based on a regular grid. *Computers & Geosciences* 17, 413–422.
- Gallant, J.C., Dowling, T.I., 2003. A multiresolution index of valley bottom flatness for mapping depositional areas. *Water Resources Research* 39, 1–13.
- Grace, J., 2004. Understanding and managing the global carbon cycle. *Journal of Ecology* 92, 189–202.
- Grimm, R., Behrens, T., Marker, M., Elsenbeer, H., 2008. Soil organic carbon concentrations and stocks on Barro Colorado Island — digital soil mapping using random forests analysis. *Geoderma* 146, 102–113.
- IUSS Working Group WRB, 2007. World Reference Base for Soil Resources 2006, First Update 2007. World Soil Resources Reports No. 103. FAO, Rome.
- Jenny, H., 1941. Factors of soil formation: A system of Quantitative Pedology. McGraw-Hill, New-York (191 pp.).
- Johnston, C.A., et al., 2004. Carbon cycling in soil. *Frontiers in Ecology and the Environment* 2, 522–528.
- 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.
- Kheir, R.B., et al., 2010. Predictive mapping of soil organic carbon in wet cultivated lands using classification-tree based models: the case study of Denmark. *Journal of Environmental Management* 91, 1150–1160.
- Kravchenko, A.N., Robertson, G.P., Hao, X., Bullock, D.G., 2006a. Management practice effects on surface total carbon: differences in spatial variability patterns. *Agronomy Journal* 98, 1559–1568.
- Kravchenko, A.N., Robertson, G.R., Snap, S.S., Smucker, A.J.M., 2006b. Using information about spatial variability to improve estimates of total soil carbon. *Agronomy Journal* 98, 823–829.
- Lacoste, M., 2012. Evolution des sols à l'échelle du paysage sous des conditions de changement climatique et de structure du paysage. (PhD thesis) Agrocampus-Ouest, Rennes (243 pp.).
- Lacoste, M., Lemerrier, B., Walter, C., 2011. Regional mapping of soil parent material by machine learning based on point data. *Geomorphology* 133, 90–99.
- Lagacherie, P., McBratney, A., Voltz, M., 2007. Digital soil mapping — an introductory perspective. In: Hartemink, A.E., McBratney, A. (Eds.), *Developments in Soil Science*. Elsevier, Amsterdam.
- Lal, R., Pimentel, D., 2008. Soil erosion: a carbon sink or source? *Science* 319, 1040–1042.
- Lemerrier, B., Lacoste, M., Loum, M., Walter, C., 2011. Extrapolation at regional scale of local soil knowledge using boosted classification trees: a two-step approach. *Geoderma* 171, 75–84.
- Lin, L.I.-K., 1989. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 45, 255–268.
- Lin, L.I.-K., 2000. A note on the concordance correlation coefficient. *Biometrics* 56, 324–325.
- Malone, B.P., McBratney, A.B., Minasny, B., Laslett, G.M., 2009. Mapping continuous depth functions of soil carbon storage and available water capacity. *Geoderma* 154, 138–152.
- Malone, B.P., McBratney, A.B., Minasny, B., 2011. Empirical estimates of uncertainty for mapping continuous depth functions of soil attributes. *Geoderma* 160, 614–626.
- Martin, M.P., et al., 2009. Optimizing pedotransfer functions for estimating SBD using boosted regression trees. *Soil Science Society of America Journal* 73, 485–493.
- McBratney, A.B., Santos, M.L.M., Minasny, B., 2003. On digital soil mapping. *Geoderma* 117, 3–52.
- Meersmans, J., et al., 2009. Changes in organic carbon distribution with depth in agricultural soils in northern Belgium, 1960–2006. *Global Change Biology* 15, 2739–2750.
- Merot, P., et al., 2003. Testing a climato-topographic index for predicting wetlands distribution along a European climate gradient. *Ecological Modelling* 163, 51–71.
- Minasny, B., McBratney, A.B., 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Computers & Geosciences* 32, 1378–1388.
- Mishra, U., et al., 2009. Predicting Soil organic carbon stock using profile depth distribution functions and ordinary kriging. *Soil Science Society of America Journal* 73, 614–621.
- Mishra, U., Lal, R., Liu, D.S., Van Meirvenne, M., 2010. Predicting the spatial variation of the soil organic carbon pool at a regional scale. *Soil Science Society of America Journal* 74, 906–914.
- Moore, I.D., Gessler, P.E., Nielsen, G.A., Peterson, G.A., 1993. Soil attribute prediction using terrain analysis. *Soil Science Society of America Journal* 57, 1548.
- O'Callaghan, J.F., Mark, D.M., 1984. The extraction of drainage networks from digital elevation data. *Computer Vision, Graphics and Image Processing* 28, 323–344.
- Quinlan, J.R., 1992. Learning with continuous classes. *Proceedings of the 5th Australian Joint Conference on Artificial Intelligence*, pp. 343–348.
- Quinlan, J.R., 1993. Combining instance-based and model-based learning. In: Kaufmann, M. (Ed.), *Proceedings of the Tenth International Conference on Machine Learning*, pp. 236–243.
- Quinlan, J.R., 1994. C4.5: Programs for machine learning. *Machine Learning* 16, 235–240.
- Rumpel, C., Koegel-Knabner, I., 2011. Deep soil organic matter—a key but poorly understood component of terrestrial C cycle. *Plant and Soil* 338, 143–158.
- Solomatine, D.P., Shrestha, D.L., 2009. A novel method to estimate model uncertainty using machine learning techniques. *Water Resources Research* 45.

- Tranter, G., et al., 2007. Building and testing conceptual and empirical models for predicting SBD. *Soil Use and Management* 23, 437–443.
- Turner, M.G., 1989. Landscape ecology- — the effect of pattern on process. *Annual Review of Ecology and Systematics* 20, 171–197.
- Ungaro, F., Staffilani, F., Tarocco, P., 2010. Assessing and mapping topsoil organic carbon stock at regional scale: a scorpan kriging approach conditional on soil map delineations and land use. *Land Degradation & Development* 21, 565–581.
- Van Oost, K., Quine, T.A., Govers, G., Heckrath, G., 2006. Modeling soil erosion induced carbon fluxes between soil and atmosphere on agricultural land using SPEROS-C. In: Roose, E.J., Lal, R., Feller, C., Barthes, B., Stewart, B.A. (Eds.), *Soil Erosion and Carbon Dynamics*. Taylor & Francis Group, USA, pp. 37–51.
- Van Oost, K., Quine, T.A., Govers, G., De Gryze, S., Six, J., Harden, J.W., Ritchie, J.C., McCarty, G.W., Heckrath, G., Kosmas, C., Giraldez, J.V., da Silva, J.R.M., Merckx, R., 2007. The impact of agricultural soil erosion on the global carbon cycle. *Science* 318, 626–629.
- Van Oost, K., Six, J., Govers, G., Quine, T., De Gryze, S., 2008. Soil erosion: a carbon sink or source? *Response. Science* 319, 1042.
- Vasques, G.M., Grunwald, S., Comerford, N.B., Sickman, J.O., 2010. Regional modelling of soil carbon at multiple depths within a subtropical watershed. *Geoderma* 156, 326–336.
- Viaud, V., Angers, D.A., Walter, C., 2010. Toward landscape-scale modelling of soil organic matter dynamics in agroecosystems. *Soil Science Society of America Journal* 74, 1847–1860.
- Walter, C., Rossel, R.A.V., McBratney, A.B., 2003a. Spatio-temporal simulation of the field-scale evolution of organic carbon over the landscape. *Soil Science Society of America Journal* 67, 1477–1486.
- Walter, C., Merot, P., Leyer, B., Dutin, G., 2003b. The effect of hedgerows on soil organic carbon storage in hillslopes. *Soil Use and Management* 19, 201–207.
- Walton, J.T., 2008. Subpixel urban land cover estimation: comparing cubist, random forests, and support vector regression. *Photogrammetric Engineering and Remote Sensing* 74, 1213–1222.
- Zhang, Z.Q., et al., 2010. Application of categorical information in the spatial prediction of soil organic carbon in the red soil area of China. *Soil Science and Plant Nutrition* 56, 307–318.