Mapping soil properties at multiple depths from disaggregated legacy soil maps in the Brittany region, France

Yosra Ellili-Bargaoui, Christian Walter, Didier Michot, Blandine Lemercier

Title: Mapping soil properties at multiple depths from disaggregated legacy soil maps in the Brittany region, France

Authors:

Yosra Ellili-Bargaoui[1,2], Christian Walter[1], Didier Michot[1], and Blandine Lemercier[1]

[1] UMR SAS, Institut Agro, INRAE, 35000 Rennes, France

[2] Interact, UniLaSalle, 60 000 Beauvais, France

Corresponding Author: Yosra Ellili-Bargaoui

Corresponding Author's Institution: Interact, UniLaSalle, 60 000 Beauvais, France

Corresponding Author's contact (email) Yosra.ellili@unilasalle.fr

Abstract

Digital Soil Mapping (DSM) is increasingly needed to improve existing soil information and derive soil property maps at the suitable spatial resolution for sustainable soil landscape management. However, predicting several soil properties while preserving specific pedological process is a great challenge, particularly when only coarse soil information is available over large areas. Spatial disaggregation seems to be an effective technique to extract pedological information by downscaling the original information to produce soil maps at finer resolutions. In a previous study, legacy soil maps of Brittany (France) were disaggregated at a 50 m spatial resolution using the DSMART (Disaggregation and Harmonisation of Soil Map Units Through Resampled Classification Trees) algorithm and pedological knowledge. The present study had two main objectives: (i) assess the preservation of the relationships between soil properties when soil properties are estimated at standard depths by applying the equal-area spline method on soil data at pedon scale, and (ii) combine disaggregated soil maps and spline-function results to estimate spatial patterns of nine soil properties for six regular soil-depth intervals down to 200 cm across Brittany, an area of 27 040 km². To this end, soil properties were first generated for standard soil-depth intervals using spline functions. Then, for mapping soil properties at the six standard depths, weighted mean of each soil attribute was calculated for each grid cell from reference soil-property values of the three most probable predicted soil types. Their associated probabilities of occurrence were used as weights. To assess the ability of spline functions to preserve soil-property relationships, multiple statistical analyses were performed using original and splined soil datasets. Bivariate and multivariate analysis highlighted that spline functions preserved soil-property relationships. Derived digital soil maps showed strong spatial patterns: SOC and silt contents generally decreased with depth, while sand content and coarse fragment percentage consistently increased with depth. In addition, experimental semivariogram analysis of SOC content showed high spatial variability over short distances for all soil-depth intervals except the deepest (100-200 cm), while silt content showed high semivariance for the deepest soil layers. This study can be considered an example of harmonisation to common output specifications, which generates a geo-database of quantitative soil properties that describe lateral and vertical soil variation for

regular depth intervals. These predictions can be incorporated into environmental models to help decision makers manage landscapes.

Keywords: Digital soil mapping, Equal-area spline functions, Spatial disaggregation, DSMART, Regional scale, Legacy soil data, Multiple soil classes

1. Introduction

Addressing environmental issues requires accurate information about spatial patterns of soil types and properties. Consequently, providing quantitative soil information of known accuracy is a great challenge to satisfy the needs of end-users, especially landscape managers and stakeholders (Ellili et al., 2019, Chaney et al., 2016). In most environmental and agricultural research, accurate and continuous soil data are increasingly incorporated in soil-landscape modelling to monitor natural resources and ecosystems (Odgers et al., 2012). However, soil data are usually not available at the adequate spatial resolution, particularly over large areas, where only legacy soil maps at coarse spatial resolution are available.

From a pedological viewpoint, soil attributes vary either continuously or sharply down a soil profile (Ponce-Hernandez et al., 1986) as well as across the landscape. As soils are often described in terms of morphological horizons, it is often difficult to derive meaningful comparisons of soil phenomena when dealing with a collection of soil profile information because soil horizonation varies from one profile to the next. Soil depth functions are useful here because they facilitate the harmonization of depths within a profile, allowing easier comparisons of soil properties from site to site because they now have the same depth support. There are a number of different types of soil depth functions that could or have been used for certain applications. For instance, Jenny (1941) made the earliest known soil depth function by drawing freehand curves between data points that represented the mid-point depth of a given horizon. More sophisticated approaches have been used, such as fitting exponential decay functions (Russell and Moore, 1968). However, the main disadvantage of these methods is that each local variation in the soil profile affects the shape of the fitted function at all depths. Consequently, the low flexibility of these functions results in variable quality of fit over soil depth (Malone et

al., 2009; Odgers et al., 2012). Moreover, exponential decay function is specifically applicable for soil variables like soil carbon, as done by Minasny et al. (2006) who fitted these functions to soil organic carbon (SOC) data in the soil profile to map carbon storage in the Lower Namoi Valley, Australia, and achieved an acceptable quality of fit.

To compensate for the lack of flexibility in depth functions, the use of spline functions seems to be a good alternative. In fact, certain spline functions, such as smoothing splines (Erh, 1972) and equal-area splines (Ponce-Hernandez et al., 1986), allowed a series of independent local functions to be fit over small intervals of a soil profile. Bishop et al. (1999) improved the approach of Ponce-Hernandez et al. (1986) by fitting quadratic polynomial splines to soil horizons. Their modified equal-area quadratic splines effectively predicted depth functions for soil pH, electrical conductivity, clay content, SOC content, and gravimetric water content at -33 kPa. However, their method required input data from contiguous soil horizons. One decade later, Malone et al. (2009) solved this issue by generalizing Bishop et al. (1999)'s approach to be able to use soil input data from non-contiguous soil horizons.

Using soil-depth functions allows soil attributes to be predicted at specific soil depths. Spline functions are applied to individual soil observations and predict attributes only for a single geographic point. To characterize three-dimensional (3D) variation in soil properties, intensive soil sampling is needed. The quadratic smoothing spline is amply capable of addressing the vertical variation challenges of soil profile harmonization but to characterize the lateral variability, sufficient spatial sampling is required. Therefore, combining digital soil mapping (DSM) techniques and soil-depth functions appears to be a good option to capture both lateral and vertical variations in soil properties. DSM predicts soil properties based on their relationships with environmental variables (Minasny et al., 2008) to address soil variability, even in areas with limited soil data

(McBratney. 2003). Malone et al. (2009) combined equal-area smoothing splines and neural network models to map SOC storage and available water capacity based on limited soil data in the lower valley of the Namoi River, Australia. Lacoste et al. (2014) also combined DSM techniques and equal-area splines to derive 3D maps of SOC stock at high spatial resolution across an agricultural landscape in Brittany, France. Other researchers (Bishop et al., 2015; Vaysse et al., 2015) adopted the same strategy to map soil properties at specific soil depths across a defined study area by using existing soil databases.

Overall, spline functions are generally coupled with DSM methods to characterize the spatial distribution of soil properties while respecting the consortium GlobalSoilMap specifications (Arrouays et al., 2014). Over large areas, disaggregation approaches are strongly recommended to confront the scarcity of both accurate and site soil data. These approaches attempt to downscale the Soil Map Unit (SMU) information to delineate unmapped Soil Type Unit (STU) (Bui and Moran, 2001; Odgers et al., 2014; Holmes et al., 2015) and then derive soil property maps. Several techniques have been demonstrated through soil science literature and tested in different case studies around the world. For instance, Odgers et al. (2014) have developed the Disaggregation and harmonization of Soil Map Units Through Resampled Classification Trees (DSMART) algorithm to spatially delineated the STU within each SMU in Australia. The DSMART algorithm generates a set of rasters depicting the spatial distribution of STU and their associated probability of occurrence at each pixel. Then Odgers et al. (2015) designed the PROPR algorithm which convert the probability information into thematic soil property maps, particularly pH and Cation Exchange Capacity (CEC), using soil profile information for each STU (Odgers et al., 2015a; Odgers et al., 2015b). The DSMART algorithm was also implemented by Chaney et al. (2016) to disaggregate the soil map of the contiguous United States at a 30m spatial resolution.

More importantly, integrating local pedological knowledge has been recognized as an effective way to enhance the performance of DSM approaches (Cook et al., 1996; Walter et al., 2006; Stoorvogel et al., 2017; Møller et al., 2019, Jamshidi et al., 2019; Ellili-Bargaoui., 2020; Pallegedara Dewage et al., 2020), Vincent et al. (2018) have implemented the DSMART algorithm with soil landscape rules describing soil distribution in the local context of the Brittany region (France).

The aims of this study were i) to develop a method to map nine soil properties at multiple soil depths ii) to assess the ability of spline functions to preserve relationships among soil properties and iii) to assess the stability of a pedotransfer function, which predict the Effective Cation Exchange Capacity (ECEC) from some soil attributes. We did this by i) fitting spline functions to STU soil information to express soil attributes for six regular depth intervals respecting the GlobalSoilMap specifications and ii) performing bivariate and multivariate statistical analysis of soil attributes before and after fitting splines to assess the goodness of fit and degree of preservation of soil-property relationships. Overall, maps of nine soil properties of major interest for agronomic and environmental purposes were generated at the regional scale by combining DSM techniques with equal-area splines. STU soil attributes were first standardized for regular depth intervals down to 200 cm and then mapped using disaggregated STU maps obtained from a previous study at a spatial resolution of 50 m (Ellili-Bargaoui et al. 2020; Vincent et al., 2018).

2. Materials and methods

The overall workflow followed in this study is as follows:

i) Extract soil-property information for all strata of each STU and identify the nature of the bottom STU strata (i.e. lithic or paralithic contact). In our context, the strata refers to a set of spatial soil layers describing the vertical structuration of STU.

ii) Fit equal-area spline functions to soil properties using STU features to obtain estimates for standard depth intervals.

iii) Assess preservation of soil-property relationships after fitting equal-area spline functions

iv) Assess the ability of splined data to correctly predict a composite variable that depends on additional soil properties (ECEC). The main idea was to check the stability of the ECEC pedotransfer function by fitting spline functions.

v) Map soil properties over the study area by combining disaggregated STU maps, their associated probability of occurrence maps and fitted equal-area soil-property values.

2.1. The study area

The study area was the Brittany region, in north-western France. Its covers approximately 27 040 km², with a variety of physical and geographic features. Its climate is oceanic, with mean annual precipitation ranging from 650 mm in the east to 1300 mm in the west, and mean annual temperature ranging from 11-13°C (CLIMATIK database: INRA, 2019). Agriculture, especially annual crops interspersed with temporary and permanent grasslands, represents the major land use. Brittany has a dense hydrographic network and natural wetlands.

The relief is relatively gentle and strongly correlated with geological formations, with elevations of 0-382 m (Figure 1). Brittany is part of the Armorican Massif, which has complex geology: sedimentary rocks, rocks metamorphosed to differing degrees (sandstone, schist), metamorphic rocks (gneiss), igneous rocks (granite), and loess deposits. This high geological diversity generates a wide range of soils. According to the World Reference Base of Soil Resources (IUSS Working Group WRB, 2014), the main soils include Cambisols, Stagnic Fluvisols, Hisotsols Podzols, Luvisols, and Leptosols.

191 Soils are organized along toposequences: relatively well-drained deep soils in upslope and
192 plateau locations, shallower well-drained soils on mid-slopes, and soils with marked
193 redoximorphic features in valleys. (**Insert Figure 1**)

194 **2.2. Regional soil database at 1:250 000 scale**

195 Soil mapping in Brittany is represented in a regional geographic database called the
196 "Référentiel Régional Pédologique" (RRP) available at 1:250 0000 scale (INRA Infosol,
197 2014). It contains a set of polygons with crisp boundaries, commonly called soil map units
198 (SMUs), which are defined as areas with homogeneous soil-forming factors, such as
199 morphology, geology, and climate. Each SMU contains known proportions of several
200 STUs, each of which is described in the RRP by a set of strata that describe the vertical
201 variation in the soil. The strata are spatial horizons describing the vertical organisation of
202 STUs. Pedological features of SMU, STU, and strata, including depth, thickness, SOC
203 content, ECEC, pH, and five particle-size fractions, are contained in a relational database
204 called DoneSol (INRA Infosol, 2014). Furthermore, each STU has a representative soil
205 profile with a full pedological description. These profiles, which cover a wide pedological
206 diversity in Brittany, are contained in a separate database. In Brittany, 341 SMUs, 320
207 STUs, and 1020 strata are currently defined in 1984 polygons.

208 **2.3. Assessing and mapping soil properties at the regional scale**

209 2.3.1. Disaggregation of complex mapping units

210 The first step of the workflow (Figure 2) was performed in previous studies. Using the
211 DSMART algorithm (Odgers et al., 2014) and soil-landscape relationships, Vincent et al.
212 (2018) disaggregated the existing legacy 1:250 000 soil map of Brittany at a resolution of
213 50 m. Disaggregation yielded estimates of the probability of occurrence of the 320 STUs
214 in each pixel. Finally, Vincent et al. (2018) retained only the three STUs with the highest

probabilities of occurrence for each pixel and produced soil-type grids for the region indicating the first, second, and third most probable STU. (**Insert Figure 2**)

2.3.2. Modelling depth variation using spline functions

Because STU strata were described according to morphological features, they have variable upper and lower depths. Strata were vertically interpolated to express vertical variation in soil properties of each STU for GlobalSoilMap standard depth intervals by applying equal-area spline functions (Bishop et al., 1999; Malone et al., 2009) using the GSIF package (Hengl et al., 2013) of R software (R Core Team, 2019). The equal-area spline function respects the mean of the target soil property and ensures continuous variation with soil depth (Malone et al., 2009). Its result is a set of interpolated values that reflect the target property's mean for regular depth intervals down to 200 cm.

To avoid spline-function drawbacks, which consist of overestimating error at the extremes of a given STU, we followed previous recommendations (Odgers et al., 2012; Vaysse et al., 2015) and added two pseudo-strata to each STU. The first was created by subdividing the first strata into two elementary strata, which created a thin topsoil layer from 0-3 cm. The second was added from the bottom strata down to 200 cm, but only for STUs that had no strata down to 200 cm but whose depth could reach 200 cm. When present, the bottom pseudo- strata had the same characteristics as the strata just above it. More importantly, STU depth can be defined as the distance from the soil surface to the contact with bedrock (lithic) or weathered rock (paralithic), which contains fissures that make it easier for roots to penetrate the bedrock (Soil Survey Staff, 1993).

2.3.3 Treatment of censored soil dataset

In the present study, STU depth was determined using a previously developed classification tree (Styc and Lagacherie, 2019, Vaysse et al., 2015) applied to the bottom horizon of the representative soil profile of each STU. In the classification tree, each

horizon is classified as “censored”, “may be censored”, or “not censored” (Figure 3), the first two of which mean that the bottom horizon could be duplicated down to 200 cm. In general, the bottom horizon was classified as “not censored” when it had lithic or paralithic contact (R, M, F, C/R, C/M, C/D horizons) and “censored” when was identified as a pedogenetic horizon (A, B, E horizons). Since ambiguities in the C horizon mean that it cannot be directly classified as “may be censored” or “not censored”, the tree applies pedological rules that assess additional soil properties such as weathering, internal structure, classification, and compactness. The tree thus uses a binary approach to assess the C horizon and classifies it as “may be censored” when at least one of three rules is satisfied. (**Insert Figure 3**)

2.3.3. Correlation analysis among soil properties

Statistical analyses were performed to assess the degree to which fitting spline functions preserved soil-property relationships. As far as we are aware, this is the first time that preservation of soil attribute relationships using spline functions has been investigated for legacy soil maps. First, a Pearson correlation matrix was calculated to demonstrate the degree of interaction between soil properties before and after fitting spline functions. The matrix was calculated for two soil datasets: (i) all STU strata identified in the DoneSol database (1020 strata) and (ii) soil properties estimated for the six regular depth intervals down to 200 cm (1710 horizons). A two-tailed parametric t-test was also performed to assess the significance of correlation coefficients. Next, a normalized principal component analysis (PCA) was performed on both original and splined soil-property values to visualize organization of soil data and correlations among target soil properties simultaneously. The main advantage of this statistical method was that the relative relationships among soil attributes was preserved and all variables were projected onto a factorial plan.

265 2.3.4. Prediction of a composite soil variable (ECEC)

266 Two pedotransfer functions were calibrated using soil property values before and after
267 fitting spline functions to predict a composite soil variable, the ECEC, as a function of
268 soil attributes. The main idea was to check the stability of pedotransfer functions,
269 determined using stepwise multiple linear regression (Hocking, 1976). Both models were
270 selected using the Akaike information criterion, which allowed selecting the best set of
271 soil variables to predict the ECEC. This approach assessed the ability of the splined soil
272 data to accurately predict a composite soil attribute that depends strongly on explanatory
273 soil variables. Overall, the method achieved three objectives simultaneously by comparing
274 (i) the performance of models that best predict a target variable (ECEC) as a function of
275 additional soil properties for both soil datasets, (ii) selected explanatory soil properties
276 before and after fitting spline functions, and (iii) the partial coefficient of each soil
277 property for both models based on their associated confidence intervals.

278 2.3.5. Digital soil mapping of soil properties

279 The first step of the workflow in this study standardized the depth of the STU
280 characteristics to be expressed for six regular depth intervals down to 200 cm. By using
281 the three most probable STU maps, with their associated probabilities of occurrence at a
282 50 m spatial resolution (Ellili et al., submitted; Vincent et al., 2018), and STU soil-
283 property estimates for six regular depth intervals, several soil-property maps were
284 generated to estimate the spatial pattern of each soil property in Brittany. To this end, the
285 weighted mean of each soil attribute was calculated for each grid cell based on reference
286 soil-property values of the three most probable STU, and their associated probabilities of
287 occurrence were used as weights, as follows (Eq 1):

288 $$\hat{y}\,(x_{ij}) = \sum_{k=1}^{3}\left(P_{STU_{kj}}, x_{ij}\right) * y(STU_{kj,}\ x_{ij}) / \sum_{k=1}^{3}(P_{STU_{kj}}, x_{ij}) \qquad (Eq\ 1)$$

where $\hat{y}(x_{ij})$ is the predicted soil property for grid cell (x_{ij}) and soil depth j=1, 2, 3, 4, 5, 6; $y(STU_k, xi)$ is the reference soil property estimated at soil depth j associated with STU k=1, 2, 3 predicted at all (x_{ij}); and $(P_{STU_{kj}}, x_{ij})$ is the probability of occurrence of STU_k in the given grid cell (x_{ij}).

The workflow produced maps of nine soil properties – SOC; particle-size distribution of fine sand, coarse sand, fine silt, coarse silt, and clay; coarse fragments; pH; and ECEC – at a fine spatial resolution of 50 m over the entire area from legacy soil maps.

2.3.6. Spatial correlation of soil properties

To analyze the spatial correlation of disaggregated soil properties, we calculated experimental semivariograms of each predicted variable for the six standard depth intervals. To this end, we first generated a grid with a resolution of 50 m covering the entire area and randomly selected 30 000 cells from it using the Spatially Balanced sampling tool of ArcGis 10.6 GIS software (ESRI, 2012). Next, soil-property estimates were extracted for each cell of the selected dataset using the multiple extraction tool of ArcGIS. Before semivariogram analysis, SOC content was log-transformed, and ECEC was square-root-transformed. By applying the variogram function in the R package "gstat" (Pebesma, 2004), several experimental semivariograms were generated for all soil properties and all standard depth intervals to characterise the spatial pattern of the soil-property maps produced.

3. Results

3.1. Spline-function outputs

As an example of spline-function outputs, the equal-area splines fit to strata silt content for four STUs depended strongly on the initial silt content of each strata (Figure 4). Abrupt changes in content between adjacent strata influenced the shape of the spline curve

considerably. However, the spline curves were constrained at the top and bottom strata because adding the two thin pseudo-strata minimized overfitting. (**Insert Figure 4**)

3.2. Correlation between soil properties before and after fitting smoothing spline functions

For both Pearson correlation matrices, well-known correlations between soil properties were significant (Table 1). For instance, SOC content correlated positively with clay content and ECEC but negatively with pH and sand content ($P < 0.001$). The correlation between ECEC and clay content was also positive ($P < 0.001$). In contrast, pH was weakly but not significantly correlated with ECEC and clay content. The correlation matrices showed similar correlation coefficients between soil properties in both the original and splined datasets, indicating that applying spline depth functions to pedological strata preserved soil attribute correlations.

3.3. Multicollinearity among soil properties

The first two dimensions of the PCA performed on 1020 strata features explained 57% of the total variance (Figure 5a). The main contributing variables were sand content (30%) and total silt content (25%) on the first dimension and clay content (26%), CEC (22%), and SOC content (15%) on the second. Similarly, the first two dimensions of the PCA performed on 1710 standard horizon features explained 58% of the total variance (Figure 5b). The main contributing variables were sand content (30%) and total silt content (23%) on the first dimension and SOC content (15%), total silt content (17%), and ECEC (17%) on the second. Thus, the first PCA dimension for the two datasets (34.3% and 34.2%, respectively) depended on soil-texture fractions, contrasting coarse texture and relatively fine texture. The second PCA dimension for the two datasets (22.7% and 23.3%, respectively) depended on physico-chemical properties and fine texture, particularly clay content. (**Insert Figure 5**)

Table 1 : Pearson's correlation matrices between soil properties before (1020 strata) and after fitting spline functions (1710 standard horizons). SOC = soil organic carbon, CEC = cation exchange capacity. * P < 0.05; ** P < 0.01; *** P < 0.001.

Dataset	Property	SOC	Sand	Clay	Silt	Coarse fragments	Fine silt	Coarse silt	pH
Before	Soil organic carbon	1							
	Sand	-0.17***	1						
	Clay	0.26***	-0.53***	1					
	Silt	0.03	-0.83***	-0.04	1				
	Coarse fragments	-0.11**	0.33***	-0.12**	-0.30***	1			
	Fine silt	0.10*	-0.71***	0.33***	0.62***	-0.08	1		
	Coarse silt	-0.04	-0.51***	-0.30***	0.80***	-0.32***	0.02	1	
	pH	-0.14**	-0.07	0.01	0.07	-0.25***	-0.21***	0.25***	1
	CEC	0.48***	-0.26***	0.48***	-0.01	-0.17***	0.12**	-0.11**	0.04
After	Soil organic carbon	1							
	Sand	-0.19***	1						
	Clay	0.30***	-0.55***	1					
	Silt	0.03	-0.82***	-0.02	1				
	Coarse fragments	-0.13***	0.31***	-0.10***	-0.30***	1			
	Fine silt	0.13***	-0.71***	0.39***	0.59***	-0.11***	1		
	Coarse silt	-0.05	-0.49***	-0.31***	0.80***	-0.29***	-0.02	1	
	pH	-0.15***	0.01	-0.05	0.02	-0.21***	-0.26***	0.22***	1
	CEC	0.40***	-0.28***	0.50***	0.00	-0.21***	0.18***	-0.14***	0.01

3.4. Stepwise multiple linear regression

The best regression model for the 1020 strata features (before fitting spline functions) contained SOC content ($P < 0.001$), clay content ($P < 0.001$), and pH ($P < 0.01$) and explained 37% (multiple R^2) of variation in ECEC (Adj. $R^2 = 0.36$, $F = 193$, $P < 0.005$). The best regression model for the 1710 standard horizons (after fitting spline functions) also contained SOC content, clay content, and pH (all ($P < 0.001$) and explained 33% of variation in ECEC (Adj. $R^2 = 0.32$, $F = 213$, $P < 0.001$). The 90% confidence intervals of

both models appeared similar, which confirms that fitting spline functions to pedological strata information preserved the relationships among soil attributes (Table 2). For instance, for SOC content, the 90% confidence interval before and after fitting spline functions was 0.42-0.52 and 0.21-0.34 g kg^{-1}, respectively.

Table 2 : Regression coefficients of effective cation exchange capacity (ECEC) regression models before and after fitting spline functions. SOC = soil organic carbon. * P < 0.05; ** P < 0.01; *** P < 0.001

Model	Variable	Estimate	Standard error	t-value	Significance	5% limit	95% limit
Before splines	Constant	-3.33	1.34	-2.48	0.01*	-4.55	-1.11
	SOC (g kg^{-1})	0.47	0.03	14.80	<0.05***	0.42	0.52
	Clay (%)	0.33	0.02	14.38	<0.05***	0.30	0.37
	pH	0.71	0.22	3.15	0.001**	0.34	1.08
After splines	Constant	-2.3	0.67	-1.02	0.30	-2.61	0.62
	SOC (g kg^{-1})	0.27	0.02	13.74	<0.05***	0.21	0.34
	Clay (%)	0.33	0.01	17.56	<0.05***	0.30	0.37
	pH	0.55	0.15	3.47	<0.005***	0.29	0.82

3.5. Soil-property mapping at multiple depths

Depth functions successfully preserved relationships among physico-chemical soil properties. This implies the reliability of soil-property estimates and their spatial patterns over the study area; the estimates remained highly informative and reflected information in the legacy soil maps.

In soil-property maps, SOC content tended to decrease down the soil profile across most of Brittany, being higher in surface layers (0-5 and 5-15 cm) than in subsurface layers (60-100 and 100-200 cm) (Figure 6). Furthermore, the spatial pattern of SOC content showed a strong increase from east to west and north to south. Overall, soils not used for annual crops – located in permanent grasslands, wooded areas, and along watercourses –

367 tended to have higher SOC content than those used for annual crops. In surface layers (0-
368 5 and 5-15 cm), cropping areas (northern, western, and central Brittany) had the lowest
369 SOC content (usually 10-40 g kg^{-1}), while grasslands and wooded areas had SOC content
370 in excess of 150 g kg^{-1}.
371 Likewise, silt content always decreased as depth increased, except for soils developed
372 from Aeolian loam deposits in northern Brittany (Figure 7). This trend was also related to
373 the depth of these soils, which did not exceed 100 cm. For surface layers (0-30 cm), silt
374 content usually ranged from 30-60% but exceeded 70% for soils developed from Aeolian
375 loam, which coincide with areas of annual crops. For subsurface layers, particularly 30-
376 60 cm and 60-100 cm, the lowest silt contents (< 30%) were located in areas of natural
377 vegetation in south-western and north-central Brittany. **(Insert Figure 6), (Insert Figure**
378 **7)**
379 Overall, the disaggregated data covered a wide range of soil property values (Table 3).
380 For instance, median SOC content decreased from 30 g kg^{-1} at 0-5 cm to 4 g kg^{-1} at 100-
381 200 cm. Most soil properties followed the same trend, except for sand content, and coarse
382 fragments, which increased down soil profiles.
383
384
385
386
387
388
389
390
391

392 **Table 3 :** Descriptive statistics of the disaggregated soil maps of the soil properties
393 studied.

Soil property	Depth interval (cm)	Min	1st quantile	Median	Mean	3rd quantile	Max
Soil organic carbon ($g\ kg^{-1}$)	0-5	3	20	30	32	37	761
	5-15	4	19	28	30	34	663
	15-30	2	19	25	28	32	789
	30-60	0	7	10	13	14	807
	60-100	0	3	4	6	5	583
	100-200	0	2	4	5	7	686
Clay (%)	0-5	2	15	17	18	21	65
	5-15	1	16	17	18	21	66
	15-30	1	15	17	18	20	68
	30-60	1	14	16	16	19	69
	60-100	2	15	17	18	21	65
	100-200	1	12	16	16	19	59
Sand (%)	0-5	1	23	28	32	43	88
	5-15	0	23	29	32	41	88
	15-30	0	23	29	32	42	90
	30-60	2	23	31	33	42	99
	60-100	0	23	29	32	41	88
	100-200	2	22	34	33	42	94
Silt (%)	0-5	5	41	50	50	57	73
	5-15	4	42	50	50	57	74
	15-30	4	42	51	50	57	74
	30-60	2	43	50	50	58	77
	60-100	1	33	44	44	55	79
	100-200	1	34	43	42	49	75
Fine silt (%)	0-5	1	21	25	26	30	45
	5-15	1	21	25	25	29	45
	15-30	1	21	25	26	30	44
	30-60	1	21	24	25	30	40
	60-100	1	17	21	22	25	41
	100-200	1	17	21	20	23	42
Coarse silt (%)	0-5	3	20	22	25	26	64
	5-15	3	20	22	24	26	65
	15-30	2	20	22	25	26	65
	30-60	0	20	23	25	26	63
	60-100	1	16	20	22	25	63
	100-200	1	16	21	22	24	58
Coarse fragments (%)	0-5	0	3	6	7	9	55
	5-15	0	3	5	6	8	55
	15-30	0	3	6	7	10	58
	30-60	0	4	8	11	15	83
	60-100	0	6	10	13	15	92
	100-200	0	5	9	11	14	98
Cation exchange capacity ($cmol^{+}kg^{-1}$)	0-5	2	10	12	12	14	69
	5-15	2	10	13	13	15	39
	15-30	2	10	11	11	13	40
	30-60	1	4	5	6	7	32
	60-100	1	3	4	4	5	95
	100-200	1	3	5	5	5	58
pH	0-5	4	6	6	6	6	9
	5-15	4	6	6	6	6	9
	15-30	4	6	6	6	6	9
	30-60	4	5	6	6	6	9
	60-100	4	5	5	5	6	9
	100-200	4	5	5	5	6	9

Overall, the experimental semivariograms highlighted significant differences among soil properties and soil-depth intervals. For all soil properties, subsurface layers, particularly 60-100 and 100-200 cm, varied more than surface layers (0-5, 5-15 and 15-30 cm), which had nearly the same semivariogram shape (Figure 8). Furthermore, subsurface layers had a "nugget effect" (i.e. value of the semivariogram at zero lag size) twice as large as those of surface layers. For instance, the nugget effect for sand content was nearly 50 $(g\ kg^{-1})^2$ at 0-5, 5-15, and 30-60 cm and reached 100 $(g\ kg^{-1})^2$ in the bottom layer (100-200 cm). In contrast, the range of the semivariogram varied by soil property and depth interval and was a mean of 30 000 m. The sill:nugget ratio also varied greatly by soil property and depth interval (e.g. for SOC content, from 2.0 for surface layers (0-60 cm) to 2.5 for the bottom layer). **(Insert Figure 8)**

4. Discussion

4.1. Equal-area spline method: advantages and limitations

Reconstructing STU information as a function of depth by fitting spline functions to soil-property values was a convenient way to estimate soil attributes for regular depth intervals. Several studies (Malone et al., 2009; Lacoste et al., 2014; Odgers et al., 2012; Stycs and Lagacherie 2019; Vaysse et al., 2014) have used this method to standardize many soil properties by depth. In the present study, it effectively preserved existing relationships between all soil properties, as well as the pedological information. Due to their flexibility, the quadratic polynomial spline functions captured local variation in each soil interval (Bishop et al., 1999; Ponce-Hernandez et al., 1986). More importantly, equal-area splines predict several soil properties simultaneously for a set of common depth intervals. These predictions are similar for a given soil depth and reflect specific pedological process. Soil-property estimates can be used in environmental models and

418 decision-making tools to derive consistent proxies of ecosystem functions and help guide
419 stakeholders' choices.
420 The equal-area spline method estimated variation in a given soil property with depth
421 satisfactorily for many kinds of soil profiles, but it also had some limitations. For instance,
422 soil properties (e.g. SOC content, clay content) of some soil profiles changed abruptly
423 with depth, which sometimes generated negative values at points close to the abrupt
424 boundary. The simplest solution was to insert two thin pseudo-strata at the top and bottom
425 strata to restrict the abrupt transition in the spline function. Despite its potential
426 effectiveness, most studies do not follow this approach because it requires additional time
427 and effort to harmonize the input data.

428 **4.2. Spatial structure of disaggregated soil-property maps**

429 The disaggregated digital soil maps depicting soil properties for all depth intervals were
430 generally consistent with landscape features, including land use. For instance, SOC
431 content peaked in wooded areas and permanent grasslands but was lowest in cropping
432 areas. More importantly, Histosols, especially in north-western Brittany, showed high
433 SOC content in deeper soil layers. Thus, SOC content was influenced mainly by structural
434 factors, such as parent material, topography, hydromorphic soil conditions, and some
435 management practices (e.g., soil tillage, crop cover, and land use (Ellili et al., 2019)).
436 As expected, the spatial pattern of silt content highlighted well the geological and
437 pedological knowledge about soils in Brittany (Lacoste et al., 2011). Silt content increased
438 with depth for soils developed in deep Aeolian loam deposits, especially in north-central
439 Brittany, and was lowest in soils in the north and south, where soils are derived from
440 granite. Predictions of sand content and coarse fragments by soil-depth interval also
441 appeared to be consistent with existing knowledge about the study area, particularly the
442 map of predicted soil parent material created in a previous study (Lacoste et al., 2011).

We found high variability in the spatial pattern of each soil attribute studied, all of which had a high nugget effect. According to the geostatistical literature (Journel and Huijbregts, 1978), the nugget effect is related to measurement errors or spatial variation at distances smaller than the sampling interval, or both. However, the experimental semivariograms showed higher variability for deeper layers than surface layers, which can be explained mainly by the degree of weathering of soil parent material, which varies greatly among the contrasting parent materials in Brittany.

4.3. Uncertainties of disaggregated soil maps: legacy soil data and disaggregation procedure limitations

In a previous study (Ellii-Bargaoui., 2019), the accuracy of generated soil property maps was assessed using an independent validation soil dataset. The validation procedure was performed at two soil depth intervals (5-15 and 30-60 cm) using 260 soil samples. In general, soil property predictions were unbiased except for coarse fragments and CEC in the 5-15 cm layer. Validation statistics (R^2, RMSE, RRMSE and ME) were better for the 30-60 cm layer except for soil particle-size distribution.

Overall, the major part of uncertainties was related to the performance of the disaggregation procedure (DSMART with soil landscape relationships). Splines functions also explains some of these uncertainties, but it remains marginal compared to the uncertainties generated by the disaggregation procedure. Moreover, the legacy soil data presents some limitations which amplify these uncertainties, Indeed, STUs were not sampled equally across the landscape, which leads to unequal representation of soil samples among STUs in the model calibration soil dataset, especially when some STUs are more abundant than others. In addition, differences in sampling designs followed to sample legacy soil profiles may have influenced the determination of SMU proportions. For instance, some projects sampled soil profiles to study redoximorphic soil conditions,

468 while other projects did so to study agricultural soils. Finally, the decrease in data
469 availability with depth could also have biased our predictions, especially those for deeper
470 soil intervals.

471 5. Conclusion and perspectives

472 We developed a method to produce soil-property maps from legacy soil data for the
473 regular intervals defined by GlobalSoilMap specifications. By combining spline functions
474 and the DSMART algorithm, it achieved continuity in both lateral and vertical variation
475 in soil properties. Moreover, our predictions for each depth interval were consistent with
476 the geological and pedological heterogeneity across the landscape. However, some
477 challenges remain beyond creation of the soil-property maps themselves, particularly the
478 assessment of uncertainties in soil-property estimates. Therefore, estimating prediction
479 intervals of target soil properties and their potential distributions worthwhile to be
480 investigated in the future as earlier done by Odgers et al., (2015a). This could allow to
481 distinguish different sources of uncertainties generated by the implemented
482 disaggregation approach, fitting spline functions procedure and inherent uncertainties of
483 the legacy soil dataset.

484 **Acknowledgments**

485 This research was performed in the framework of the INRA “Ecoserv” metaprogram. This
486 work was also supported by Sols de Bretagne project and Soilserv program funded by
487 ANR (Agence Nationale de la Recherche) (ANR-16- CE32-0005-01).

Figure captions

Figure 1 : Location of the Brittany study area, showing boundaries of soil map units (SMU) and elevation.

Figure 2 : The workflow used to derive disaggregated soil-property maps at the regional scale. STU = soil typology unit.

Figure 3 : The classification tree used to characterise lithic and paralithic horizons. STU = soil typological unit (adapted from Vaysse, 2015).

Figure 4 : Strata silt content (open black rectangles), equal-area spline curves (red), and estimated mean silt content for GlobalSoilMap regular soil-depth intervals (solid grey rectangles) for four soil typological units

Figure 5 : Principal component analysis based on physico-chemical and texture characteristics of soil typological unit strata (a) before and (b) after fitting spline functions. SOC: soil organic carbon content, ECEC: effective cation exchange capacity, sand: sand content, CF: coarse fragments, F-silt: fine silt, C: silt: coarse silt, and silt: silt content.

Figure 6 : Disaggregated maps of soil organic carbon (SOC) content in Brittany at the six soil-depth intervals.

Figure 7: Disaggregated maps of silt content in Brittany for the six soil-depth increments.

Figure 8 : Experimental semivariograms derived from disaggregated soil maps for seven soil properties for six regular soil-depth intervals.

References

Arrouays, D., Grundy, M.G., Hartemink, A.E., Hempel, J.W., Heuvelink, G.B.M., Hong, S.Y., Lagacherie, P., Lelyk, G., McBratney, A.B., McKenzie, N.J., Mendonca-Santos, M.D.L.,

Minasny, B., Montanarella, L., Odeh, I.O.A., Sanchez, P.A., Thompson, J.A., Zhang, G.L., 2014. GlobalSoilMap: toward a fine-resolution global grid of soil properties. In: Sparks, D.L. (Ed.), Advances in Agronomy vol 125. Elsevier Academic Press Inc, San Diego, pp. 93–134.

Bishop, T.F.A., Horta, A., Karunaratne, S.B., 2015. Validation of digital soil maps at different spatial supports. Geoderma 241–242, 238–249. https://doi.org/10.1016/j.geoderma.2014.11.026

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. https://doi.org/10.1016/S0016-7061(99)00003-8

Bui, E.N., Moran, C.J., 2001. Disaggregation of polygons of surficial geology and soil maps using spatial modelling and legacy data. Geoderma 103, 79–94. https://doi.org/10.1016/S0016-7061(01)00070-2

Chaney, N.W., Wood, E.F., McBratney, A.B., Hempel, J.W., Nauman, T.W., Brungard, C.W., Odgers, N.P., 2016. POLARIS: A 30-meter probabilistic soil series map of the contiguous United States. Geoderma 274, 54–67. https://doi.org/10.1016/j.geoderma.2016.03.025

Climatik., 2019. Climatik databases, INRA.

Cook, S., Corner, R., Groves, P., Grealish, G.: Use of airborne gamma radiometric data for soil mapping. Soil Research 34, 183. https://doi.org/10.1071/SR9960183, 1996.

Ellili-Bargaoui, Y., Malone, B. P., Michot, D., Minasny, B., Vincent, S., Walter, C., and Lemercier, B. 2020. Comparing three approaches of spatial disaggregation of legacy soil maps based on the Disaggregation and Harmonisation of Soil Map Units Through Resampled Classification Trees (DSMART) algorithm. SOIL, 6, 371–388. https://doi.org/10.5194/soil-6-371-2020.

Ellili Bargaoui, Y., Walter, C., Michot, D., Saby, N.P.A., Vincent, S., Lemercier, B., 2019. Validation of digital maps derived from spatial disaggregation of legacy soil maps. Geoderma 356, 113907. https://doi.org/10.1016/j.geoderma.2019.113907

Ellili, Y., Walter, C., Michot, D., Pichelin, P., Lemercier, B., 2019. Mapping soil organic carbon stock change by soil monitoring and digital soil mapping at the landscape scale. Geoderma 351, 1–8. https://doi.org/10.1016/j.geoderma.2019.03.005.

Erh, K.T., 1972. Application of spline functions to soil science. Soil Science 114, 333–338.

ESRI, 2012. ArcMap 10.1. Environmental Systems Resource Institute, Redlands, California

Hengl, T., Kempen, B., Heuvelink, G., Malone, B., Hannes, R., 2013. GSIF: Global Soil Information Facilities. R package version 0.3–2.

Hocking, R., 1976. A Biometrics Invited Paper. The Analysis and Selection of Variables in Linear Regression. Biometrics, 32.1, 1-49. doi:10.2307/2529336

Holmes, K.W., Griffin, E.A., Odgers, N.P., 2015. Large-area spatial disaggregation of a mosaic of conventional soil maps: evaluation over Western Australia. Soil Research 53, 865. https://doi.org/10.1071/SR14270.

INRA Infosol, 2014. Donesol Version 3.4.3. Dictionnaire de données

IUSS Working Group WRB: World reference base for soil resources 2014. World Soil Resources Reports No. 103. FAO, Rome,116 pp.

Jamshidi, M., Delavar, M.A., Taghizadehe-Mehrjerdi, R., Brungard, C: Disaggregation of conventional soil map by generatingmulti realizations of soil class distribution (case study:Saadat Shahr plain, Iran). Environ Monit Assess, 191-769, https://doi.org/10.1007/s10661-019-7942-x, 2019.

Jenny, H., 1941. Factors of Soil Formation: a System of Quantitative Pedology. McGraw-Hill, New York and London.

Journel, A. G. and Huijbregts, Ch. J., 1979. Mining Geostatistics. London & New York (Academic Press), 1978. x 600 pp., 267 figs. Mineralogical Magazine, 43(328), 563-564. doi:10.1180/minmag.1979.043.328.34

Lacoste, M., Lemercier, B., Walter, C., 2011. Regional mapping of soil parent material by machine learning based on point data. Geomorphology 133, 90–99. https://doi.org/10.1016/j.geomorph.2011.06.026

Lacoste, M., Minasny, B., McBratney, A., Michot, D., Viaud, V., Walter, C., 2014. High resolution 3D mapping of soil organic carbon in a heterogeneous agricultural landscape. Geoderma 213, 296–311. https://doi.org/10.1016/j.geoderma.2013.07.002

Malone, B.P., McBratney, A.B., Minasny, B., Laslett, G.M., 2009. Mapping continuous depth functions of soil carbon storage and available water capacity. Geoderma 154, 138–152. https://doi.org/10.1016/j.geoderma.2009.10.007

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

Minasny, B., McBratney, A.B., Mendonca-Santos, M.L., Odeh, I.O.A., Guyon, B., 2006.Prediction and digital mapping of soil carbon storage in the Lower Namoi Valley.Australian Journal of Soil Research 44, 233–244.

Minasny, B., McBratney, A.B., Lark, M.R., 2008. Digital Soil Mapping Technologies for Countries with Sparse Data Infrastructures. In: Hartemink, A.E., McBratney, A.B., Mendonca-Santos, M.L. (Eds.), Digital Soil Mapping with Limited Data. Springer Science, Australia, pp. 15–30.

Minasny, B., McBratney, A.B., 2010. Methodologies for Global Soil Mapping, in: Boettinger, J.L., Howell, D.W., Moore, A.C., Hartemink, A.E., Kienast-Brown, S. (Eds.), Digital Soil Mapping. Springer Netherlands, Dordrecht, pp. 429–436. https://doi.org/10.1007/978-90-481-8863-5_34

Minasny, B., McBratney, A.B., 2007. Spatial prediction of soil properties using EBLUP with the Matérn covariance function. Geoderma 140, 324–336. https://doi.org/10.1016/j.geoderma.2007.04.028

Minasny, B., McBratney, A.B., Malone, B.P., Wheeler, I., 2013. Digital Mapping of Soil Carbon, in: Advances in Agronomy. Elsevier, pp. 1–47. https://doi.org/10.1016/B978-0-12-405942-9.00001-3

Møller, A.B., Malone, B., Odgers, N.P., Beucher, A., Iversen, B.V., Greve, M.H., Minasny, B.: Improved disaggregation of conventional soil maps. Geoderma 341, 148–160. https://doi.org/10.1016/j.geoderma.2019.01.038, 2019.

Odgers, N., McBratney, A., Minasny, B., Sun, W., Clifford, D., 2014. Dsmart: An algorithm to spatially disaggregate soil map units, in: Arrouays, D., McKenzie, N., Hempel, J., de Forges, A., McBratney, Alex (Eds.), GlobalSoilMap. CRC Press, pp. 261–266. https://doi.org/10.1201/b16500-49

Odgers, N.P., Holmes, K.W., Griffin, T., Liddicoat, C., 2015a. Derivation of soil-attribute estimations from legacy soil maps. Soil Res. 53, 881. https://doi.org/10.1071/SR14274

Odgers, N.P., Libohova, Z., Thompson, J.A., 2012. Equal-area spline functions applied to a legacy soil database to create weighted-means maps of soil organic carbon at a continental scale. Geoderma 189–190, 153–163. https://doi.org/10.1016/j.geoderma.2012.05.026

Odgers, N.P., McBratney, A.B., Minasny, B., 2015b. Digital soil property mapping and uncertainty estimation using soil class probability rasters. Geoderma 237–238, 190–198. https://doi.org/10.1016/j.geoderma.2014.09.009

Pallegedara Dewage, S. N. S., Minasny, B., Malone, B.: Disaggregating a regional-extent digital soil map using Bayesian area-to-point regression kriging for farm-scale soil carbon assessment. SOIL,6,359-369. https://soil.copernicus.org/articles/6/359/2020/10.5194/soil-6-359-2020, 2020.

Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. Comput. Geosci. 30, 683–691

Ponce-Hernandez, R., Marriott, F.H.C., Beckett, P.H.T., 1986. An improved method for reconstructing a soil profile from analyses of a small number of samples. Journal of Soil Science 37, 455–467. https://doi.org/10.1111/j.1365-2389.1986.tb00377.x

R Core Team, 2019. R: A Language and Environment for Statistical Computing. **R. Foundation** for Statistical Computing, Vienna, Austria 3-900051-07-0. http://www.R-project.org.

Russell, J.S., Moore, A.W., 1968. Comparison of different depth weightings in the numerical analysis of anisotropic soil profile data. Transactions of the 9th International Congress of Soil Science 4, 205–213.

Soil Science Division Staff. 1993. Soil survey manual. C. Ditzler, K. Scheffe, and H.C. Monger (eds.). USDA Handbook 18. Government Printing Office, Washington, D.C.

Stoorvogel, J.J., Bakkenes, M., Temme, A.J.A.M., Batjes, N.H., ten Brink, B.J.E.: S-World: A Global Soil Map for Environmental Modelling. Land Degradation & Development 28, 22–33. https://doi.org/10.1002/ldr.2656, 2017.

Styc, Q., Lagacherie, P., 2019. What is the Best Inference Trajectory for Mapping Soil Functions: An Example of Mapping Soil Available Water Capacity over Languedoc Roussillon (France). Soil Syst. 3, 34. https://doi.org/10.3390/soilsystems3020034

Vaysse, K., Lagacherie, P., 2015. Evaluating Digital Soil Mapping approaches for mapping GlobalSoilMap soil properties from legacy data in Languedoc-Roussillon (France). Geoderma Regional 4, 20–30. https://doi.org/10.1016/j.geodrs.2014.11.003

Vincent, S., Lemercier, B., Berthier, L., Walter, C., 2018. Spatial disaggregation of complex Soil Map Units at the regional scale based on soil-landscape relationships. Geoderma 311, 130–142. https://doi.org/10.1016/j.geoderma.2016.06.006

Walter, C., Lagacherie, P., Follain, S.: Integrating pedological knowledge into digital soil mapping. In: Lagacherie, P., McBratney, A., Voltz, M. (Eds.), Digital Soil Mapping. An Introductory Perspective. Development in Soil Science vol. 31. Elsevier, pp. 289–310 (ISBN-13: 978-0-444-52958-9), 2006.

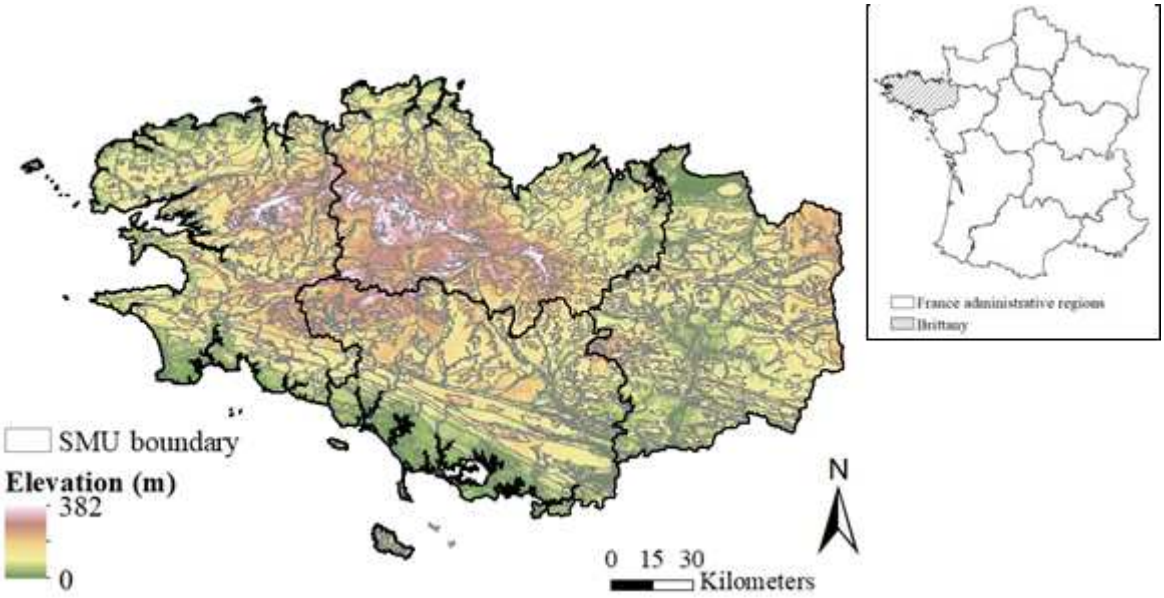

Figure 1 : Location of the Brittany study area, showing boundaries of soil map units (SMU) and elevation.

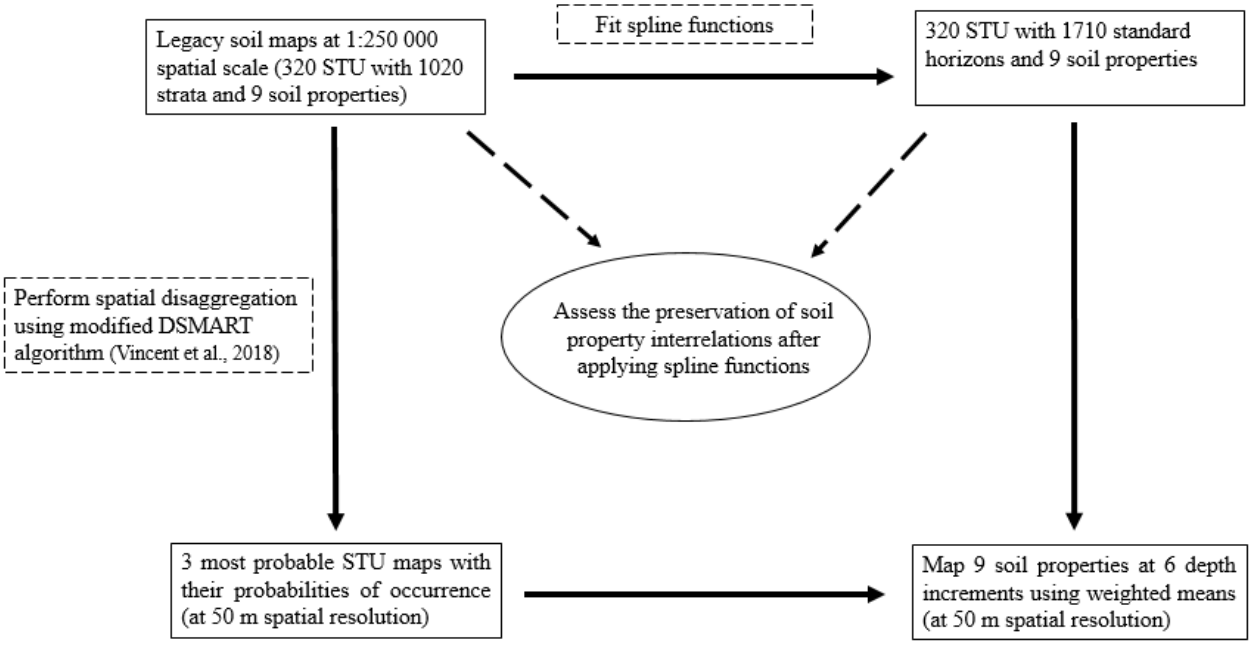

Figure 1 : The workflow used to derive disaggregated soil-property maps at the regional scale. STU = soil typology unit.

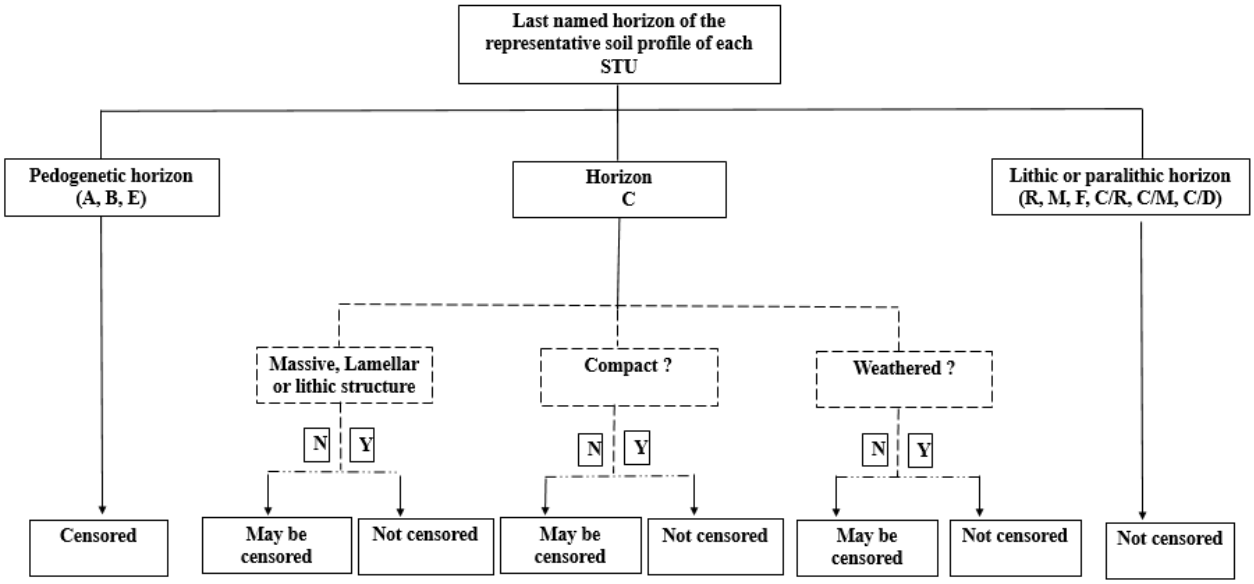

Figure 1 : The classification tree used to characterise lithic and paralithic horizons. STU = soil typological unit (adapted from Vaysse, 2015).

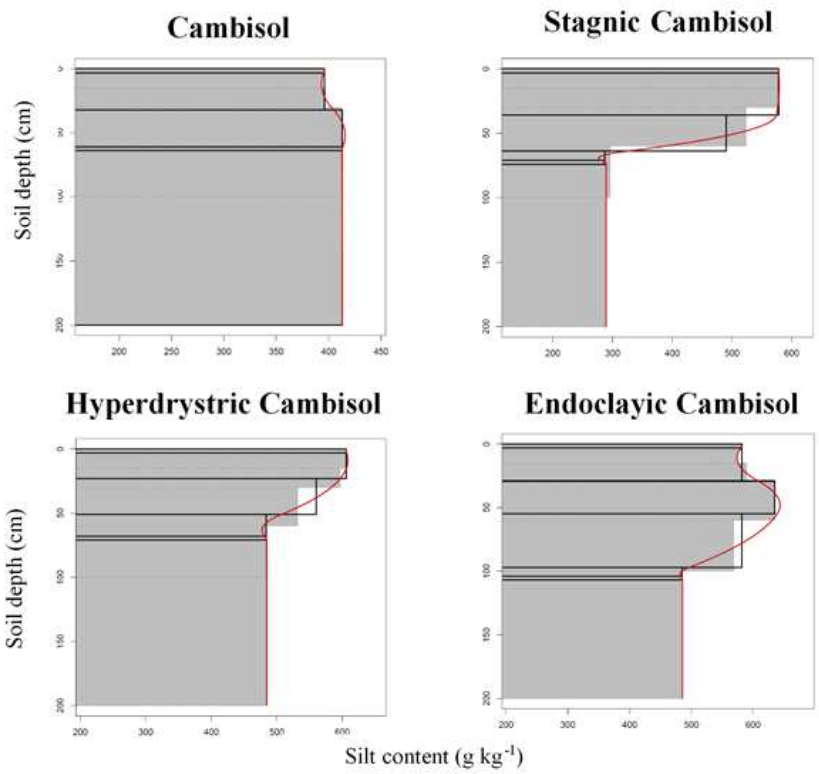

Figure 1 : Strata silt content (open black rectangles), equal-area spline curves (red), and estimated mean silt content for *GlobalSoilMap* regular soil-depth intervals (solid grey rectangles) for four soil typological units.

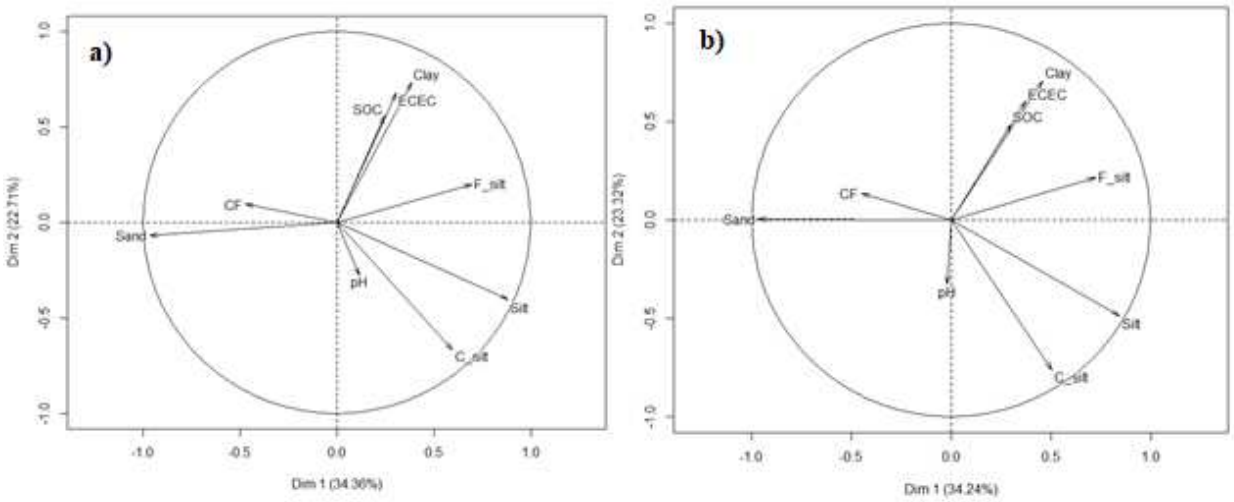

Figure 1 : Principal component analysis based on physico-chemical and texture characteristics of soil typological unit strata (a) before and (b) after fitting spline functions. SOC: soil organic carbon content, ECEC: effective cation exchange capacity, sand: sand content, CF: coarse fragments, F-silt: fine silt, C: silt: coarse silt, and silt: silt content.

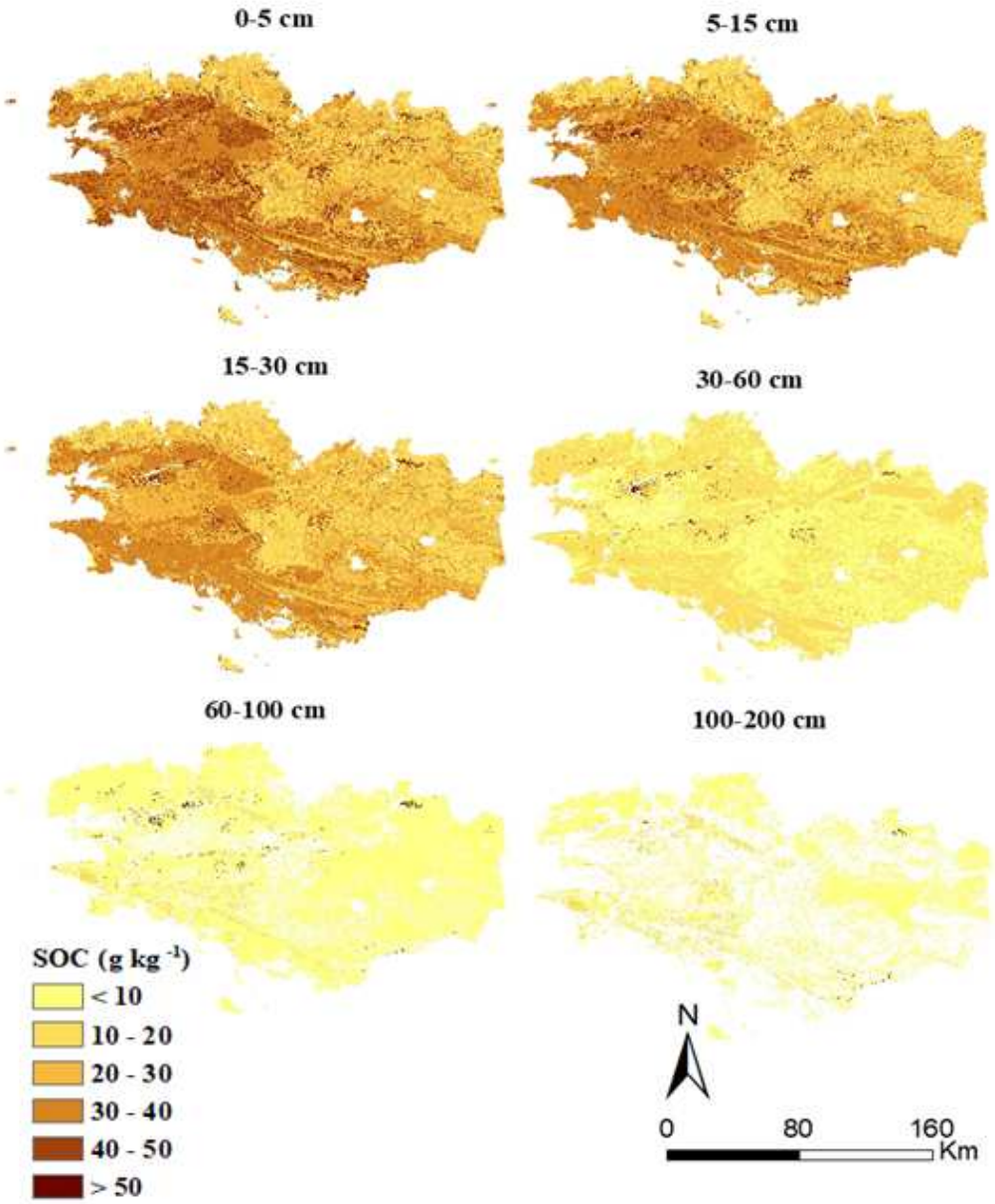

Figure 1 : Disaggregated maps of soil organic carbon (SOC) content in Brittany at the six soil-depth intervals.

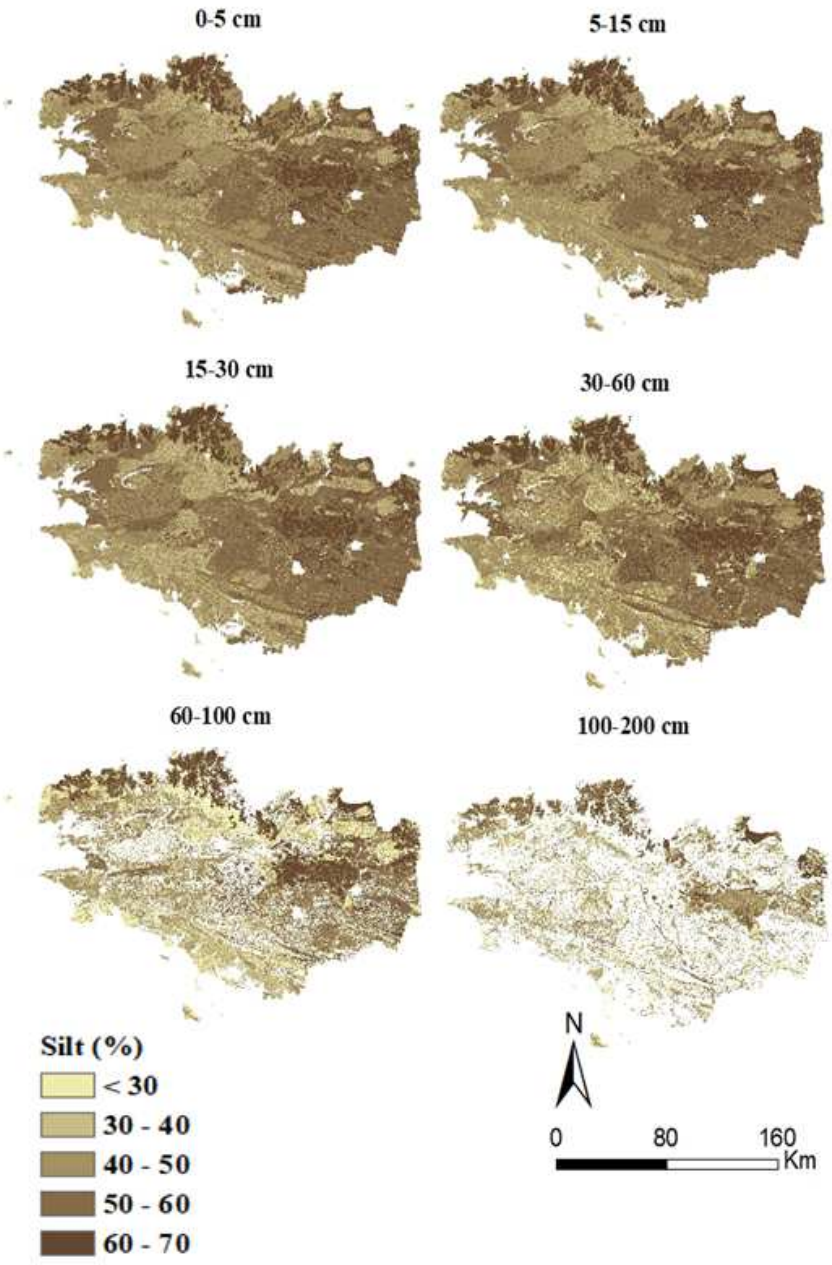

Figure 1: Disaggregated maps of silt content in Brittany for the six soil-depth increments.

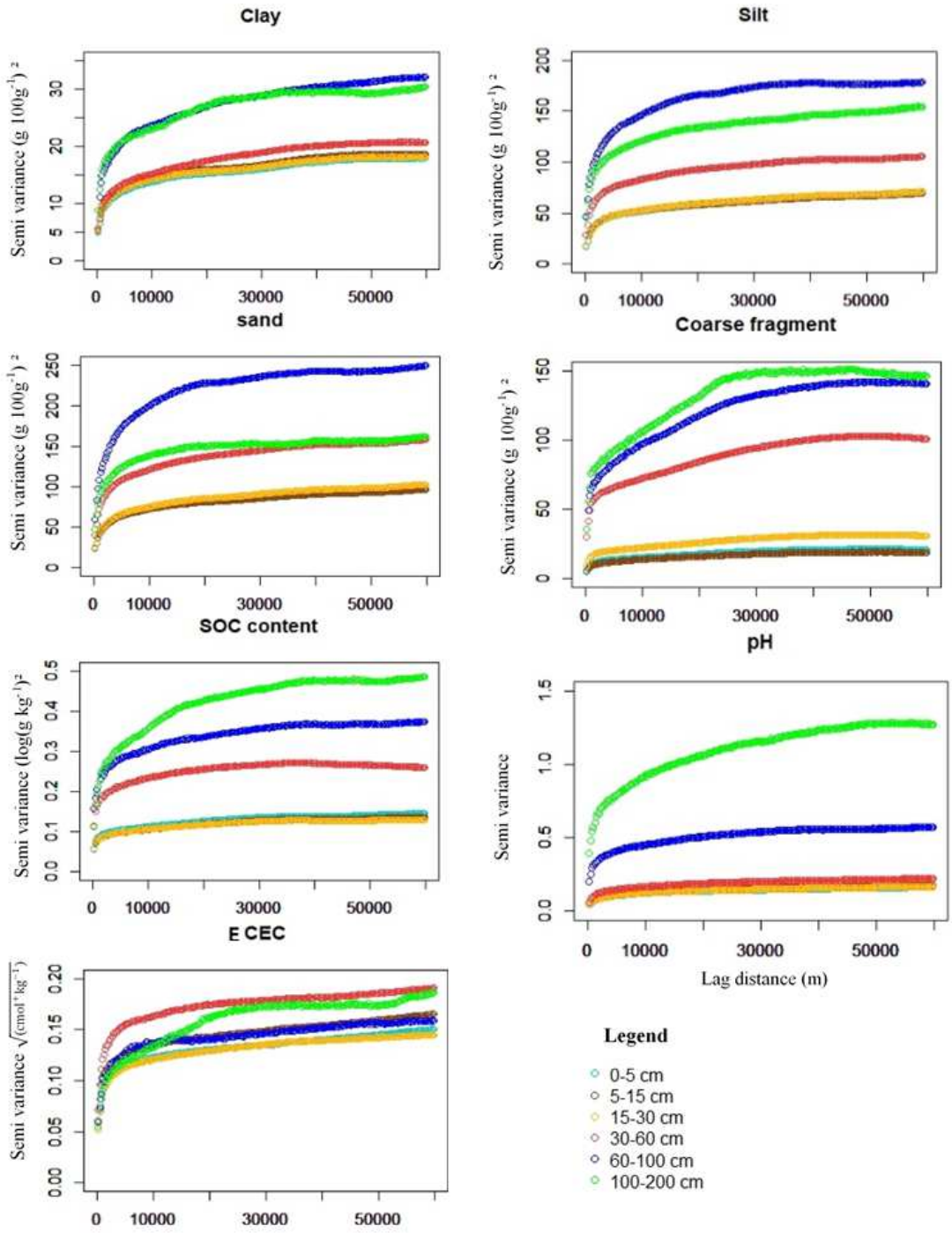

Figure 1 : Experimental semivariograms derived from disaggregated soil maps for seven soil properties for six regular soil-depth intervals.