



**HAL**  
open science

## Validation of digital maps derived from spatial disaggregation of legacy soil maps

Yosra Ellili Bargaoui, Christian Walter, Didier Michot, Nicolas P.A. Saby, Sébastien Vincent, Blandine Lemerancier

► **To cite this version:**

Yosra Ellili Bargaoui, Christian Walter, Didier Michot, Nicolas P.A. Saby, Sébastien Vincent, et al.. Validation of digital maps derived from spatial disaggregation of legacy soil maps. *Geoderma*, 2019, 356, pp.113907. 10.1016/j.geoderma.2019.113907 . hal-02283023

**HAL Id: hal-02283023**

<https://institut-agro-rennes-angers.hal.science/hal-02283023>

Submitted on 23 Aug 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

1 **Manuscript title: Validation of digital maps derived from spatial disaggregation of legacy**  
2 **soil maps**

3 **The author affiliation respect the following order: author name, author affiliation and**  
4 **author address**

5 Yosra Ellili, UMR SAS, INRA, AGROCAMPUS OUEST 35000 Rennes, France  
6 INRA, UMR1069 SAS,65 rue de Saint-Brieuc 35042 Rennes cedex, France  
7

8 Christian Walter, UMR SAS, AGROCAMPUS OUEST, INRA 35000 Rennes, France  
9 AGROCAMPUS OUEST, UMR1069 SAS, 65 rue de Saint-Brieuc 35042 Rennes Cedex, France  
10

11 Didier Michot, UMR SAS, AGROCAMPUS OUEST, INRA 35000 Rennes, France  
12 AGROCAMPUS OUEST, UMR1069 SAS, 65 rue de Saint-Brieuc 35042 Rennes Cedex, France  
13

14 Nicolas P.A. Saby, INRA, InfoSol Unit  
15 US 1106, 45075 Orléans, France  
16  
17

18 Sébastien Vincent, UMR SAS, INRA, AGROCAMPUS OUEST 35000 Rennes, France  
19 INRA, UMR1069 SAS,65 rue de Saint-Brieuc 35042 Rennes cedex, France  
20

21 Blandine Lemerrier, UMR SAS, AGROCAMPUS OUEST, INRA 35000 Rennes, France  
22 AGROCAMPUS OUEST, UMR1069 SAS, 65 rue de Saint-Brieuc 35042 Rennes Cedex, France  
23

24 **Corresponding Author:** Yosra Ellili

25 **Corresponding Author's Institution:** UMR SAS, INRA, AGROCAMPUS OUEST 35000  
26 Rennes, France

27 **Corresponding Author's contact** (email) ellili.yosra@gmail.com

28  
29

30 **Keywords**

31 Digital soil mapping, spatial disaggregation, soil sampling, validation, France

32

33 **Abstract**

34 Spatial disaggregation of soil map units involves downscaling existing information to produce  
35 new information at a finer scale than that of the original source. Currently, it is becoming a  
36 powerful tool to address the spatial distribution of soil information over large areas, where legacy  
37 soil polygon maps are the only source of soil information. Because of the high expense of  
38 additional resampling, few studies have sought to validate disaggregated soil maps using  
39 independent sampling. This study implemented spatial disaggregation approach to measure the  
40 quality of soil property predictions derived from disaggregated soil maps, using stratified simple  
41 random sampling of a study area of 6 848 km<sup>2</sup> (11 strata and 135 soil profiles). In a previous  
42 study, the existing legacy soil polygon map of Brittany (France) at 1:250,000 scale was spatially  
43 disaggregated at 50 m resolution using an algorithm called Disaggregation and Harmonisation of  
44 Soil Map Units Through Resampled Classification Trees (DSMART), which uses soil-landscape  
45 expert rules of soil distribution in space. By fitting equal-area spline functions, soil properties  
46 were then estimated at six depth intervals according to GlobalSoilMap specifications. To validate  
47 disaggregated soil maps, two approaches were developed according to the soil attribute nature  
48 (continuous or categorical). For categorical soil properties (soil parent material, soil drainage  
49 class, soil type and soil depth class), the overall strict purity (the degree to which all classification  
50 criterion are respected) by the most probable STU (Soil Typological Unit) map was estimated at  
51 34%, while the overall average purity reached 70%. The overall partial soil-type purity reached  
52 60%, the overall partial parent material purity reached 78% and the overall partial soil drainage

53 class as well as soil depth class purities reached 65% and 78%, respectively. Continuous soil  
54 properties (clay content, fine silt content, coarse silt content, total silt content, fine sand content,  
55 coarse sand content, coarse fragments, Cation Exchange Capacity (CEC) and pH) were validated  
56 at two soil depth intervals (5-15 and 30-60 cm) using 260 soil samples. In general, soil property  
57 predictions were unbiased except for coarse fragments and CEC in the 5-15 cm layer. Validation  
58 statistics ( $R^2$ , RMSE, RRMSE and ME) were better for the 30-60 cm layer except for soil  
59 particle-size distribution. Thus, differences in prediction accuracies among strata (the validation  
60 support) denote areas where more soil data or better soil prediction models are needed to improve  
61 the disaggregation process.

62

63

64

65

66

67

68

69

70

71

72

## 73        **1. Introduction**

74        Soil information derived from legacy maps is often at a spatial resolution too coarse to be useful  
75        for soil management decisions and for solving a variety of agricultural and environmental issues  
76        (Bui and Moran, 2001). Spatial disaggregation of legacy soil polygon maps has been introduced  
77        recently to produce enhanced soil information at a finer scale than that of the original source  
78        (Yang et al., 2011; Kerry et al., 2012; Odgers et al., 2014; Subburayalu et al., 2014; Feng et al.,  
79        2016; Møller et al., 2019; Ellili et al., 2019; Zeraatpisheh et al., 2019). Such disaggregation  
80        techniques are based on machine-learning algorithms modelling the relation between a soil class  
81        and a suite of environmental covariates, which are assumed to represent soil forming factors. This  
82        approach has been shown to be a powerful tool to deliver soil information over large areas where  
83        complex soil polygons, which contain several STU (Soil Typological Unit), are the only source  
84        of soil information (Bui and Moran, 2001). Nevertheless, disaggregated digital soil maps, like  
85        available conventional soil maps, are inherently uncertain, and their quality has been  
86        insufficiently investigated.

87        To measure the quality of soil maps, it is recommended to compare predictions to independent  
88        data not used in the modelling (Chatfield, 1995, Brus et al., 2011) (i.e. “external” or “test”  
89        accuracy). To this end, one needs to define the sampling scheme for validation, which comprises  
90        two important aspects: sampling design and sampling support.

91        For the validation strategy, three common approaches are usually found in the literature. The first  
92        consists of applying internal validation by splitting the data. This approach randomly selects  
93        validation subsamples from the calibration dataset to be used to estimate the accuracy of the fitted  
94        model. Generally, these “holdback” data represent a small part (20-30%) of the full dataset. This  
95        validation approach was used, for example, by Ramirez et al. (2014) to assess the quality of

96 digital soil maps across a study area covering a total area of 5 km<sup>2</sup> in Sao Paulo state in Brazil. It  
97 was also used by Veronesi et al. (2012) to validate 3D soil compaction maps of the Czech  
98 University of life Sciences farm (21 ha) located in the Czech Republic. The second approach  
99 consists of using a k-fold cross-validation. It differs from the previous method in that the splitting  
100 procedure is repeated several times, which makes it more efficient (Hastie et al., 2009;  
101 Stoorvogel et al., 2009; Biswas and Zhang, 2018). Leave-one-out cross-validation is the most  
102 common form of cross-validation, ensuring that each sample point can be used as a validation  
103 dataset: one sample point is left, while the rest is used to calibrate the model. This approach was  
104 explored by Malone et al. (2009) to validate digital soil maps depicting soil organic carbon  
105 (SOC) content and available water capacity across a 1500 km<sup>2</sup> area in Sydney, Australia. Lacoste  
106 et al. (2014) also used cross-validation on their training datasets to assess the quality of high-  
107 resolution 3D maps of SOC content over a heterogeneous agriculture landscape in Brittany,  
108 France. In contrast to the two previous methods, the third method involves collecting new  
109 samples using probability sampling to assess model accuracy.

110 The first two approaches are often used because quality measures and their standard errors can be  
111 obtained easily without additional sampling. However, they may provide unbiased and valid  
112 estimates of map accuracy only if the locations of sampling sites are selected by probability  
113 sampling, even if the subset is randomly selected from the dataset (Brus et al., 2011). Indeed,  
114 spatial auto-correlation will generally occur within the prediction errors, and the calibration  
115 dataset itself can be biased because the splitting process is not statistically optimised. Therefore,  
116 Brus et al. (2011) recommended collecting additional independent data according to a probability  
117 sampling strategy. Nevertheless, due to the high expense of additional resampling, few studies  
118 have validated the accuracy of spatial model using independent sampling as done by Walter

119 (1990), Kempen et al. (2009), Brus et al. (2011) and Keskin and Grunwald (2018). In 2009,  
120 Kempen et al. explored the use of multinomial logistic regression for digital soil mapping to  
121 update an existing soil map and used additional fieldwork to assess the quality of the updated soil  
122 map. Brus et al. (2011) validated a soil-class map of the Drenthe province of the Netherlands  
123 (268,000 ha) using 150 validation locations selected by stratified simple random sampling.  
124 Stoorvogel et al. (2009) also used random transect sampling to assess the accuracy of a  
125 quantitative soil map of the Niore study area (81,600 ha), Senegal, depicting SOC content of the  
126 topsoil.

127 Another important aspect for validating a digital soil map is the type of sampling units and  
128 sample support. In most digital soil mapping studies, using a point support is the standard  
129 practice to assess the quality of the soil maps produced. Therefore, even at large areal extents,  
130 digital soil maps are validated over a small area using bulk soil samples and soil cores rather than  
131 management-related spatial supports (e.g. entire fields or management units). In the digital soil  
132 mapping literature, few studies have considered a non-point support in the validation procedure.  
133 For example, Bishop et al. (2015) assessed the quality of digital soil maps depicting clay content  
134 using an independent validation dataset collected following stratified random sampling. They  
135 investigated three spatial supports: i) point, ii) 48 m blocks and iii) soil-land use complexes. They  
136 found that point supports yielded the lowest measures for assessing digital soil map quality, while  
137 soil-land use complexes achieved the highest. In another study, Stoorvogel et al. (2009) validated  
138 a map of topsoil SOC content in Senegal using validation samples collected at a block support of  
139 30 x 30 m, which corresponds to the mean size of fields in their study area. Each composite soil  
140 sample had a support of 900 m<sup>2</sup> and was derived by thoroughly mixing five subsamples taken  
141 within a 12 m radius to capture soil short-range variability. Saby et al. (2008) clearly showed a

142 strong relation between within-site variability and site area for soil monitoring sites in Europe. To  
143 integrate short-range spatial variation, soil inventory programs usually follow a composite  
144 sampling approach. For example, in the framework of the Land Use and Cover Area frame  
145 Statistical survey (LUCAS) a harmonized soil dataset was collected over the extent of the  
146 European Union following a composite sampling approach (Ballabio et al., 2016). At the national  
147 scale, the French soil monitoring network (RMQS) adopted a composite sampling strategy as  
148 well (Saby et al., 2014).

149 The quality of digital soil maps can be assessed by several measures, which depend mainly on  
150 whether the soil information is categorical or continuous. To assess the quality of soil maps in the  
151 Netherlands, Marsman and De Gruijter (1986) calculated multiple quality indicators for  
152 categorical information such as the partial purity (the degree of concordance between observed  
153 and predicted of a classification criterion) of several classification criterion (subgroup, sand  
154 classes, loam classes and groundwater classes), the mean purities of soil classification criterion  
155 and the strict purity. The accuracy of continuous soil property predictions by soil maps is  
156 generally depicted by common statistics of the cumulative spatial error such as the mean error,  
157 the variance error, absolute error and the mean squared error.

158 The objective of this study was to develop a method to measure the accuracy of a 50 m resolution  
159 disaggregated soil maps and their derived soil properties maps over a large area (6,848 km<sup>2</sup>),  
160 using an independent validation dataset whose sampling locations were selected using a  
161 probability sampling strategy. Like Odgers et al. (2015) continuous soil properties were derived  
162 from disaggregated soil maps to address the spatial distribution of continuous soil properties (pH,  
163 SOC content, soil distribution size, Cation Exchange Capacity (CEC)) over regular depths up to  
164 200 cm. Moreover, the accuracy of disaggregated maps depicting predicted STU (Odgers et al.,

165 2015; Chaney et al., 2016, Møller et al., 2019) as well as categorical soil attributes (soil type, soil  
166 drainage class, parent material and soil depth class) were assessed to determine how the accuracy  
167 of soil maps varies with classification criteria selected. In our context, some classification  
168 criterion like soil depth and soil drainage behaviour are relevant to characterize the agronomic  
169 potential of soils. For example, insufficient soil drainage increases the risk of soil compaction and  
170 reduces nutrient availability. In addition, predictions of soil properties are the main inputs of  
171 decision-making tools to sustainably manage and solve environmental issues. Thus, maps of  
172 functional soil properties with known accuracy are needed to provide a simple guide for non-soil  
173 specialist agriculture and stockholders. Overall, two approaches were developed to validate both  
174 categorical maps depicting predicted soil typological units (STUs) with their associated  
175 classification criterion and continuous maps depicting soil properties at two soil depths.

## 176 **2. Materials and methods**

### 177 **2.1 Study area**

178 The study area is the Ille-et-Vilaine department, in eastern Brittany, France (NW France, 47° 40'  
179 to 48° 40' N, 1° to 2° 20' W) (Fig. 1). It has a total area of 6,848 km<sup>2</sup>, which is drained mainly by  
180 its major rivers (Ille and Vilaine rivers ) and their tributaries. The central and coastal parts of the  
181 study area have low elevation, usually less than 50 m above sea level in the coastal zone and the  
182 valleys and less than 100 m elsewhere. The western part of the department has higher elevations,  
183 peaking at 256 m. The mean annual rainfall is about 650 mm and the annual temperatures  
184 average 11.4 °C. The study area is part of the Armorican Massif (BRGM, 2009), whose geology  
185 is complex: intrusive rocks (granite, gneiss and micaschist) in northern and north-western zones,  
186 sedimentary rocks (sandstone, Brioverian schist) in central and southern zones, and superficial  
187 deposits (aeolian loam, alluvial and colluvium deposits) overlaying bedrock formations with

188 decreasing thickness from north to south. This high geological heterogeneity generates high soil  
189 variability over short distances. The soils in the study area include Cambisols, Stagnic Fluvisols,  
190 Histosols, Podzols, Luvisols and Leptosols according to the World Reference Base of Soil  
191 Resources (IUSS Working Group WRB, 2014). The main land uses are annual crops (e.g. maize,  
192 wheat, barley) and temporary or permanent grasslands, but the study area also includes woods  
193 and natural areas. In this study, the built environment was not considered.

## 194 **2.2 Disaggregated soil map**

195 To address soil spatial distribution in Brittany, a regional database at 1:250,000 scale called the  
196 “Référentiel Régional Pédologique” was developed in 2012 in the “Sols de Bretagne” project.  
197 This regional database defines a set of polygons with crisp boundaries commonly called Soil Map  
198 Units (SMUs). SMUs are complex, namely each SMU contains several soil types called Soil Type  
199 Units (STUs) in known proportions. SMUs are defined as areas with homogeneous soil-forming  
200 factors, such as morphology, geology, and climate. Each STU is vertically organized into strata.  
201 The strata are spatial horizons describing the vertical structuration of STUs. Pedological features  
202 of SMU, STU and strata including depth and thickness, soil organic carbon content, CEC, pH,  
203 and 5-particle size fractions are collated in a relational database called Donesol (INRA Infosol,  
204 2014). The Brittany coarse soil map contains 341 soil map units and 320 STUs

205 In an earlier study (Vincent et al., 2018), the existing 1:250,000 legacy soil map of Brittany was  
206 disaggregated at 50 m resolution using the algorithm DSMART and soil-landscape relations. This  
207 study yielded a set of rasters depicting the three most probable STUs and their probability of  
208 occurrence within each pixel of 0.25 ha. These three most probable STUs allowed capturing the  
209 most variability in predictions. The most probable STU in a grid cell is the one that was most  
210 frequently predicted based on 50 iterations (Fig. 2). Similarly, the 2<sup>nd</sup> STU and the 3<sup>rd</sup> STU were

211 respectively, the 2<sup>nd</sup> and the 3<sup>rd</sup> most frequently predicted STU in a grid cell. Therefore, the three  
212 most probable STUs are those predicted with the three highest probabilities of occurrence.  
213 Restricted to the study area, 98 SMUs were disaggregated to spatially delineate 158 STUs  
214 occurring within the SMUs. In our study, independent soil data was used to validate the three  
215 most probable soil maps with their derived soil properties maps, covering the study area.

### 216 **2.3 Soil property estimation**

217 Continuous soil properties were mapped at the regional scale from the 1:250,000 disaggregated  
218 map (Vincent et al., 2018) according to GlobalSoilMap depth intervals (0-5 cm, 5-15 cm, 15-30  
219 cm, 30-60 cm, 60-100 cm and 100-200 cm). The first step of our study standardized the depth of  
220 STU horizons by fitting equal-area spline functions (Bishop et al., 1999) using the R package  
221 GSIF (Hengl., 2006). The equal-area spline function respects mean values of soil properties and  
222 ensures continuous variation in soil properties with depth (Malone et al., 2009). The result is, for  
223 each STU, a set of interpolated values of soil properties for the required depth intervals up to 200  
224 cm that equals the mean of the intervals.

225 Soil properties were estimated as the weighted mean of modal values of the reference soil  
226 property, whose weights were the probabilities of occurrence of the relevant STU (Eq 1). In our  
227 study, we estimated soil properties in space using the property values and probabilities of  
228 occurrence of the three most probable STUs, as follows:

$$229 \quad \hat{y}(x_i) = \frac{\sum_{k=1}^3 (P_{STU_k, x_i}) * y(STU_k, x_i)}{\sum_{k=1}^3 (P_{STU_k, x_i})} \quad [1]$$

230 where  $\hat{y}(x_i)$  is the predicted soil property for grid cell  $(x_i)$ ,  $y(STU_k, x_i)$  is the reference soil  
231 property value associated with STU  $k=1, 2, 3$  predicted at all  $(x_i)$  and  $(P_{STU_k}, x_i)$  is the  
232 probability of occurrence of  $STU_k$  at the given grid cell  $(x_i)$ .

## 233 **2.4 Development of soil dataset for map validation**

234 Sampling locations were selected using a stratified simple random sampling strategy, in which  
235 the mapped area was subdivided into subareas called strata, and from each stratum a simple  
236 random sample was selected. Eleven strata were obtained by grouping the 96 SMUs of the  
237 original soil map by similar dominant soil parent material: granite or gneiss, Aeolian loam, soft  
238 schist, sandstone, gritty schist, alluvial terrace, alluvial deposits, continental alluvium, and  
239 alluvial marsh (Fig. 3). A total of 45 sampling locations were selected with a per-stratum sample  
240 size proportional to the area of each stratum. Each stratum had a minimum of two sampling  
241 locations (Table 1). Locations for which sampling permission was denied or which proved  
242 otherwise impossible to sample were replaced with randomly selected locations within the same  
243 stratum.

244 At each selected site, a transect along the hillslope was defined and divided into three sections:  
245 upslope, midslope and downslope (Fig. 3). A principal sampling location was then randomly  
246 selected within each section and two additional points were selected in a random direction at 20  
247 m distance from the main (principal) location. This approach was followed to capture the local  
248 short-range spatial variability. The set of three locations, characterising each section, defined a  
249 plot in our study (Fig. 3). The validation dataset comprises a total of 135 plots.

250 Validation sites were located using a global positioning system (GPS) with a mean absolute error  
251 of 2 m in each direction. Within each section, the main site was described from auger borings up

252 to 200 cm in depth, and the two other soil profiles were described up to 120 m in depth. Soil  
253 morphology was described at each sampling site to determine each horizon: upper and lower  
254 limits, organic matter content, soil moisture, compacity, matrix and mottle colours according to  
255 Munsell soil colour chart, coarse fragment percentage, and soil texture class according to the  
256 GEPPA texture triangle (Baize, 2000). Finally, soil depth, soil type, soil drainage class and soil  
257 parent material were defined according to Rivière et al. (1992) and the French soil classification  
258 system (Baize and Girard, 2008). The soil type refers to the identification of diagnostic horizons  
259 depicting pedogenetic processes. The soil type can be for example Cambisol, Fuvisol,  
260 Albeluvisol. Meanwhile, the STU nomenclature reflects different information at the same time as  
261 the weathering degree of soil parent material, the redoximorphic conditions, and the soil depth. In  
262 our study, three drainage classes were distinguished: well drained, moderately drained and poorly  
263 drained. Similarly, only two soil depth classes were distinguished: deep ( $\geq 60$  cm) and shallow ( $<$   
264 60 cm). After field description, each of the 405 soil profiles was individually allocated to a  
265 suitable STU.

266 Soil samples for physico-chemical analyses were collected at six depth intervals according to  
267 GlobalSoilMap specifications: 0-5, 5-15, 15-30, 30-60, 60-100, and 100-200 cm. For each plot  
268 and each soil depth interval, a composite soil sample was collected from the three points sampled  
269 within the 20 m radius.

## 270 **2.5 Laboratory analysis**

271 In our study, for budgetary raisons, only two soil layers were considered (5-15 cm and 30-60 cm).  
272 The 5-15 cm soil layer allowed characterizing soil properties of topsoil genetic horizons while the  
273 second soil layer 30-60 cm allowed characterizing soil properties of deeper soil horizons. Overall,  
274 considering the maximum soil depth, a total of 260 soil samples were air dried then sieved to 2

275 mm; the resulting fine earth and gravel were weighed to determine the percentage of coarse  
 276 material present in soils. All samples were analysed according to standard methods to determine  
 277 their particle-size distribution in five classes (NF X 31-107), SOC content by dry combustion  
 278 (Thermoscientific Finnigan EA 1112 Flash elemental analyser) (NF ISO 10694), CEC (NF X 31-  
 279 130) and pH 1:5 H<sub>2</sub>O (soil pH in water with 1:5 soil-to-water ratio, NF ISO 10390).

## 280 **2.6 Measures of map quality**

281 Fig. 4 shows the general diagram followed to compute different maps quality measures according  
 282 to the type of soil property map produced. The following sections detailed each approach as well  
 283 as the quality indicators retained.

### 284 **2.6.1 Quality measures for continuous soil properties**

285 The prediction performance of continuous soil properties was assessed by several quality  
 286 measures relating the observed value of a soil property derived from in situ sampling to its  
 287 corresponding prediction for each depth interval. These statistics comprise the root-mean-squared  
 288 error (RMSE) (Eq 2), relative-root-mean-squared error (RRMSE) (Eq 3), mean error (ME) (Eq  
 289 4), and coefficient of determination (R<sup>2</sup>) (Eq 5).

290 For stratified simple random sampling (De Gruijter et al., 2006), the quality indicator is estimated  
 291 as the weighted mean per stratum:

$$292 \text{ RMSE} = \sum_{h=1}^H w_h \widehat{rmse}_h = \sum_{h=1}^H w_h \sqrt{\frac{1}{n_h} \sum_{i=1}^{n_h} (\hat{y}(x_i) - y(x_i))^2} \quad [2]$$

$$293 \text{ RRMSE} = \sum_{h=1}^H w_h \widehat{rrmse}_h = \sqrt{\frac{1}{n_h} \sum_{i=1}^{n_h} \frac{1}{y(x_i)^2} (\hat{y}(x_i) - y(x_i))^2} \quad [3]$$

$$294 \text{ ME} = \sum_{h=1}^H w_h \widehat{me}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} (\hat{y}(x_i) - y(x_i)) \quad [4]$$

295 For  $R^2$ , there is no estimator for stratified simple random sampling.

$$296 \quad R^2 = \sqrt{\frac{\sum_{i=1}^n (y(x_i) - \widehat{y}(x_i))^2}{\sum_{i=1}^n (y(x_i) - \bar{y}(x_i))^2}} \quad [5]$$

297 where  $W_h$  denotes the relative area of stratum  $h$ ,  $h=1 \dots H=11$ ,  $y(x_i)$  denotes the observed value  
298 of the soil property at validation site  $(x_i)$ ,  $\widehat{y}(x_i)$  denotes the predicted value of the soil property  
299 for the 50 m cell containing  $x_i$  and  $n_h$  is the number of validation soil samples collected at each  
300 depth interval within each stratum.

## 301 **2.6.2 Quality measures for categorical soil properties**

### 302 **2.6.2.1 Validation supports**

303 In our study, quality measures of categorical soil attributes were estimated at three different  
304 scales:

- 305 • Section, to detect whether the quality of soil maps varied by hillslope position
- 306 • Mapping stratum, the basis of the stratified simple random sampling
- 307 • Global considering the entire study area.

308 Thus, we calculated measures both globally over the entire study area and on average within each  
309 stratum. This procedure provided an overall assessment of predictive performance according to  
310 the type of support.

### 311 **2.6.2.2 Quality measures**

312 Categorical soil attributes were validated using three disaggregated soil maps depicting the 1<sup>st</sup>,  
313 2<sup>nd</sup> or 3<sup>rd</sup> most probable STUs, respectively. We used several methods to calculate purity values,  
314 considering four classification criteria: parent material, soil type, soil drainage class (i.e.

315 redoximorphic conditions of soils) and soil depth class. All quality measures were detailed in the  
316 following section:

- 317 • **Partial purity:** the percentage of field observations of a given criterion that equal the  
318 value on the disaggregated map. Partial purity was calculated for each soil profile and  
319 equalled 1 if the predicted classification criterion equalled the observed classification  
320 criterion or 0 if not.
- 321 • **Strict purity:** the percentage of field observations for which all four-classification criteria  
322 equal those of the disaggregated map. This measure was calculated for each soil profile.
- 323 • **Mean purity:** the arithmetic mean of the partial purities of the four criteria.

324 The partial purity of each classification criterion equalled 1 if correctly predicted for at least one  
325 of the three profiles prospected within a 20 m radius. Hence, each section had a single purity  
326 value for each criterion, corresponding to a binary value of 0 or 1.

327 At the transect level, the partial, strict and mean purities were calculated by averaging the purity  
328 values of all sections included.

329 At the stratum level, partial purities were calculated for each of the four criteria by averaging the  
330 corresponding purity values of all  $n$  transects included in a given stratum. As with partial purity,  
331 the strict purity of the 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> STU maps equalled the mean of all strict purities for all  $n$   
332 transects within the stratum. Mean purity was derived from the partial purities and equals the  
333 arithmetic mean of the four partial purities. The standard deviation (Se) of each partial purity  
334 within strata was also computed using Eq 6:

$$335 \text{ Se } (P_{hk}) = \sqrt{\sum_{i=1}^{n_h} \frac{1}{(n_h-1)} (P_{ihk} - \bar{P}_{hk})^2} \quad [6]$$

336 where  $P_{hk}$  denotes the partial purity of stratum  $h$  for a given map  $k$  (1<sup>st</sup>, 2<sup>sd</sup> and 3<sup>rd</sup> predicted  
 337 STU),  $P_{ihk}$  denotes the purity value of all transects within stratum  $h$  of map  $k$ ,  $n_h$  denotes the  
 338 number of sections in stratum  $h$  and  $\bar{P}_{hk}$  denotes the mean purity value in the given stratum.

339 For the stratified simple random sampling, the overall partial purity of each classification  
 340 criterion (parent material, soil type, soil drainage class, and soil depth class), the overall strict  
 341 purity and the overall mean purity were estimated as the mean of the strata purities weighted by  
 342 their respective area using Eq 7 (De Gruijter et al., 2006):

$$343 \quad \bar{P}_k = \sum_{h=1}^H w_h P_{hk} \quad [7]$$

344 where  $\bar{P}_k$  denotes the overall purity,  $w_h = A_h/A$  where  $A$  denotes the study site area and  $A_h$  denotes  
 345 the stratum area,  $P_{hk}$  denotes the mean of purity values (partial, average and strict) in stratum  $h$   
 346 and  $k$  denotes the STU map. The associated standard error (Se) was estimated using Eq 8:

$$347 \quad \text{Se}(\bar{P}_k) = \sqrt{\sum_{h=1}^H \frac{1}{(n_h)} (w_h^2 s_h^2)} \quad [8]$$

348 where  $s_h^2$  denotes the variance of purity values in stratum  $h$  calculated using Eq 6.

### 349 **3. Results**

#### 350 **3.1 Soil spatial variability over short distances**

351 Overall, 92% of the plots showed only one parent material, thus short-range variability appeared  
 352 less pronounced for this criterion; similar percentages were observed for soil drainage class and  
 353 soil depth class (Fig. 5). In contrast, STUs generally appeared highly variable, as only 40% of the  
 354 validation plots had a single STU, while 55% and 5% of the validation plots had two or three

355 STUs, respectively. Overall, short-range spatial variation in STU allocated was higher than for  
356 the other classification criteria.

### 357 **3.2 Purity measures**

358 In general, the 1<sup>st</sup> most probable STU map had the highest overall purities (Table 2). For instance,  
359 overall purity of soil parent material was estimated at 78%, 44% and 31% for the 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup>  
360 STU maps, respectively. Likewise, overall purity of soil depth class was higher for the 1<sup>st</sup> STU  
361 maps (78%) than for the 2<sup>nd</sup> (56%) or 3<sup>rd</sup> STU maps (48%). For overall purity of soil type, the 1<sup>st</sup>  
362 STU map reached 60% and those of the 2<sup>nd</sup> and 3<sup>rd</sup> STU maps reached 31% and 26%,  
363 respectively. For drainage class purity, the 1<sup>st</sup> STU map remained the most accurate (65%),  
364 followed by the 2<sup>nd</sup> (35%) and 3<sup>rd</sup> STU maps (29%). STU purity appeared to be low for the three  
365 STU maps, being 23%, 5% and 3% for the 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> STU maps, respectively.

366 Over the entire study area, strict and mean purities had a positive linear relation with the mean  
367 probability of STU occurrence (Fig. 6). For the three soil maps, strict purity appeared to be  
368 relatively low (Table 2, Fig. 6). The highest strict purity (34%) was reached by the most probable  
369 STU map followed by the 2<sup>nd</sup> STU map (12%). The 3<sup>rd</sup> most probable STU map had the lowest  
370 strict purity (5%). Meanwhile, mean purity was almost three times that of strict purity for all soil  
371 maps except for the 1<sup>st</sup> STU map, for which it was only two times that of strict purity. Mean  
372 purity equalled 70%, 42% and 34% for the 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> STU map, respectively.

373 Aeolian loam, hard schist and alluvial terrace strata had the lowest strict purities (< 20%) (Table  
374 3). In contrast, medium schist, continental alluvium and alluvial marsh strata had the highest  
375 strict purities (> 67%). Mean purity exceeded 65% for all strata except hard schist and alluvial  
376 deposits, for which it was 54% and 58%, respectively. Continental alluvium (92%) and alluvial  
377 marsh strata (96%) had the highest mean purities. For drainage class, the partial purity for granite

378 or gneiss, soft schist, medium schist, sandstone, alluvial terrace, continental alluvium and alluvial  
379 marsh strata were the highest (> 65%). Aeolian loam, gritty schist and alluvial deposit strata had  
380 partial purities  $\geq 50\%$ , while that of hard schist was estimated at 33%.

381 For soil type purity, soil type was generally predicted well, except for alluvial terrace strata,  
382 considering both the 1<sup>st</sup> and 2<sup>nd</sup> probable STU (Fig. 7). Alluvial terrace strata had the lowest  
383 purity (17%), which markedly affected the strict purity (17%), whereas the mean purity remained  
384 high (71%). For alluvial marsh and continental alluvium strata, the distribution functions were on  
385 the maximum, indicating that all soils developed from these parent materials were correctly  
386 predicted by the 1<sup>st</sup> most probable STU.

387 When comparing the distribution functions of soil type purity within strata, medium schist and  
388 gritty schist strata had similar distributions for the 1<sup>st</sup> STU. Meanwhile, medium schist and  
389 alluvial terrace strata had similar distributions for the 2<sup>sd</sup> STU. Furthermore, hard schist,  
390 sandstone and alluvial deposit strata had a wide range of partial purities. This variation was more  
391 pronounced for granite or gneiss, hard schist and gritty schist strata for the 2<sup>sd</sup> STU (Fig. 7).

392 At the local scale, the partial purity of each classification criterion did not vary significantly  
393 among hillslope positions (upslope, midslope and downslope) for any of the three disaggregated  
394 soil maps (Table 4). The  $\chi^2$  test comparing the proportions of good prediction of each  
395 classification criterion was not significant ( $\alpha=0.05$ ). This suggested that the effect of hillslope  
396 position in the disaggregation process is not important for all classification criteria (Table 4). In  
397 addition, confidence intervals of good soil type predictions overlapped for all sections and soil  
398 maps considered (Fig. 8); therefore, there was no significant effect of hillslope position ( $\alpha=0.05$ ).  
399 Overall, the disaggregation algorithm predicted soil spatial distribution along hillslope positions

400 with the same performance, and the 1<sup>st</sup> most probable STU has the best validation measures over  
401 the study area.

### 402 **3.3 Descriptive statistics of continuous soil properties for the validation dataset**

403 In general, the validation dataset covered a wide range of soil property values. For instance, pH  
404 ranged from 3.95 to 8.58 (mean = 6.07) at 5-15 cm and from 4.41 to 8.9 (mean = 6.25) at 30-60  
405 cm (Table 5). The variability expressed by the standard deviation was relatively constant for all  
406 texture fractions and for both depth intervals. SOC content varied more at 5-15 cm, particularly  
407 for alluvial deposits, medium schist, and granite or gneiss (Fig. 9). SOC content decreased  
408 sharply and varied less with increasing depth, except for alluvial marsh. The mean of SOC  
409 contents, except those of soil sampled from alluvial marsh, did not differ significantly at 5-15 cm  
410 even though they clearly differed among parent materials.

411 Considering both soil depth intervals, hard schist parent material had the lowest pH, whereas  
412 marsh parent material had the highest pH. In addition, for medium schist parent material, the  
413 distribution of associated pH values was wider at 5-15 cm than at 30-60 cm.

414 Clay content had a wider distribution for marsh parent material at 5-15 cm, and for medium schist  
415 and alluvial deposit parent materials at 30-60 cm. Overall, clay content increased slightly with  
416 increasing depth for all parent materials except aeolian loam, hard schist and alluvial marsh.

### 417 **3.4 Validation of continuous soil property predictions**

418 At 5-15 cm, MEs of the models were close to zero, suggesting unbiased predictions for pH  
419 (0.29), sand content (-0.20%) and fine silt content (-0.20%) (Table 6). The MEs of soil texture  
420 fractions increased slightly with increasing depth, but those of CEC and SOC content strongly  
421 decreased.

422  $R^2$  was larger for pH and clay content at both 5-15 and 30-60 cm. For instance,  $R^2$  for clay  
423 content was moderate (0.65 and 0.37 at 5-15 and 30-60 cm, respectively). SOC content and  
424 coarse fragment percentage had the smallest  $R^2$ . For SOC content, the  $R^2$  was estimated at 0.07 at  
425 5-15 cm and 0.05 at 30-60 cm soil layer. Meanwhile, the  $R^2$  for coarse fragment percentage was  
426 estimated at 0.13 and 0.05 at 5-15 cm and 30-60 cm soil depths, respectively.

427 RRMSE was less than 1 for all soil properties except coarse fragments at 30-60 cm. RRMSE and  
428  $R^2$  were generally of opposite magnitude. For instance,  $R^2$  was 0.43 for pH at 5-15 cm, with an  
429 RRMSE of about 0.13. An opposite trend characterised SOC content, which had the lowest  $R^2$   
430 (0.07) and an RRMSE of 0.65.

431 Overall, sand, fine silt and total silt contents were underestimated regardless of depth interval,  
432 whereas coarse fragments were consistently overestimated at both depth intervals. Therefore, the  
433 quality of soil property predictions depended on the property considered and was generally better  
434 at 30-60 cm than at 5-15 cm except for particle-size distribution.

## 435 **4 Discussion**

### 436 **4.1 Soil map quality measures**

437 In our study, the quality of disaggregated soil property maps was tested using the common  
438 accuracy measures of partial, strict and mean purity (Wilding et al., 1965; Beckett and Webster,  
439 1971; Walter, 1990; Kempen et al., 2009). Partial purity reveals the percentage of the  
440 disaggregated map for which each classification criterion equals the field observation. This  
441 measure is a strict assessment in which each error in the validation set is given equal weight.  
442 Moreover, it does not consider existing pedological similarities between categories of each  
443 classification criterion. For example, the partial purity of soil type is the same regardless of

444 whether there is confusion in the prediction of two taxonomically distant or taxonomically similar  
445 soils. Even the limits of these quality measures, they remained relevant to assess the accuracy of  
446 disaggregated soil maps, which are mainly used in soil management decisions as well as in the  
447 decision-making tools.

448 The main shortcoming of strict purity is that incorrect prediction of a single classification  
449 criterion has the same influence as that of all classification criteria. The lowest strict purity  
450 reported in our study resulted from incorrect prediction of one or more classification criteria of  
451 soil profiles. For instance, strict purity for the alluvial terrace stratum was 17%, while its  
452 associated partial purities exceeded 83%, except for soil type purity (17%). The hard schist strata  
453 followed the same trend: its strict purity was 17%, while its partial purities were  $\geq 50\%$  except  
454 for drainage class purity (33%). Therefore, low strict purity was driven by the low partial purity  
455 of a single classification criterion although the partial purities for the three other criterion were  
456 high.

#### 457 **4.2 Performance of categorical soil attributes prediction by spatial disaggregation**

458 As collecting new soil samples to validate disaggregated soil maps is expensive and time  
459 consuming, particularly for large areas, few studies have validated the digital soil maps produced  
460 (Minasny et al., 2013). Most of the studies that did were based on relatively small datasets (e.g.  
461  $n=150$  (Kempen et al., 2009);  $n=123$  (Collard et al., 2014),  $n=53$  new + 10 legacy soil profiles  
462 (Nauman et al., 2014)), which may have limited their ability to cover all soil distributions. An  
463 alternative approach is to use legacy soil profiles, as done by Odgers et al. (2014) and Nauman  
464 and Thompson (2014). Although commonly used in digital soil mapping, this approach may have  
465 significant hidden sampling bias because rare soils are poorly represented, and legacy datasets  
466 were usually not designed probabilistically.

467 In our study, the overall purity of soil type was estimated at 60% by the 1<sup>st</sup> most probable STU.  
468 Similar findings were reached by Kempen et al. (2009, 58%) and Collard et al. (2014, 65.9%)  
469 using multinomial logistic regression and probabilistic sampling designs. The overall strict purity,  
470 assessed using independent soil data, was estimated at 34% (1<sup>st</sup> STU) and reached 51%  
471 considering the three most probable STUs. Odgers et al. (2014) reported similar results using 285  
472 validation legacy profiles and the DSMART algorithm; 22.5% of profiles were predicted  
473 correctly with the most probable STU and 50% were predicted as being one of the three most  
474 probable STUs. Other studies, such as Holmes et al. (2015), reported low strict purity (20%, 1<sup>st</sup>  
475 STU), while Subburayalu and Salter (2013) reported strict purity that reached 23% using decision  
476 trees and 26% using random forest model. However, comparing our partial purity results to those  
477 reported in previous studies remains difficult because of the complexity of soils in our study area  
478 and differences in the taxonomic classification system considered.

479 The accuracy of traditional soil maps is also assessed using statistics such as purity and the kappa  
480 index (Kempen et al., 2009). Previously, using stratified random sampling, Marsman and De  
481 Gruijter (1986) reported strict purity values of 8.0-10.7% depending on the mapping method.  
482 Much higher strict purities (39-95% depending on the level of classification) were reported by  
483 Wilding et al. (1965). However, the comparison between digital soil maps and conventional maps  
484 remains unsatisfactory because the taxonomic classification systems differ, and the validation  
485 sampling design was not always defined in each study.

486 The accuracy of the disaggregated soil map appeared to be lower than that of local existing  
487 1:25,000 soil maps (Walter, 1990). These maps depict the southern part of the Ille-et-Vilaine  
488 department and cover almost 2% of our study area. Accurate maps were obtained from field  
489 description of soil profiles selected following a dedicated sampling design and produced using

490 the local Armorican Massif auger method (Rivière et al., 1992). To assess the quality of existing  
491 1:25,000 soil maps, Walter (1990) validated the three most abundant SMUs by randomly  
492 selecting 120 validation sites per SMU. Soils were described using the Armorican Massif auger  
493 method, and the quality of soil maps was assessed according to the partial purity of each  
494 classification criterion, and the strict and mean purities. The study of Walter (1990) and ours  
495 differed greatly in the partial purity of soil type. The partial purities of the existing map exceeded  
496 84% for the three SMUs vs. 60% in our study. Despite differences in soil drainage and soil depth  
497 class categories, strict purities and mean purities were of the same order of magnitude as those in  
498 our study. The mean purities for the three SMUs ranged from 61-77% vs. 72% in our study,  
499 while the strict purities for three SMUs ranged from 12-37% vs. 32% in our study. Therefore, the  
500 disaggregated map cannot replace high-resolution soil maps based on intensive sampling but is an  
501 efficient way to address the spatial variability of dominant soil types across the study area. An  
502 important advantage of disaggregated soil maps is their ability to cover large areas even where  
503 soil data are scarce but environmental covariates are available at high spatial resolution. Thus, the  
504 disaggregated map is an effective way to produce maps of soil properties for large areas, with a  
505 measure of prediction uncertainty from the probability layers.

#### 506 **4.3 Accuracy of quantitative soil property predictions by spatial disaggregation**

507 For soil maps depicting quantitative soil properties, the quality of prediction was assessed using  
508 the common statistical parameters  $R^2$ , RMSE, ME and RRMSE. In general, the variability of  
509 predicted soil properties within the study area seemed plausible when compared to those of  
510 previous studies, such as that of Lacoste et al. (2014). For SOC content, like Lacoste et al. (2014),  
511 results were better (lower absolute ME and RMSE) at 30-60 cm than at 5-15 cm. Using an  
512 independent validation dataset, these authors validated high-resolution 3D SOC distribution maps

513 across a heterogeneous agricultural landscape in Brittany. Overall, they found an increase in SOC  
514 prediction accuracy with soil increasing depth. For example, the RMSE was  $12.6 \text{ g kg}^{-1}$  at 0-7.5  
515 cm but markedly lower ( $2.2 \text{ g kg}^{-1}$ ) at 60-75 cm. However, our results do not agree with those of  
516 Minasny et al. (2013) or Malone et al. (2009), which highlighted a decreasing accuracy of  
517 prediction with increasing depth. The influence of agriculture practices (e.g. tillage, fertilisation),  
518 differences in soils and the quality of legacy soil data might explain this difference. Even though  
519 human influences were proxied using land cover in the DSMART algorithm, this covariate  
520 appears to be too poor for large land use classes to represent human influences in our study.  
521 Spline functions fitted to STUs to estimate soil properties at GlobalSoilMap depth intervals might  
522 also explain some of the inaccuracy of predicted soil properties.

523 Odgers et al. (2015a) performed a similar study in which they applied the Digital Soil Property  
524 Mapping Using Soil Class Probability Rasters algorithm (PROPR) to address the spatial pattern  
525 of clay content over their study area (South Australia) using soil class probability rasters.  
526 Interestingly, the  $R^2$  of Odgers et al. (2015a) is 67% smaller (0.22) than that in our study (0.65) at  
527 5-15 cm and 52% smaller (0.18) than that in our study (0.37) at 30-60 cm. The RMSE and ME  
528 are higher in our study than those of Odgers et al. (2015a) by a factor of  $\sim 2$ . In another study,  
529 Odgers et al. (2015b) mapped pH following the same method as the clay content over the Shire of  
530 Dalrymple in Australia. Their study yielded a small  $R^2$  between observed and predicted pH at  
531 both 5-15 and 30-60 cm.  $R^2$  reached 0.06 vs. 0.43 in our study at 5-15 cm and 0.10 vs. 0.54 in our  
532 study at 30-60 cm. However, the RMSE was relatively similar in the two studies even if the  
533 sampling design and strategy were different.

534 The main shortcoming revealed by this research is that soil classes with low probabilities of  
535 occurrence have no influence on predictions of soil properties. Only the three most probable

536 STUs were considered for validation and to derive quantitative soil property maps. Although  
537 DSMART attempts to predict the most probable STUs for a specific soil-landscape context,  
538 prediction accuracy varies greatly and depends strongly on the soil heterogeneity of the study  
539 area. In addition, the resulting rasters depicting the probability of occurrence of STU maps  
540 themselves depend greatly on polygon component proportions, which are derived from expert  
541 soil knowledge and brings more uncertainty to the predicted STUs. Thus, close examination of  
542 the soil regional database remains crucial to increase prediction performance for both  
543 disaggregated soil maps and derived soil maps depicting quantitative soil properties.

#### 544 **4.4 Improvement and future study**

545 In this study, statistical validation of disaggregated soil maps provided good overall purities.  
546 Each classification criterion was judged individually, and an overall measure (strict purity)  
547 assessed the overall accuracy of digital soil maps. However, expert soil knowledge was not  
548 included when calculating the validation statistics. Soil taxonomic distance, especially for soils  
549 occurring in similar environmental conditions or having similar morphological criteria, was not  
550 considered but is worth investigating in future mapping studies.

#### 551 **5. Conclusion**

552 To validate a soil map, using an independent dataset is a safe option that leads to unbiased  
553 estimates of quality measures without using models to assess the uncertainty (Brus et al., 2011).  
554 This was applied at the regional scale in north-western France, where the three most probable  
555 disaggregated STU maps at 50 m resolution were available. A total of 260 soil samples, collected  
556 at 5-15 and 30-60 cm depths, were analysed to determine their particle-size distributions, CEC,  
557 pH, SOC content and coarse fragment percentages. For soil maps depicting quantitative soil

558 properties derived from spatial disaggregation, soil texture fractions remained the least biased at  
559 both depth intervals. In general, validation statistics of soil properties depended on the soil  
560 property considered as well as on the soil depth interval. For categorical soil maps, the partial  
561 purity of soil class, and the strict and mean purities were informative. Overall, the predictive  
562 quality of disaggregated soil maps (1<sup>st</sup>, 2<sup>sd</sup> and 3<sup>rd</sup> STUs) varied among strata, but continental  
563 alluvium and marsh strata were predicted well for the three soil maps. Furthermore, the 1<sup>st</sup> most  
564 probable STU, with high probability of occurrence, had the best validation measures regardless of  
565 soil parent material. Differences in prediction accuracies among strata denote areas where more  
566 soil data or better soil prediction models have to be obtained or applied first to improve the  
567 disaggregation process.

#### 568 **Acknowledgements**

569 The authors gratefully acknowledge all farmers at the Ille-et-Vilaine department involved in our  
570 research. We thank technical staff who actively participated in field sampling and laboratory  
571 analysis. This research was performed in the framework of the INRA “Ecoserv” metaprogram.  
572 This work was also funded by the Soilserv program funded by ANR (Agence Nationale de la  
573 Recherche) (ANR-16- CE32-0005-01).

574

575

576

577

578

579

580

581

582

583

584

585

586

587

588 **References:**

589 Baize, D., 2000. Guide des analyses en pédologie, INRA. Seconde édition revue et augmentée. Pp  
590 257.

591 Baize, D., Girard, M.C., 2008. Référentiel pédologique 2008. Association française pour l'étude du  
592 sol.

593 Beckett, P.H.T., Webster, R., 1971. Soil variability: A review. *Soils and Fertiliz.* 34, 1: 1-15.

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

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

598 Biswas, A., Zhang, Y., 2018. Sampling Designs for Validating Digital Soil Maps: A Review.  
600 *Pedosphere* 28, 1–15. [https://doi.org/10.1016/S1002-0160\(18\)60001-3](https://doi.org/10.1016/S1002-0160(18)60001-3)

601 Ballabio, C., Panagos, P., Monatanarella, L., 2016. Mapping topsoil physical properties at European  
602 scale using the LUCAS database. *Geoderma* 261, 110–123.  
603 <https://doi.org/10.1016/j.geoderma.2015.07.006>

604 BRGM. 2009. <http://sigesbre.brgm.fr/Histoire-geologique-de-la-Bretagne-59.html>

605 Brus, D.J., Kempen, B., Heuvelink, G.B.M., 2011. Sampling for validation of digital soil maps.  
606 *European Journal of Soil Science* 62, 394–407. <https://doi.org/10.1111/j.1365-2389.2011.01364.x>

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

610 Chatfield, C., 1995. *Problem Solving. A statistician's guide*, second edition. New York: Chapman  
611 and Hall/CRC. <https://doi.org/10.1201/b15238>

- 6122 Collard, F., Kempen, B., Heuvelink, G.B.M., Saby, N.P.A., Richer de Forges, A.C., Lehmann, S.,  
613 Nehlig, P., Arrouays, D., 2014. Refining a reconnaissance soil map by calibrating regression  
614 models with data from the same map (Normandy, France). *Geoderma Regional* 1, 21–30.  
615 <https://doi.org/10.1016/j.geodrs.2014.07.001>
- 6163 De Gruijter, J.J., Brus, D.J., Bierkens, M.F.P., Knotters, M., 2006. *Sampling for Natural Resource*  
617 *Monitoring*. Springer, Berlin.
- 6184 De Gruijter, J.J., McBratney, A.B., Minasny, B., Wheeler, I., Malone, B.P., Stockmann, U., 2016.  
619 Farm-scale soil carbon auditing. *Geoderma* 265, 120–130.  
620 <https://doi.org/10.1016/j.geoderma.2015.11.010>
- 621 Ellili, Y., Malone, B. P., Michot, D., Minasny, B., Vincent, S., Walter, C., and Lemercier, B., 2019.  
622 Comparing three approaches of spatial disaggregation of legacy soil maps based on DSMART  
623 algorithm, *SOIL Discuss.*, <https://doi.org/10.5194/soil-2019-36>, in review.
- 624 Feng, L., Xiaoyuan, G., A-Xing, Z., Fraser, W., Xiaodong1, S., Ganlin, Z., 2016. Soil polygon  
625 disaggregation through similarity-based prediction with legacy pedons. *J Arid Land* 8(5): 760–  
626 772.
- 6275 Holmes, K.W., Griffin, E.A., Odgers, N.P., 2015. Large-area spatial disaggregation of a mosaic of  
628 conventional soil maps: evaluation over Western Australia. *Soil Research* 53, 865.  
629 <https://doi.org/10.1071/SR14270>
- 6306 INRA Infosol, 2014. Donesol Version 3.4.3. Dictionnaire de données.
- 6317 IUSS Working Group WRB, 2007. *World Reference Base for Soil Resources 2006, first update*  
632 2007. *World Soil Resources Reports No. 103*. FAO, Rome. 116 pp.
- 6338 Kempen, B., Brus, D.J., Heuvelink, G.B.M., Stoorvogel, J.J., 2009. Updating the 1:50,000 Dutch  
634 soil map using legacy soil data: A multinomial logistic regression approach. *Geoderma* 151, 311–  
635 326. <https://doi.org/10.1016/j.geoderma.2009.04.023>
- 636 Kerry, R., Goovaerts, P., Rawlins, B.G., Marchant, Ben P., 2012. Disaggregation of legacy soil data  
637 using area to point kriging for mapping soil organic carbon at the regional scale *Geoderma* 170,  
638 347–358.
- 6399 Keskin, H., Grunwald, S., 2018. Regression kriging as a workhorse in the digital soil mapper's  
640 toolbox. *Geoderma* 326, 22–41. <https://doi.org/10.1016/j.geoderma.2018.04.004>
- 6420 Lacoste, M., Minasny, B., McBratney, A., Michot, D., Viaud, V., Walter, C., 2014. High resolution  
642 3D mapping of soil organic carbon in a heterogeneous agricultural landscape. *Geoderma* 213,  
643 296–311. <https://doi.org/10.1016/j.geoderma.2013.07.002>
- 6441 Malone, B.P., McBratney, A.B., Minasny, B., Laslett, G.M., 2009. Mapping continuous depth  
645 functions of soil carbon storage and available water capacity. *Geoderma* 154, 138–152.  
646 <https://doi.org/10.1016/j.geoderma.2009.10.007>

- 6422 Marsman, B.A., De Gruijter, J.J., 1986. Quality of soil maps: a comparison of soil survey methods  
648 in a sandy area, *Soil survey papers*. Soil Survey Inst, Wageningen.
- 649 Minasny, B., McBratney, A.B., Malone, B.P., Wheeler, I., 2013. Digital Mapping of Soil Carbon, in:  
650 *Advances in Agronomy*. Elsevier, pp. 1–47.
- 651 Møller, A.B., Malone, B., Odgers, N.P., Beucher, A., Iversen, B.V., Greve, M.H., Minasny, B.,  
652 2019. Improved disaggregation of conventional soil maps. *Geoderma* 341, 148–160.  
653 <https://doi.org/10.1016/j.geoderma.2019.01.038>.  
654
- 655 Nauman, T.W., Thompson, J.A., 2014. Semi-automated disaggregation of conventional soil maps  
656 using knowledge driven data mining and classification trees. *Geoderma* 213, 385–399.  
657 <https://doi.org/10.1016/j.geoderma.2013.08.024>
- 658 Nauman, T.W., Thompson, J.A., Rasmussen, C., 2014. Semi-Automated Disaggregation of a  
659 Conventional Soil Map Using Knowledge Driven Data Mining and Random Forests in the  
660 Sonoran Desert, USA. *Photogrammetric Engineering & Remote Sensing* 80, 353–366.  
661 <https://doi.org/10.14358/PERS.80.4.353>
- 662 Odgers, N., McBratney, A., Minasny, B., Sun, W., Clifford, D., 2014. Dsmart: An algorithm to  
663 spatially disaggregate soil map units, in: Arrouays, D., McKenzie, N., Hempel, J., de Forges, A.,  
664 McBratney, Alex (Eds.), *GlobalSoilMap*. CRC Press, pp. 261–266.  
665 <https://doi.org/10.1201/b16500-49>
- 666 Odgers, N.P., Holmes, K.W., Griffin, T., Liddicoat, C., 2015a. Derivation of soil-attribute  
667 estimations from legacy soil maps. *Soil Research* 53, 881. <https://doi.org/10.1071/SR14274>
- 668 Odgers, N.P., McBratney, A.B., Minasny, B., 2015b. Digital soil property mapping and uncertainty  
669 estimation using soil class probability rasters. *Geoderma* 237–238, 190–198.  
670 <https://doi.org/10.1016/j.geoderma.2014.09.009>
- 671 Ramirez-Lopez, L., Schmidt, K., Behrens, T., van Wesemael, B., Demattê, J.A.M., Scholten, T.,  
672 2014. Sampling optimal calibration sets in soil infrared spectroscopy. *Geoderma* 226–227, 140–  
673 150. <https://doi.org/10.1016/j.geoderma.2014.02.002>
- 674 Rivière, J.M., Tico, S., Dupont, C., 1992. Méthode tarière Massif Armoricaïn. Caractérisation des  
675 sols. INRA, Rennes, 20 p.
- 676 Saby, N.P.A., Bellamy, P.H., Morvan, X., Arrouays, D., Jones, R.J.A., Verheijen, F.G.A.,  
677 Kibblewhite, M.G., Verdoodt, A., Üveges, J.B., Freudenschuß, A., Simota, C., 2008. Will  
678 European soil-monitoring networks be able to detect changes in topsoil organic carbon content?  
679 *Global Change Biology* 14, 2432–2442. <https://doi.org/10.1111/j.1365-2486.2008.01658.x>
- 680 Saby, N.P.A., Bellamy, Arrouays, D., Jolivet, C., Martin, M.P., Lacoste, M., Ciampalini, R., Richer  
681 de Forges, A.C., Laroche, B., Bardy, M., 2014. National soil information and potential for  
682 delivering GlobalSoilMap products in France: A review. *Globalsoilmap: Basis of the Global*  
683 *Spatial Soil Information System*. Pp.69-72.

6843 Stoorvogel, J.J., Bakkenes, M., Temme, A.J.A.M., Batjes, N.H., ten Brink, B.J.E., 2017. S-World:  
685 A Global Soil Map for Environmental Modelling: S-World: A Global Soil Map for  
686 Environmental Modelling. *Land Degradation & Development* 28, 22–33.  
687 <https://doi.org/10.1002/ldr.2656>

6884 Stoorvogel, J.J., Kempen, B., Heuvelink, G.B.M., de Bruin, S., 2009. Implementation and  
689 evaluation of existing knowledge for digital soil mapping in Senegal. *Geoderma* 149, 161–170.  
690 <https://doi.org/10.1016/j.geoderma.2008.11.039>

6915 Subburayalu, S.K., Slater, B.K., 2013. Soil series mapping by knowledge discovery from an Ohio  
692 county soil map. *Soil Science Society of America Journal* 77, 1254–1268.

693 Subburayalu, S.K., Jenhani, I., Slater, B.K., 2014. Disaggregation of component soil series on an  
694 Ohio County soil survey map using possibilistic decision trees. *Geoderma* 213, 334–345

6956 Hastie, T., Tibshirani, R., Friedman, J., 2009. *The Elements of Statistical Learning: Data Mining,  
696 Inference, and Prediction*. Second Edition. Springer Series in Statistics.

6977 Veronesi, F., Corstanje, R., Mayr, T., 2012. Mapping soil compaction in 3D with depth functions.  
698 *Soil and Tillage Research* 124, 111–118. <https://doi.org/10.1016/j.still.2012.05.009>

6998 Vincent, S., Lemercier, B., Berthier, L., Walter, C., 2018. Spatial disaggregation of complex Soil  
700 Map Units at the regional scale based on soil-landscape relationships. *Geoderma* 311, 130–142.  
701 <https://doi.org/10.1016/j.geoderma.2016.06.006>

7029 Wilding, L.P., Jones, R.B., Schafer, G.M., 1965. Variation of Soil Morphological Properties within  
703 Miami, Celina, and Crosby Mapping Units in West-Central Ohio<sup>1</sup>. *Soil Science Society of  
704 America Journal* 29, 711. <https://doi.org/10.2136/sssaj1965.03615995002900060033x>

7050 Walter, C., 1991. Estimation de propriétés du sol et quantification de leur variabilité à moyenne  
706 échelle: cartographie pédologique et géostatistique dans le sud de l’Ile et Vilaine (France). PhD  
707 Thesis. Ecole nationale supérieure agronomique de Rennes.

708 Yang, L., Jiao, Y., Fahmy, S., Zhu, A., Hann, S., Burt, J.E., Qi, F., 2011. Updating conventional soil  
709 maps through digital soil mapping. *Soil Sci. Soc. Am. J.* 75 (3),  
710 1044–1053.

711 Zeraatpisheh, Z., Ayoubi, S., Brungard, C.W., Finke, P., 2019. Disaggregating and updating a legacy  
712 soil map using DSMART, fuzzy cmeans and k-means clustering algorithms in Central Iran.  
713 *Geoderma* 340, 249–258.

714

715

716

717

718

719

720

721

722

723

724

725 **Figure captions**

726 Fig. 1. Locations of the 45 transects selected by stratified simple random sampling based on  
727 dominant soil parent material strata.

728 Fig. 2. The 50 m resolution disaggregated soil map showing the most probable Soil Typological  
729 Unit and associated probability of occurrence.

730

731 Fig. 3. Diagram of the transect sampling strategy along the hillslope.

732 Fig. 4. Diagram of the validation strategy to compute quality measures of digital soil maps and  
733 their derived soil properties maps

734 Fig. 5. Percentage of validation plots in which 1, 2 or 3 categories of a given soil classification.  
735 criterion (soil depth, soil drainage class, parent material, soil type, soil typological unit) were  
736 observed within a 20 m radius.

737 Fig. 6. Relation over the entire study area between strict and mean purities (%) and probability of  
738 occurrence of the three most probable soil typological units (STUs).

739 Fig. 7. Distribution functions of soil type purity and their mean values (black dots) by stratum for  
740 the (left) 1<sup>st</sup> and (right) 2<sup>sd</sup> soil typological unit (STU) predicted.

741 Fig. 8. Mean percentage of good soil type prediction and their 95% confidence intervals for the  
742 three most probable soil typological units (STUs) according to hillslope position. Dashed lines  
743 denote the overall soil type purity for the entire study area.

744 Fig. 9. Distribution of SOC content, pH and clay content according to observed parent material  
745 from the validation dataset. G: granite, I: gneiss, L: aeolian loam, N: soft schist, O: medium  
746 schist, P: hard schist, Q: sandstone, R: gritty schist, T: alluvial terrace, U: colluvium deposits, V:  
747 alluvial deposits, Vm: marsh. Whiskers represent 1.5 times the interquartile range. The number in  
748 brackets is the number of soil samples collected from each parent material. Means of groups  
749 sharing a letter in the group label do not differ significantly at  $\alpha=5\%$ .

750

751

752

753

754

755

756

757

758

759

760

761

762

763

764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785

**Table headings**

Table 1. Characteristics of the 11 strata used for stratified simple random sampling to produce the validation dataset, dominant parent material, percentage of the study area and number of transects sampled.

Table 2. Estimated partial, strict and mean purities (%) for the three soil maps over the entire study area.

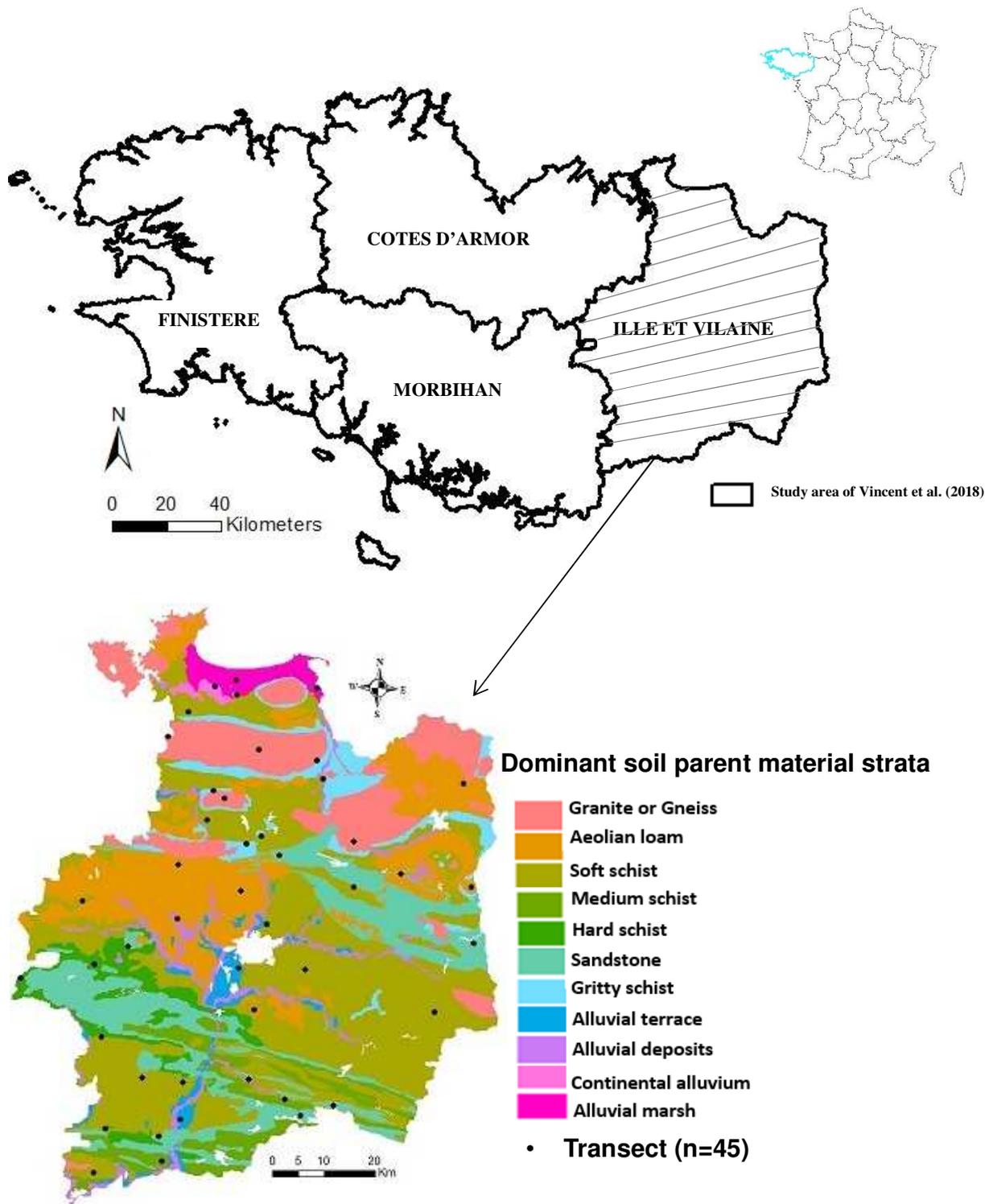
Table 3. Design-based estimates of purities of soil classification criteria, strict purity and mean purity for selected strata for the 1<sup>st</sup> most probable STU map.

Table 4. Results of the  $\chi^2$  test performed to compare proportions of good prediction of classification criteria along hillslope positions for the three soil typological unit (STU) maps ( $\alpha=5\%$ ).

Table 5. Descriptive statistics of soil properties for the validation dataset at 5-15 and 30-60 cm (n = 135 and 125, respectively).

786 Table 6. Accuracy indicators of soil property prediction at 5-15 and 30-60 cm (n = 135 and 125,  
787 respectively, root-mean-squared error (RMSE), relative-root-mean-squared error (RRMSE),  
788 mean error (ME) and coefficient of determination ( $R^2$ )).

789



790

791 Fig.1. Locations of the 45 transects selected by stratified simple random sampling based on

792 dominant soil parent material strata.

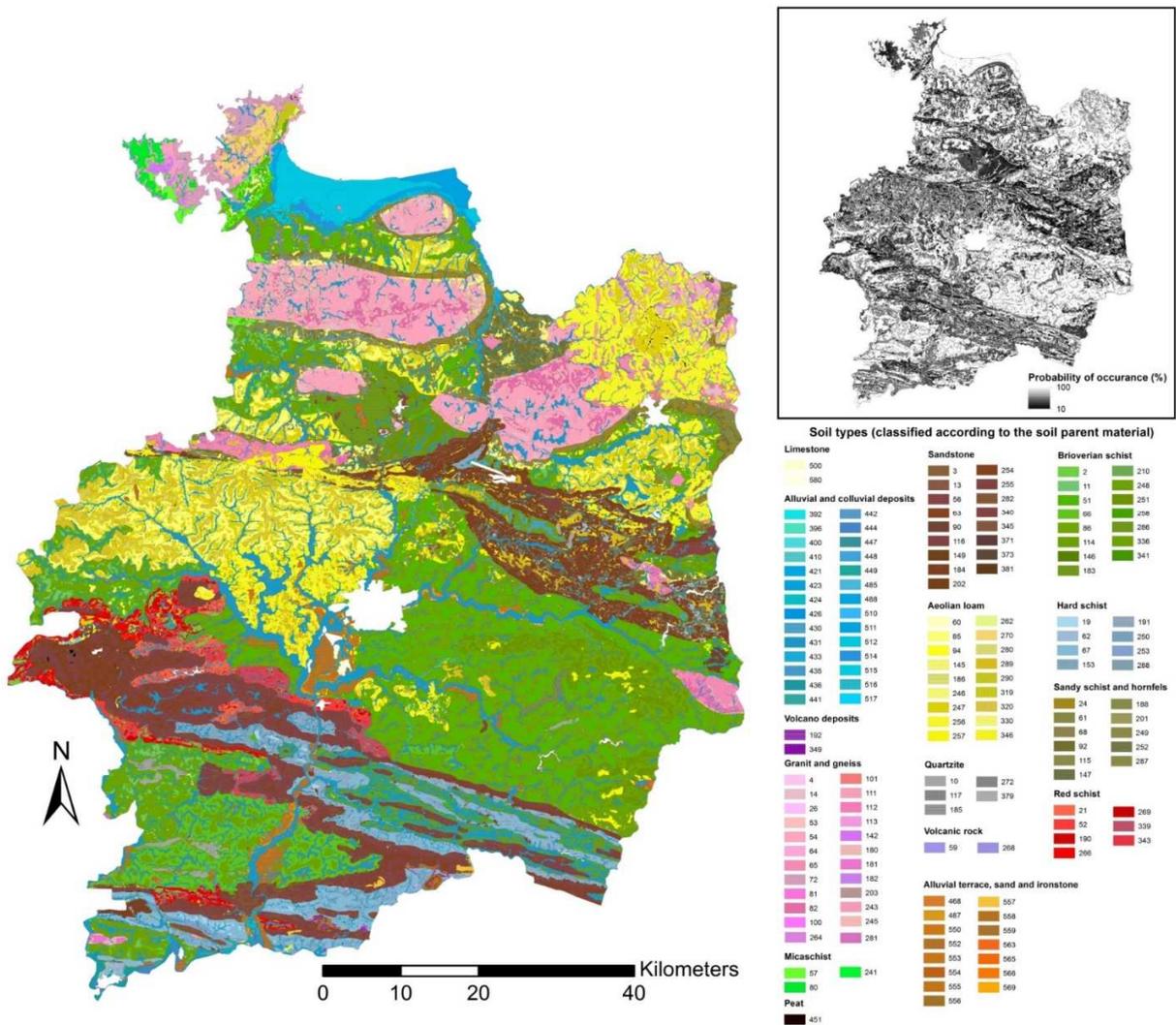


Fig.2. The 50 m resolution disaggregated soil map showing the most probable soil typological unit and associated probability of occurrence.

793

794

795

796

797

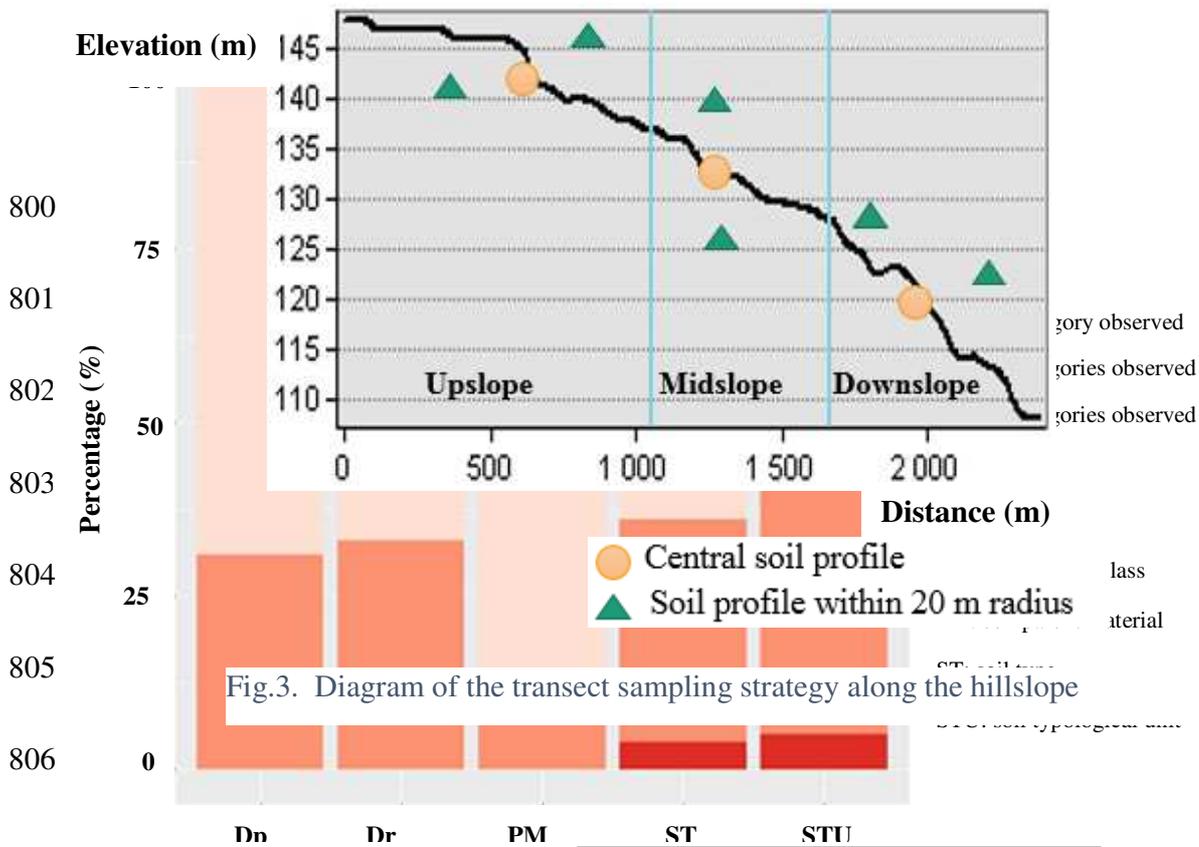


Fig.3. Diagram of the transect sampling strategy along the hillslope

Fig  
(sc  
rac

ion  
) m

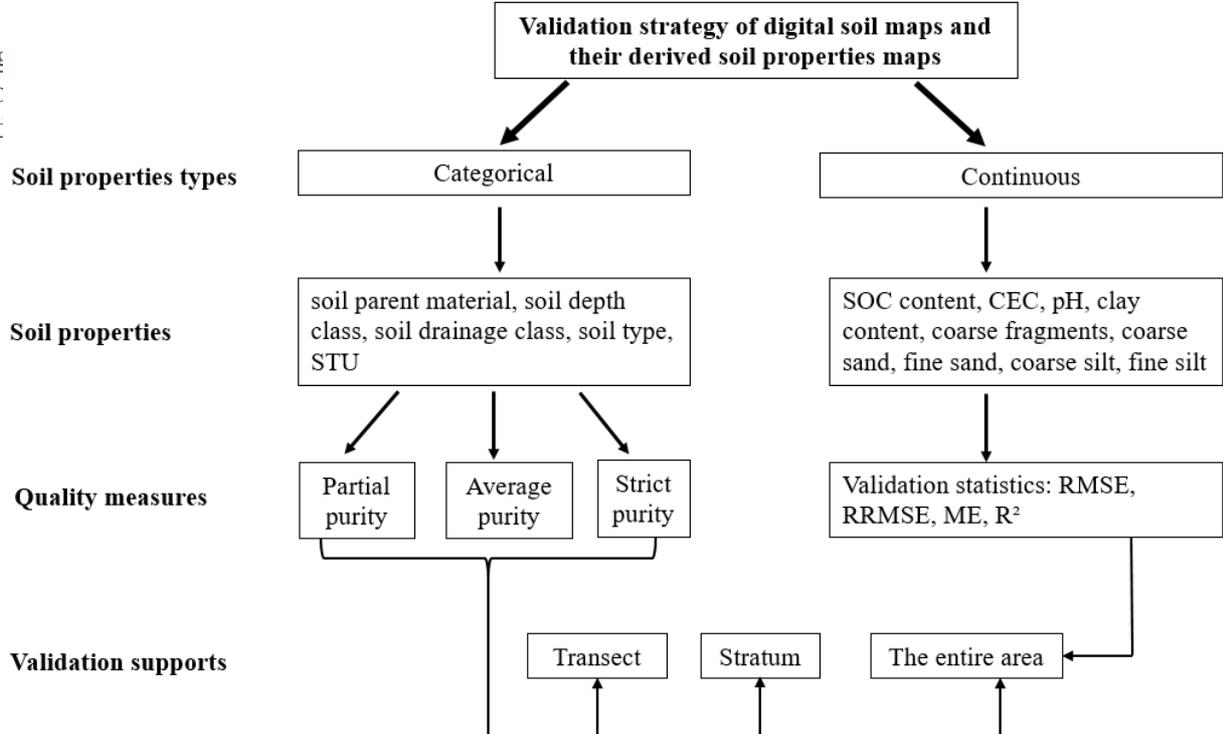


Fig. 4: Diagram of the validation strategy to compute quality measures of digital soil maps and their derived soil properties maps

807

808

809

810

811

812

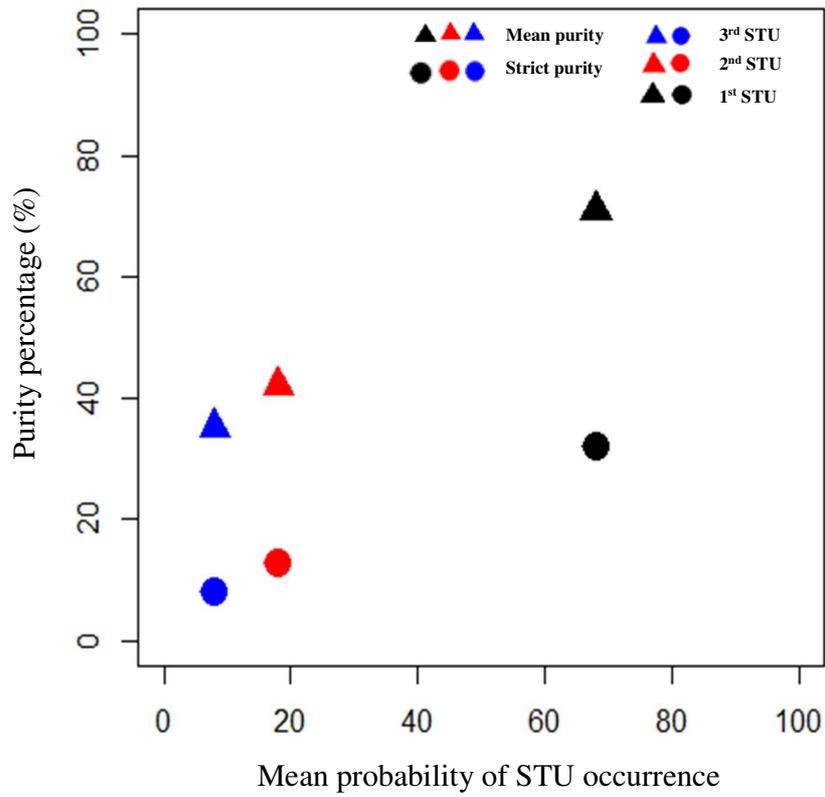
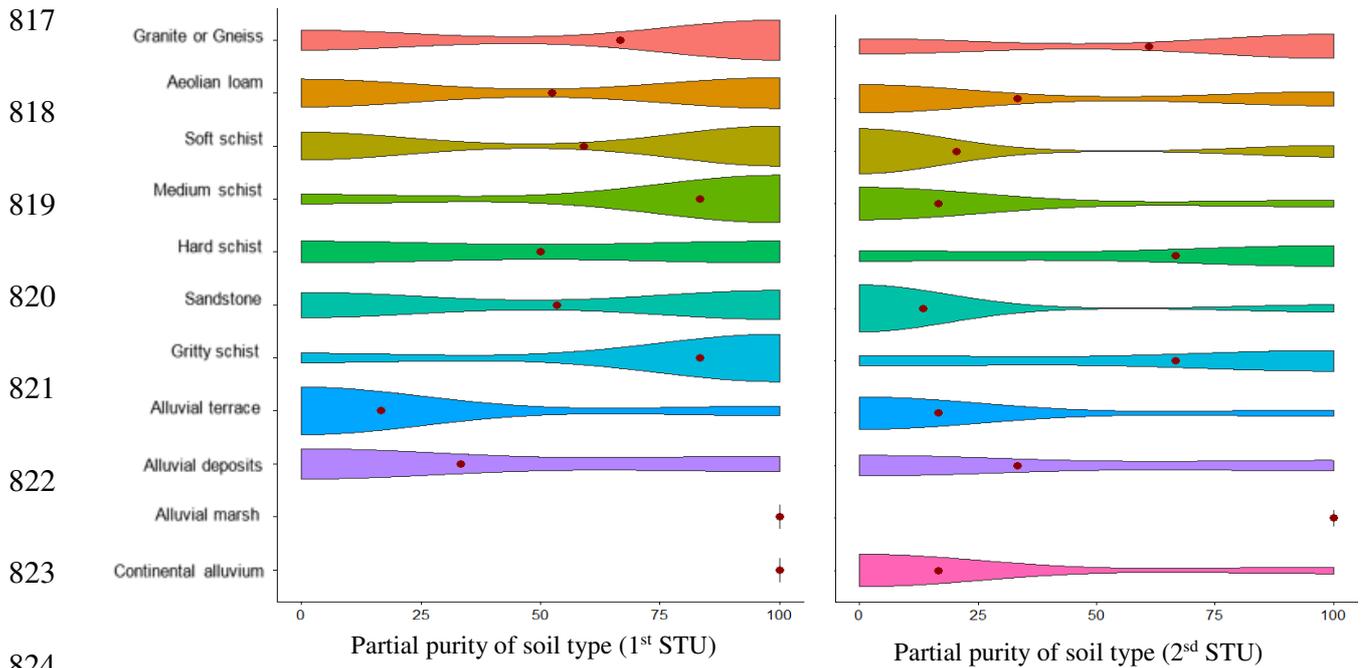
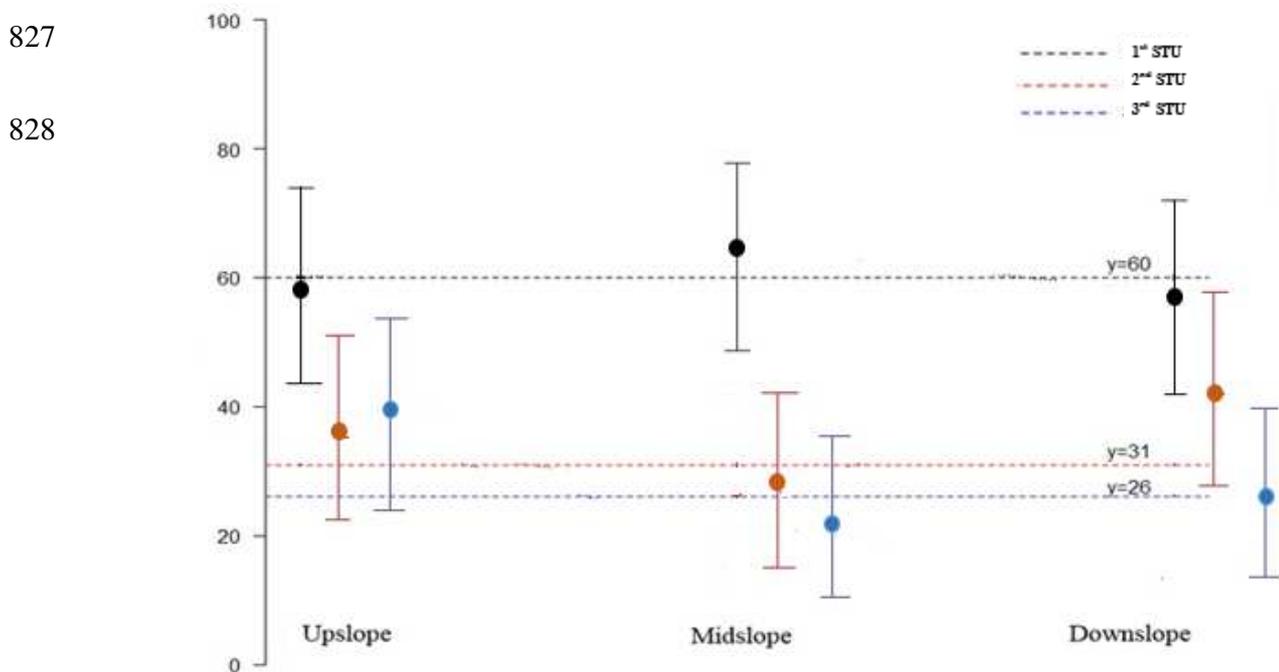


Fig. 6. Relation over the entire study area between strict and mean purities (%) and probability of occurrence of the three most probable soil typological units (STUs)

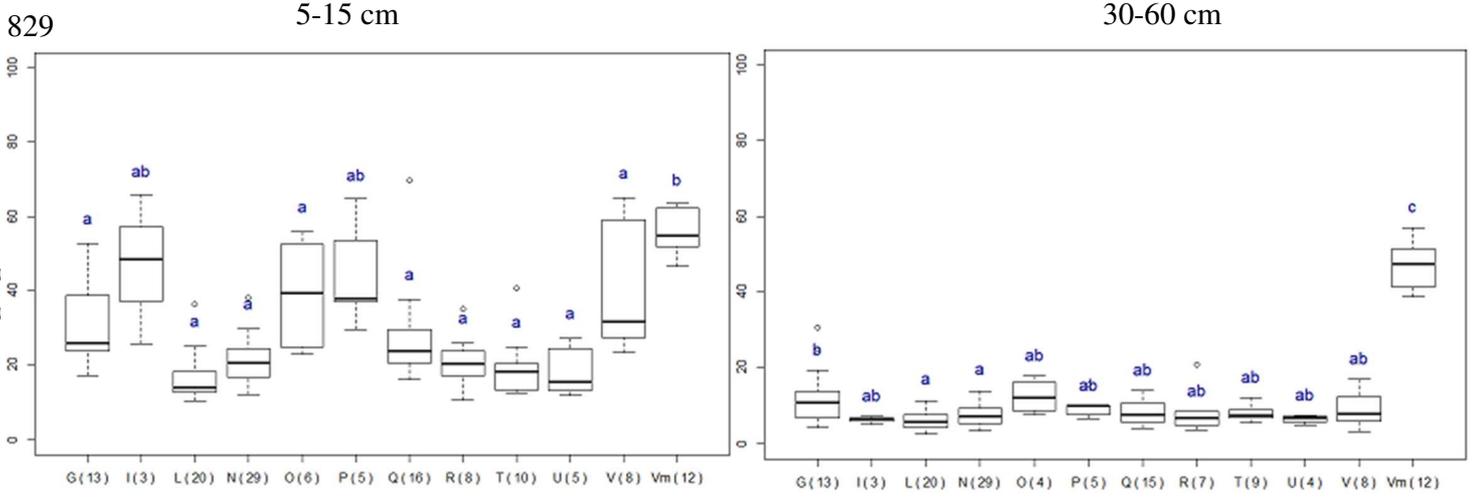


825 Fig. 7. Distribution functions of soil type purity and their mean values (red dots) by stratum  
 826 for the (left) 1<sup>st</sup> and (right) 2<sup>nd</sup> soil typological unit (STU) predicted

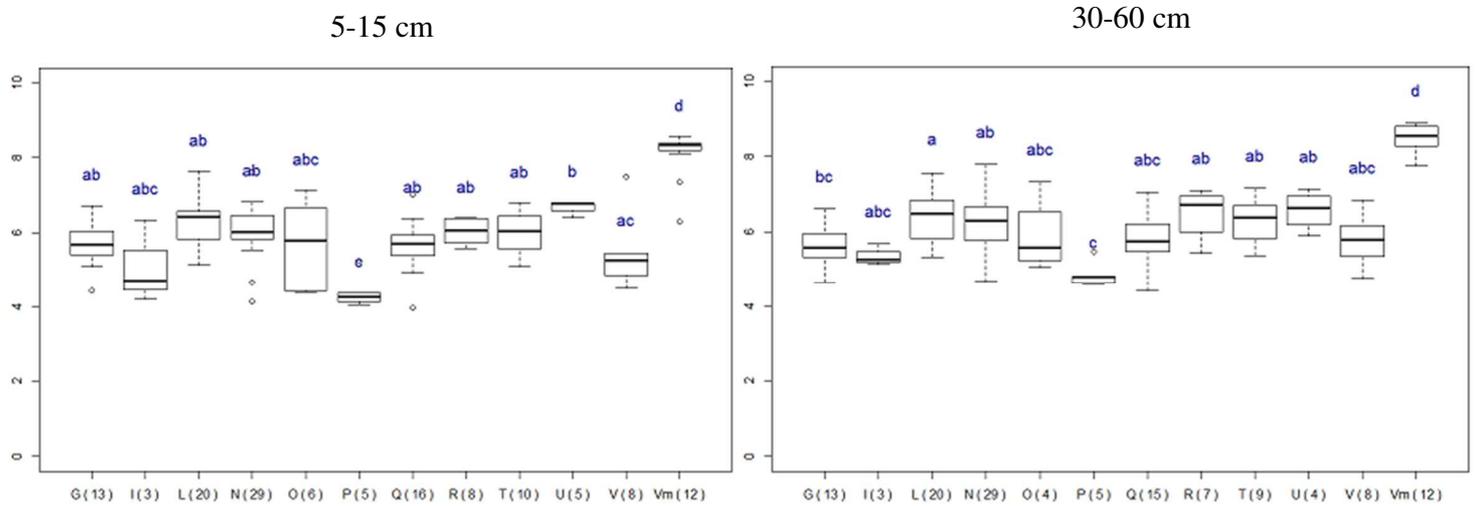


827  
 828 Fig. 8. Mean percentage of good soil type prediction and their 95% confidence intervals for the  
 three most probable soil typological units (STUs) according to hillslope position. Dashed lines  
 denote the overall soil type purity for the entire study area.

SOC content ( $\text{g kg}^{-1}$ )



pH



Clay content ( $\text{g kg}^{-1}$ )

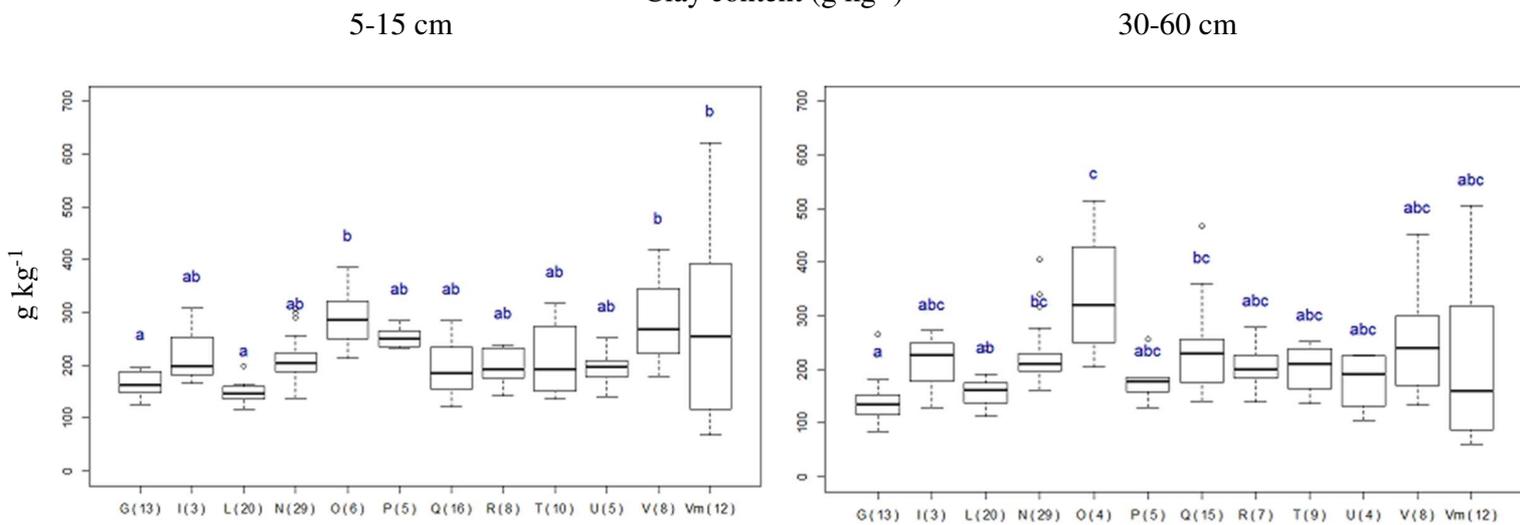


Fig. 9. Distribution of SOC content, pH and clay content according to observed parent material for the validation dataset. G: granite, I: gneiss, L: aeolian loam, N: soft schist, O: medium schist, P: hard schist, Q: sandstone, R: gritty schist, T: alluvial terrace, U: colluvium deposits, V: alluvial deposits, Vm: marsh. Whiskers represent 1.5 times the interquartile range. The number in brackets is the number of soil samples collected from each parent material. Means of groups sharing a letter in the group label do not differ significantly at  $\alpha=5\%$

830 Table1: Characteristics of the 11 strata used for stratified simple random sampling to produce the  
 831 validation dataset, dominant parent material, percentage of the study area and number of transects  
 832 sampled.

<b>Stratum</b>	<b>Dominant parent material</b>	<b>Percentage of study area (%)</b>	<b>Transects (n=45)</b>
1	Granite or gneiss	12	5
2	Aeolian loam	17	7
3	Soft schist	35	13
4	Medium schist	5	2
5	Hard schist	4	2
6	Sandstone	14	6
7	Gritty schist	5	2
8	Alluvial terrace	2	2
9	Alluvial deposits	3.6	2
10	Continental alluvium	0.4	2
11	Alluvial marsh	2	2

833

834

835 Table 2: Estimated partial, strict and mean purities (%) for the three soil maps over the entire  
 836 study area.

Soil map	Partial purities					Strict purity	Mean purity
	Parent material purity	Drainage class purity	Soil type purity	Soil depth class purity	STU purity		
<b>STU1</b>	78 (6.6)	65 (4.5)	60 (4.5)	78 (4)	23 (4.4)	34	70
<b>STU2</b>	44 (4.7)	35 (3.3)	31 (3.6)	56 (4.9)	5 (2)	12	42
<b>STU3</b>	31 (4.1)	29 (2.9)	26 (3.5)	48 (4.8)	3 (1.8)	5	34

837 Numbers in brackets are design-based estimates of standard deviations for the 45 transect purity  
 838 values

839 Table 3: Design-based estimates of purities of soil classification criteria, strict purity and mean  
 840 purity for selected strata for the 1<sup>st</sup> most probable STU map

<b>Stratum</b>	Parent material purity	Drainage class purity	Soil type purity	Soil depth class purity	Strict purity	Mean purity
Granite or gneiss	72 (46)	72 (25)	67 (21)	78 (17)	39	72
Aeolian loam	81 (40)	52 (37)	52 (32)	81(37)	20	66
Soft schist	77 (42)	67 (19)	59 (27)	80 (25)	33	71
Medium schist	100 (0)	83 (23)	83 (23)	83 (23)	67	87
Hard schist	83 (40)	33 (0)	50 (23)	50 (23)	17	54
Sandstone	73 (45)	73 (43)	53 (38)	67 (0)	40	67
Gritty schist	67 (51)	50 (23)	83 (23)	83 (23)	33	71
Alluvial terrace	100 (0)	83 (23)	17 (23)	83 (23)	17	71
Alluvial deposits	83 (40)	50 (23)	33 (0)	67 (0)	33	58
Continental alluvium	100 (0)	67 (47)	100 (0)	100 (0)	67	92
Alluvial marsh	100 (0)	83 (23)	100 (0)	100 (0)	83	96

841 Numbers in brackets are design-based estimates of standard errors for the 45 transect purity  
 842 values

843

844 Table 4: Results of the  $\chi^2$  test performed to compare proportions of good prediction of  
 845 classification criteria along hillslope positions for the three soil typological unit (STU) maps  
 846 ( $\alpha=5\%$ )

	1 <sup>st</sup> STU map	2 <sup>sd</sup> STU map	3 <sup>rd</sup> STU map
Classification criterion	Probability (p value)		
Parent material	0.40	0.41	0.65
Soil type	0.80	0.30	0.14
Drainage class	0.96	0.68	0.25
Soil depth class	0.30	0.48	0.51

847

848

849

850

851

852

853

854

855

856

857

858

859

860

861

862

863 Table 5: Descriptive statistics of soil properties for the validation dataset at 5-15 and 30-60 cm (n  
 864 = 135 and 125, respectively)

Depth interval (cm)	Soil property	minimum	median	mean	maximum	standard deviation
5-15	CEC (cmol <sup>+</sup> kg <sup>-1</sup> )	2.33	7.03	9.52	112	11.58
	pH	3.95	5.99	6.07	8.58	0.99
	SOC content (‰)	10.40	23.30	30.34	275	2.97
	Coarse fragments (%)	0.00	12.63	13.40	49.81	9.93
	Coarse sand (%)	0.20	9.30	12.38	48.20	10.90
	Fine sand (%)	1.20	11.90	13.00	40.60	6.24
	Coarse silt (%)	0.70	25.90	26.90	52.30	11.27
	Fine silt (%)	0.25	25.40	24.78	40.50	7.30
	Clay (%)	6.80	19.20	20.90	62.20	8.05
30-60	CEC (cmol <sup>+</sup> kg <sup>-1</sup> )	1.49	4.88	5.90	34.4	4.20
	pH	4.41	6.18	6.25	8.90	1.00
	SOC content (‰)	2.55	7.53	11.84	56.90	12.19
	Coarse fragments (%)	0.00	14.39	16.11	52.37	12.62
	Coarse sand (%)	0.00	9.90	13.47	5.78	12.46
	Fine sand (%)	0.30	11.70	12.87	54.00	6.65
	Coarse silt (%)	0.40	25.00	26.30	49.80	11.44
	Fine silt (%)	0.21	24.80	24.00	42.00	7.92
	Clay (%)	6.00	19.10	20.50	51.40	8.20

865

866

867 Table 6: Accuracy indicators of soil property prediction at 5-15 and 30-60 cm (n = 135 and 125,  
 868 respectively, root-mean-squared error (RMSE), relative-root-mean-squared error (RRMSE), mean error  
 869 (ME) and coefficient of determination (R<sup>2</sup>))

Depth interval (cm)	Soil property	ME	RMSE	RRMSE	R <sup>2</sup>
5-15	CEC (cmol <sup>+</sup> kg <sup>-1</sup> )	-3.95	5.12	0.90	0.28
	pH	0.29	0.80	0.13	0.43
	SOC content (g kg <sup>-1</sup> )	2.15	14.00	0.65	0.07
	Coarse fragments (%)	9.53	12.84	0.86	0.13
	Sand (%)	-0.20	9.94	0.50	0.41
	Coarse silt (%)	0.74	9.20	0.41	0.25
	Fine silt (%)	-0.20	6.70	0.39	0.05
	Total silt (%)	-1.20	9.80	0.23	0.28
	Clay (%)	1.47	5.50	0.26	0.65
30-60	CEC (cmol <sup>+</sup> kg <sup>-1</sup> )	0.06	2.15	0.43	0.34
	pH	0.45	0.78	0.12	0.55
	SOC content (g kg <sup>-1</sup> )	0.32	5.70	0.69	0.05
	Coarse fragments (%)	8.15	13.98	1.54	0.06
	Sand (%)	-2.06	10.32	0.59	0.39
	Coarse silt (%)	0.19	10.18	0.53	0.17
	Fine silt (%)	-1.88	6.00	0.39	0.25

---

Total silt (%)	-1.69	10.95	0.28	0.13
Clay (%)	3.75	7.53	0.27	0.37

---

870