



**HAL**  
open science

# How far can the uncertainty on a Digital Soil Map be known?: A numerical experiment using pseudo values of clay content obtained from Vis-SWIR hyperspectral imagery

Philippe Lagacherie, Dominique Arrouays, Hocine Bourennane, Cécile Gomez, Manuel P. Martin, Nicolas P.A. Saby

## ► To cite this version:

Philippe Lagacherie, Dominique Arrouays, Hocine Bourennane, Cécile Gomez, Manuel P. Martin, et al.. How far can the uncertainty on a Digital Soil Map be known?: A numerical experiment using pseudo values of clay content obtained from Vis-SWIR hyperspectral imagery. *Geoderma*, 2019, 337, pp.1320-1328. 10.1016/j.geoderma.2018.08.024 . hal-02058231

**HAL Id: hal-02058231**

**<https://hal.science/hal-02058231v1>**

Submitted on 9 Sep 2021

**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.

1 **How far can the uncertainty on a Digital Soil Map be known? : a numerical experiment**  
2 **using pseudo values of clay content obtained from Vis-SWIR Hyperspectral imagery**

3 Philippe Lagacherie<sup>1</sup>, Dominique Arrouays<sup>2</sup>, Hocine Bourennane<sup>3</sup>, Cécile Gomez<sup>1</sup>, Manuel  
4 Martin<sup>2</sup>, Nicolas P.A. Saby<sup>2</sup>

5

6 1. LISAH, Univ Montpellier, INRA, IRD, Montpellier SupAgro, Montpellier, France

7 2. Infosol, INRA, 45075 Orléans, France

8 3. UR Sols, INRA, 45075 Orléans, France

9

10 Corresponding author: Philippe Lagacherie, LISAH, INRA, 2 place Viala, 34060 Montpellier  
11 (France). philippe.lagacherie@inra.fr

12

13 **Abstract**

14

15 Digital Soil Map uncertainty is usually evaluated from a set of independent soil observations  
16 that are used to determine various uncertainty indicators. However, the number and locations  
17 of the sites that constitute these evaluations may impact the value of these indicators.

18 In this paper, a numerical experiment on uncertainty indicators was performed using the  
19 pseudo values of topsoil clay content obtained from an airborne hyperspectral image in the  
20 Cap Bon region (Tunisia). These pseudo values form a soil pattern with a large extent (46% of  
21 300 km<sup>2</sup>), high resolution (5 m) and good accuracy ( $R^2_{\text{val}} = 0.75$ ) while being free of visible  
22 artefacts and pedologically plausible. Therefore, the dataset was considered a fair  
23 representation of reality while providing a quasi-unlimited choice of sites.

24 The numerical experiment considered three Quantile Regression Forests as examples of DSM  
25 models by using inputs from relief soil covariates and geographical locations that were  
26 calibrated from 200, 2,000 and 100,000 individuals respectively (low, medium and high quality  
27 models). Their uncertainty indicators were first evaluated by calculating four uncertainty  
28 indicators (ME, MSE,  $SS_{MSE}$  and PICP) from a large independent validation set of 100,000 sites.  
29 These uncertainty indicators were then computed from independent evaluation sets of  
30 different sizes (from 50 to 500 sites) and from different locations (500 evaluation sets of each  
31 size). The independent evaluation sets were selected following a stratified random sampling  
32 using compact geographical strata.

33 The numerical experiment showed that the values of the uncertainty indicators were highly  
34 variable across numbers and locations of sites. The largest variations were observed for  
35 evaluation sets with fewer than 100 sites, but non-negligible variations remained for larger  
36 evaluation datasets. This result suggested that evaluations from independent sets convey a  
37 non-negligible error on the uncertainty indicators, which increases as the number of sites  
38 decrease.

39 Evaluations of DSM models from independent evaluation sets should be interpreted with care  
40 and uncertainty on validation results should be systematically estimated. For that, numerical  
41 experiments for benchmarking DSM models on known soil patterns across the world would  
42 be a valuable complement to the analytical expressions for the uncertainty indicators and the  
43 many DSM applications for which these analytical expressions are not valid. This would serve  
44 also to improve the sampling techniques for the calibration and evaluation datasets to reduce  
45 the error when estimating the uncertainty of a DSM product.

46

47 **Keywords:** Soil mapping, Uncertainty, Hyperspectral imagery, Random forest, Sampling

48

49

50 **1. Introduction**

51

52 Soil maps are simplified representations of more complex and partially unknown patterns of  
53 soil variations. Therefore, any prediction of a soil property that can be derived from these soil  
54 maps has an irreducible and often substantial error that is the difference between the true  
55 and estimated values of the soil property. Since there is no way to systematically measure this  
56 error, we are uncertain about the true value of the soil property at most of the locations,  
57 where uncertainty refers to the state of mind of a person who expresses a lack of confidence  
58 about reality (Heuvelink, 2014).

59 Getting accurate estimates of this uncertainty is of paramount importance for end-users to  
60 make enlightened decisions on the utility and limitations of the soil data products delivered  
61 to them. An example of a concrete translation of this exigence is provided by the  
62 GlobalSoilMap specifications (Arrouays et al., 2014). Following these specifications, each soil  
63 property estimate should be provided under the form of a 90% prediction interval (PI), which  
64 reports the range of values in which the true value is expected to occur 9 times out of 10.

65 To provide this uncertainty information, a rigorous assessment of the uncertainty is  
66 considered a mandatory component of any Digital soil mapping application. Two main groups  
67 of methods can be used to accomplish this task (Heuvelink, 2014). The first is a model-based  
68 approach that involves spatial stochastic models that can provide estimates of the uncertainty  
69 of their own predictions. The main drawback of these model-based approaches is that their  
70 uncertainty evaluations are only valid under certain assumptions of the stochastic models.

71 They also use the same data that were used to calibrate the models, which may cause an

72 underestimation of the uncertainty. The second group of uncertainty evaluation methods  
73 avoids this restriction by undertaking a model-free statistical evaluation with independent  
74 sites selected by probability sampling. The comparison between the predicted and actual  
75 values of the soil attributes of interest yields a set of uncertainty indicators, the most common  
76 of which are the mean error (ME), the mean squared error (MSE), the root mean square error  
77 (RMSE) and the coefficient of determination ( $R^2$ ). The latter is confusingly referred to in the  
78 DSM literature either as the goodness-of-fit of a linear regression between the predicted and  
79 observed values, or to how close the paired prediction-observation points are to the 1:1 line.  
80 For DSM models that deliver predictions via probability distributions, accuracy plots  
81 (Goovaerts, 2001) or Prediction Interval Coverage probability (PICP) (Shrestha and Solomatine  
82 2006), it can also be calculated to evaluate the accuracy of the uncertainty attached to each  
83 prediction.

84 Various methods for producing sets of independent sites (denoted further 'evaluation sets'),  
85 including the data-splitting of the available dataset, cross-validation over the available dataset  
86 and the probabilistic sampling of additional sites, were reviewed by Brus et al (2011). They  
87 recommended the latter to ensure unbiased uncertainty estimations and provide a more  
88 complete view of the spatial distribution of the error via the Spatial Cumulative Distribution  
89 Function (SCDF).

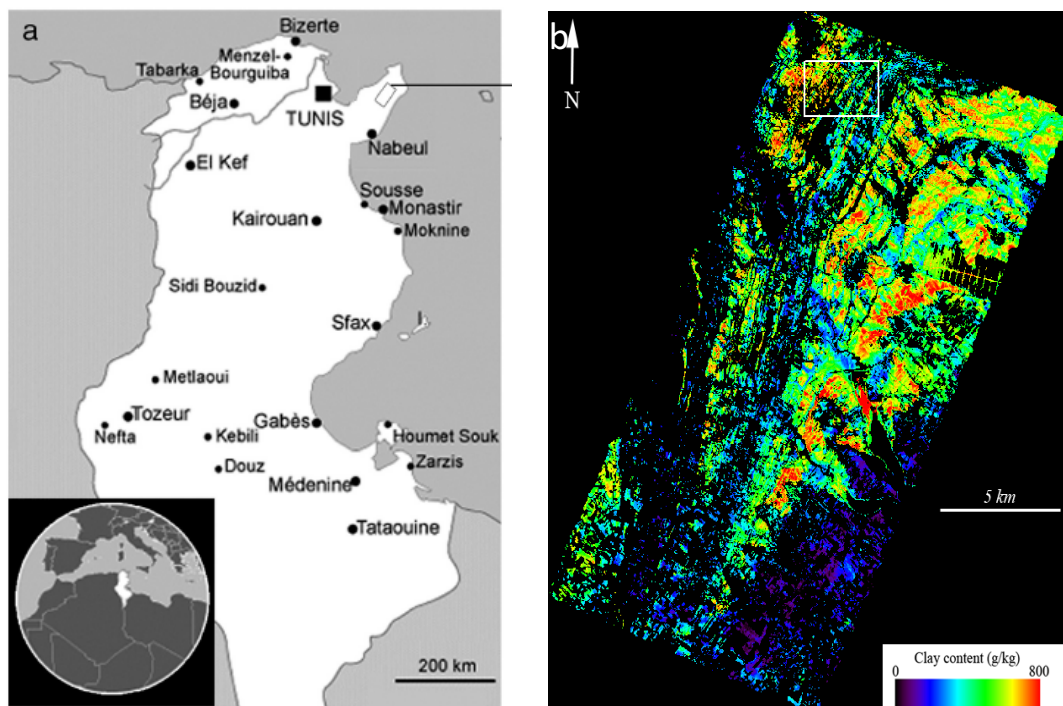
90 The main drawback of any model-free statistical evaluation is that the calculated uncertainty  
91 indicators are themselves prone to uncertainty. Indeed, similar to any statistical parameters  
92 that are derived from a set of individuals, the uncertainty indicators (e.g.,  $R^2$ , ME, and PICP)  
93 are sensitive to the number and the locations of the soil observations used for calculating  
94 them. Under the condition that a probabilistic sampling is applied, Brus et al (2011) provided  
95 an analytical expression of the mean error (ME) and the mean square error (MSE) variances

96 using the sampling fraction (the ratio between the number of samples and the number of  
97 possible sample locations) and the estimated variance of the error over a given area as inputs.  
98 Such analytical expressions of the ME and MSE variances measure the uncertainty of these  
99 uncertainty indicators, which could be very useful. For example, they could be used for  
100 determining whether the differences between the ME and MSE of two different DSM models  
101 reveal significantly different performances or for computing the size of the validation sample  
102 that is required for estimating the ME and MSE at a given precision level. However, it must be  
103 noted that i) for some indicators other than ME and MSE and ii) for some specific sampling  
104 designs (such as two-stage random sampling), there might not be readily available statistical  
105 estimates of the uncertainty of the uncertainty indicators (Brus et al. (2011)).

106 Although the uncertainty of the ME and MSE is easily computable for some sampling designs,  
107 it is far from being currently computed in DSM applications. This is because the soil datasets  
108 used as inputs of DSM models are rarely suitable for applying a probabilistic sampling. Indeed,  
109 these soil datasets are often undersized with regard to the size of the study area and the  
110 complexity of the soil cover to be modelled. This leads to substantial losses of predictive  
111 performances as soon as the sampling effort for collecting calibration sites is depleted for  
112 populating the set of independent sites required by probabilistic sampling. Furthermore, DSM  
113 applications most often use legacy data that do not respect the randomness and evenness  
114 required for a probabilistic sampling. For all these reasons, the DSM mappers most often  
115 disregard calculating any uncertainty in their uncertainty indicators and therefore neglect this  
116 issue when evaluating the DSM products.

117 This paper presents a numerical experiment for assessing the uncertainties of the uncertainty  
118 indicators of the three DSM models with contrasted predictive performances by using  
119 different probabilistic samplings of different sizes. To overcome the above-evoked limitations

120 of the current soil input data, the study used the virtual pattern of the topsoil pseudo clay  
121 content derived from airborne Vis-NIR-SWIR hyperspectral data acquired over the Cap Bon  
122 region (300 km<sup>2</sup>, Tunisia) at a five-meter resolution (Gomez et al., 2015). This pattern is  
123 constituted of well-predicted clay values ( $R^2=0.75$ ) that are free of visible artefacts and  
124 pedologically plausible, which allows it to be considered as a fair representation of the  
125 variations of a real soil property across the landscape. Such a soil dataset provided a quasi-  
126 unlimited number of pseudo-measured sites that made probabilistic sampling (and therefore  
127 ME and MSE variance calculations) applicable without any effect on the predictive  
128 performances. It also enabled the calculation of any uncertainty indicators from their  
129 empirical distributions obtained by repeating the validation process  $n$  times, which means  
130 selecting  $n$  different validation sets of a given size and determining the uncertainty indicators  
131 each time.



132  
133 Figure 1: Location of the study area (a) and the spatial pattern of pseudo values of topsoil clay content  
134 (b)

135

## 136 2. Material and methods

137

### 138 2.1. The study area

139

140 The study area is in the Cap Bon region in northern Tunisia (36°24'N to 36°53'N; 10°20'E to  
141 10°58'E), which is 60 km east of Tunis ([Figure 1a](#)). This 300 km<sup>2</sup> area includes the Lebna  
142 catchment, which is mainly rural (>90%). The Lebna catchment is devoted to the cultivation  
143 of cereals in addition to legumes, olive trees, vineyards and natural vegetation for animals.  
144 The region is characterized by its rolling hills and elevations between 0 and 226 m. The climate  
145 varies from humid to semi-arid, with an inter-annual precipitation of 600 mm and an inter-  
146 annual potential evapotranspiration of 1500 mm. The soil pattern of the Lebna catchment is  
147 mainly the result of variations in lithology. The variations in the bedrock between Miocene  
148 sandstone and Marl cause large variations in the physical and chemical soil properties ([Zante  
149 et al., 2005](#)). Furthermore, the distance between successive sandstone outcrops decreases  
150 significantly as the terrain changes from the ocean to the mountains, which also causes  
151 variations in the soil property patterns ([Gomez et al., 2012b](#)). The soil materials were  
152 redistributed laterally along the slopes during the Holocene, which adds to the complexity of  
153 the soil pattern. The main soil types are regosols (IUSS working group WRB, 2006)), which are  
154 preferentially associated with sandstone outcrops, and calcic cambisols and vertisols, which  
155 preferentially formed on marl outcrops and lowlands. The southeastern region of the study  
156 area has a flatter landscape with sandy Pliocene deposits in which calcosols and rendzinas  
157 prevail.

158



159 2.2. Data

160

161 2.2.1. Hyperspectral image and derived topsoil clay content predictions

162

163 The numerical experiment uses an image of topsoil clay content as input. The data were  
164 derived from a Vis-NIR-SWIR hyperspectral image (Gomez et al., 2012b). The approach used  
165 to produce the data is summarized below. More details about the pre- and post-processing of  
166 the hyperspectral image can be found in Gomez et al (2012b).

167 On November 2, 2010, AISA-Dual airborne-based hyperspectral data were acquired over the  
168 study area with a spatial resolution of 5 m. The area of the image is approximately 12 km x 24  
169 km. The AISA-Dual spectrometer measured the reflected radiance via 359 non-contiguous  
170 bands covering the 400 to 2450 nm spectral range, with 4.6 nm bandwidths between 400 and  
171 970 nm and 6.5 nm bandwidths between 970 and 2450 nm. The instantaneous field of view  
172 (IFOV) was 24 degrees. Topographical corrections were performed using a digital elevation  
173 model built from ASTER data and ground control points.

174 To isolate the bare soil areas, the study masked pixels with normalized difference vegetation  
175 index (NDVI) values greater than an expert-calibrated threshold (0.20). Water and Urban areas  
176 were also removed. Finally, the bare soil represented 46.3% of our study area and potentially  
177 5,889,847 measured AISA-Dual 5 m x 5 m pixels.

178 A Partial Least Square Regression (PLSR) technique (Tenenhaus, 1998) was then applied to  
179 estimate the topsoil clay contents from the 280 reflectance bands provided by the AISA-DUAL  
180 airborne sensor at each location. The PLSR was calibrated from 129 couples of Vis-NIR-SWIR  
181 reflectance spectra acquired by the AISA-DUAL sensor on bare soil surfaces associated with  
182 the topsoil clay content measured on a laboratory soil sample collected from the same bare

183 soil surfaces. Before the PLSR model was built, the reflectance was converted into  
184 “absorbance” ( $\log [1/\text{reflectance}]$ ). In addition, a Savitzky–Golay filter with second-order  
185 polynomial smoothing and window widths of 30 nm (Savitzky and Golay, 1964) and a mean  
186 centering and variance scaling was applied to the spectra to reduce noise. The calibrated PLSR  
187 model was then validated using a leave-one-out cross-validation that showed successful  
188 predictions ( $R^2 = 0.75$ ). The PLSR model was then applied to all bare soil pixels to estimate the  
189 topsoil clay content, thus providing the final predicted topsoil clay properties map (Figure 1b),  
190 which is denoted “pseudo values of Clay content”. These treatments were implemented in R  
191 (Version 1.17) using the signal and pls packages (Mevik and Wehrens, 2007).

192

#### 193 2.2.2. Digital Elevation Model and derivatives

194

195 A 30-m ASTER digital elevation model (DEM) with specific ortho-rectification and mosaicking  
196 was produced for this area. The classical geomorphometric indicators found in the DSM  
197 literature were calculated. These include Elevation, Slope, Aspect, plan Curvature, Profile  
198 Curvature and Multi-Resolution Valley Bottom Flatness (MRVBF). Sine and cosine  
199 transformations were applied to the ‘aspect’ to obtain four indices with a continuous gradient:  
200 ‘northness’, ‘easterness’, ‘north-westerness’ and ‘north-easterness’. Finally, the X and Y  
201 coordinates (the n of “scorpan” in McBratney et al., 2003) were also used as soil covariates.

202

#### 203 2.3. Checking the plausibility of the predicted soil patterns

204

205 We conducted a prior check to confirm the reliability of using the pseudo values of the clay  
206 content determined above as a realistic example of a soil property pattern. It was particularly

207 important to check the absence of any distortion of the spatial pattern due to the spectral  
208 measurements by remote sensing. Three experimental variograms showing the spatial  
209 structure of clay variations were calculated and then fitted with an exponential model using  
210 the weighted least square method (Cressie, 1993) using three different data sets: i) the 129  
211 sites with laboratory measurements of the clay content, ii) the same sites with pseudo values  
212 of the clay content and iii) 100,000 randomly selected sites with pseudo values of the clay  
213 content. Comparisons of the variograms were performed (see section 3.1.).

214

#### 215 2.4. Sampling technique

216

217 We used probability sampling for selecting the calibration set used to build the DSM model,  
218 evaluating the performance of these models and undertaking the numerical experiment. All  
219 of the probability sampling techniques followed the same sampling approach (the stratified  
220 random sampling technique) using compact geographical strata (Walvoort et al., 2010)  
221 recommended by Brus et al. (2011). The main advantage of this technique is that it ensures  
222 an even distribution of the samples over the studied area and is simple to apply.

223 The stratified random sampling approach was applied as follows. The study area was first  
224 stratified into 25 geographical strata of equal area using a K-means classification of the X and  
225 Y coordinates of each locations. We then randomly selected pixels of the grid within each  
226 stratum with a fixed number of locations in accordance with the total of samples required.

227

#### 228 2.5. Uncertainty indicators

229

230 We considered four uncertainty indicators among the possible ones that could have been  
 231 examined. The first two were the mean error (ME) and the mean squared error (MSE) that  
 232 were selected because these are the classical indicators whose variance can be calculated  
 233 from analytical expressions (Brus et al, 2011). The last two were the mean square error skill  
 234 score ( $SS_{mse}$ ) (Wilks, 2011 p 359, cited by Nussbaum et al, 2017)) and the Prediction Interval  
 235 Coverage probability (PICP) (Shrestha and Solomatine 2006).  $SS_{MSE}$  is similar to the  $R^2$  reported  
 236 in some studies (Vaysse and Lagacherie, 2015, Viscarra-Rossel, 2015) as the percentage of  
 237 variance explained by the model:

$$238 \quad SS_{mse} = 1 - \frac{\sum_{i=1}^{i=n} (z_i - z_i^*)^2}{\sum_{i=1}^{i=n} (z_i - \hat{z}_i)^2} \quad (1)$$

239 Where  $z_i$  and  $z_i^*$  are the respective observed and predicted values of property  $z$  at location  $l$ ,  
 240 and  $\hat{z}_i$  is the mean value of  $z$ .

241 The PICP expresses the probability that all observed values fit within the 90% prediction limits  
 242 provided by the DSM model (see section 2.7.1.).

243 It must be noted that the first three uncertainty indicators (ME, MSE and  $SS_{MSE}$ ) are  
 244 measurements of the accuracy of predictions, whereas PICP is a measurement of the accuracy  
 245 of the uncertainty prediction.

246 The calculations of these indicators should take into account the fact that a stratified random  
 247 sampling technique is applied for selecting the independent sites. Following Brus et al (2011),  
 248 these calculations are as follows.

249 The estimations of the ME, MSE,  $SS_{MSE}$  and PICP correspond to a global mean that was  
 250 estimated by design-based inference, particularly by the usual estimator for stratified random  
 251 sampling.

$$252 \quad \hat{y} = \sum_{h=1}^H w_h \hat{y}_h \quad (2)$$

253 
$$\hat{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi} \quad (3)$$

254 Where  $H$  is total number of strata ( $H = 25$ ),  $w_h$  is the weight of stratum  $h$  quantified by the  
 255 relative area,  $\hat{y}_h$  is the estimated mean of the stratum  $h$ ,  $n_h$  is the number of sampling points  
 256 in stratum  $h$ , and  $y_{hi}$  is the measurement of the indicator at location  $i$  in stratum  $h$ . For the  
 257 ME,  $y_{hi}$  can be replaced by the difference between the actual value  $z_{hi}$  and the  
 258 prediction  $z^*_{hi}$ . The MSE can be estimated by replacing  $y_{hi}$  with the squared difference  
 259 between the actual value  $z_{hi}$  and the prediction  $z^*_{hi}$ . The  $SS_{MSE}$  and the PICP can be estimated  
 260 by replacing  $y_{hi}$  by the  $SS_{MSE}$  and the PICP of the stratum  $h$ .

261

262 2.6. Analytical calculations of standard errors of ME and MSE

263 If probabilistic sampling is applied, it is possible to calculate the standard errors of the ME and  
 264 MSE. In the case of a stratified random sampling, the equations are (from De Gruijter, 2006)

265

266 
$$\widehat{SE}(\bar{y}) = \sqrt{\sum_{h=1}^H a_h^2 \hat{V}(\hat{y}_h)} \quad (4)$$

267 Where  $\hat{V}(\hat{y}_h)$  is the sampling variance of the stratum mean  $\hat{y}_h$ , which is estimated by :

268 
$$\hat{V}(\hat{y}_h) = \frac{s_h^2}{n_h} \quad (5)$$

269 
$$s_h^2 = \frac{1}{n_h-1} \sum_{i=1}^{n_h} (y_{hi} - \hat{y}_h)^2 \quad (6)$$

270

271

272 2.7. DSM modelling

273

274 Since the validation process tested in this paper is model free, any model used in DSM could  
 275 have been selected as an example of a DSM model. However, only two criteria were

276 considered: i) the model had to provide local uncertainty predictions for being able to test  
277 PICP, and ii) the model should be run without manual intervention and repeated a great  
278 number of times in the numerical experiment. We combine these two criteria results in  
279 selecting the Quantile Regression Forest as the example DSM model.

280

### 281 2.7.1. Random Forests and Quantile Random Forests

282

283 This section describes Random Forests and Quantile Random Forests. More details on these  
284 two machine learning algorithms are given in the seminal papers by Breiman et al (2001) and  
285 by Meinshausen (2006), respectively.

286 Let  $Y$  be a real-valued response variable and  $X$  be a covariate or predictor variable that is likely  
287 high-dimensional. A standard goal of statistical analysis is to infer the relationship between  $Y$   
288 and  $X$ . Random Forests grow a large (>500) ensemble of trees using  $n$  independent  
289 observations  $(Y_i, X_i), i = 1, \dots, n$ . Each tree grows via a recursive partitioning of the source set  
290 using one predictor variable  $X$ . At each step, the source set is split into two subsets following  
291 a test on the value of  $X$ . When  $Y$  is a quantitative variable, the selected test is the one that  
292 minimizes the within subset variance of  $Y$  (Breiman et al., 1984). The recursive partitioning is  
293 limited by a stopping rule, and the subsets are produced by the last split being the leaves of  
294 the tree. The ensemble of trees is produced by using a random sample of the training data  
295 and a random subset of the predictor variables for each tree.

296 For the regression, the prediction  $\hat{Y}_\theta(x)$  of a single tree  $\theta$  of a Random Forest for a new data  
297 point  $x$  can be represented as the weighted average of the original observations  $Y_i, i = 1, \dots,$   
298  $n$  :

299

300 
$$\hat{Y}_\theta(x) = \sum_{i=1}^n w_{\theta_i}(x, \theta) Y_i \quad [7]$$

301

302 where  $w_{\theta_i}(x, \theta)$  is the weight vector given by a positive constant that is one if the observation  
 303  $Y_i$  is part of the same leaf and is 0 otherwise.

304 By using Random Forests, the prediction is the average prediction of  $k$  single trees that were  
 305 constructed as described above.

306 
$$\widehat{Y}_T(x) = \sum_{i=1}^n w_{T_i}(x) Y_i \quad [8]$$

307 With 
$$w_{T_i}(x) = k^{-1} \sum_{t=1}^k w_{\theta_i}(x, \theta) \quad [9]$$

308

309 One could assume that the weighted observations deliver a good approximation not only of  
 310 the conditional mean but also of the full conditional distribution. This assumption is at the  
 311 heart of the Quantile Regression Forest algorithm, which estimates the conditional  
 312 distribution function of  $Y$  given  $x$  via

313

314 
$$\hat{F}(y|x) = \sum_{i=1}^n w_i(x) 1_{\{Y_i \leq y\}} \quad [10]$$

315

316 From this conditional distribution, it is possible to derive both the predicted value (the mean)  
 317 and the bound of the 90% prediction interval that predicts the associated uncertainty (the  
 318 0.05 and 0.95 quantiles).

319

### 320 2.7.2. Calibration and Validation of QRFs

321

322 Three different QRF were calibrated with 200, 2,000 and 50,000 sites with known pseudo

323 values of clay content. The locations were selected according to the stratified sampling  
324 techniques described above. The increasing number of sites was selected to obtain contrasted  
325 predictive performances (see section 3.2.)

326 After removing the calibration sites, we selected a master evaluation set of 100,000  
327 independent sites by applying the stratified random sampling technique using the compact  
328 geographical strata described in section 2.5. The reference values of the three uncertainty  
329 indicators of interest were computed from this master evaluation set. This set was then  
330 removed from the set of possible sites to ensure the independence of the further numerical  
331 experiment.

332

## 333 2.8. Empirical simulation

334

335 The empirical simulation aims to evaluate the amount of variation in the four uncertainty  
336 indicators (ME, MSE,  $SS_{MSE}$  and PICP) when different evaluation sets are selected. This  
337 variation can then be used as an estimate of the uncertainty caused by relying on the choice  
338 of a single specific evaluation set, as is always the case in reality.

339 The empirical simulation proceeds as follows:

- 340 1. Sample a set of  $n$  evaluation sites using a stratified random sampling technique using  
341 compact geographical strata,
- 342 2. Calculate the uncertainty indicators over the  $n$  sites,
- 343 3. Repeat steps 1 and 2 500 times, and
- 344 4. Compute the distributions and their summary statistics from the 500 values of the  
345 uncertainty indicators.



346 The tested numbers of sites ranged between  $n = 50$  and  $n = 500$  with an increment of 25. This  
347 represents the densities of the observations ranging between  $1/2.67 \text{ km}^2$  and  $1/0.27 \text{ km}^2$ .

348

## 349 2.9. Software

350

351 The software for Random Forests and Quantile Random Forest are made available in R (R  
352 Development Core Team, 2008) with the packages RandomForest (Liaw and Wiener, 2002)  
353 and quantregForest (Meinshausen and Schiesser, 2015), respectively. Stratified sampling  
354 using compact geographical strata is implemented in the R package "spcosa" (Walvoort et al.,  
355 2010). Variogram studies (section 3) were performed with the gstat package (Pebesma, 2004).

356

## 357 3. Results

358

### 359 3.1. Check of the predicted soil patterns

360

361 To check the plausibility of using the pseudo values of clay content derived from hyperspectral  
362 data, we compared the experimental variograms and fitted model variograms obtained using  
363 real clay content measurements at measured sites, the pseudo values of the clay content at  
364 the same locations, and a set of 10,000 sites with the pseudo values of the clay content.

365 Figure 2 shows the experimental and the fitted variograms of the topsoil clay content  
366 calculated from different inputs. The parameters of the variogram estimated from the pseudo  
367 values of clay content (figure 2b) were similar to those estimated with the real clay content  
368 measurements at the same locations (figure 2a). Indeed, the shapes, ranges and sills were

369 close to one another. The only noticeable difference was a smaller nugget value exhibited by  
 370 the variogram of the pseudo values of the clay contents.  
 371 The experimental variogram obtained from 100,000 sites was much less noisier than but very  
 372 similar to the previous one.

373

374 Figure 2: Variograms of clay content obtained from a) 129 sites with clay content laboratory  
 375 measurements, b) the same 129 sites with pseudo values of clay content c) 100,000 sites with pseudo  
 376 values of clay content.  
 377

378

379

Number of calibration sites	Uncertainty indicators			
	ME (g/kg)	MSE(g <sup>2</sup> /kg <sup>2</sup> )	SS <sub>MSE</sub>	PICP (%)
200	-11.3 (0.5)	19555 (89)	0.29	88.5
2,000	-7.0 (0.4)	13550 (69)	0.51	90.3
50,000	-2.7 (0.3)	6090 (42)	0.78	90.3

380

381 *Table 1: Estimated ME (standard error), MSE (standard error),  $SS_{MSE}$ , and PICP from the master set of*  
382 *independent sites*

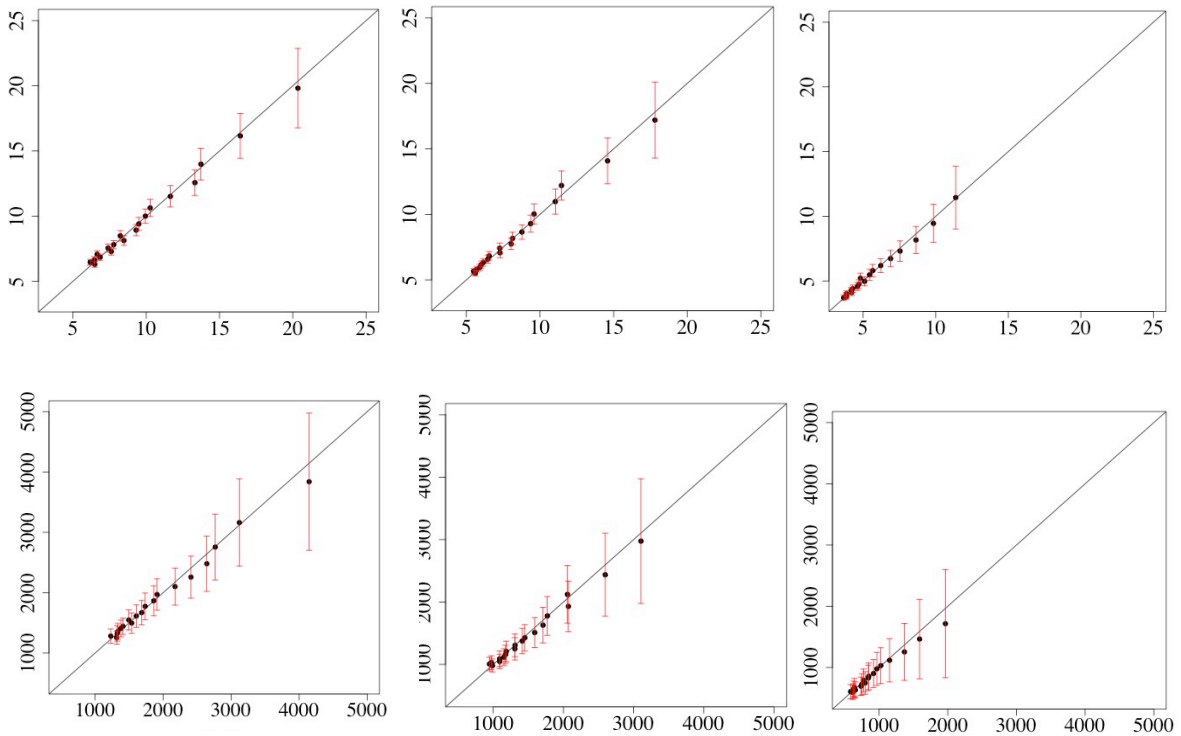
383

### 384 3.2. DSM model performances

385

386 Table 1 shows the values of the uncertainty indicators calculated from the master set of  
387 independent sites (100,000 sites) for the DSM models obtained by calibrating the quantile  
388 regression Forests with three sizes of calibration sets. As expected, the overall accuracy of the  
389 measured predictions (as measured by  $SS_{MSE}$ ) increased significantly as the number of  
390 calibration sites increased, while the bias measured by the ME and mean squared error (MSE)  
391 decreased. The PICP values were found to be close to the expected value of 90 for the two  
392 models with the greatest numbers of calibration sites. Meanwhile, the model built from 200  
393 calibration sites exhibited a PICP below 90, which revealed a slight underestimation of the  
394 uncertainty. Finally, the results obtained by the three models well covered the large range of  
395 performances of DSM models that can be encountered in the literature. It is also worth noting  
396 that the standard errors of the ME and MSE that were calculated from the variances given by  
397 equations 8 and 9 are very small, which means that the performances of the three models are  
398 significantly different from each other.

399 In the following, the QRFs calibrated from 200, 2,000 and 50,000 sites are denoted,  
400 respectively, as “low-quality QRF”, “medium-quality QRF” and “high-quality QRF”.



401

402 Figure 3: Comparisons between the estimations of standard errors of ME and MSE derived from the  
 403 numerical experiment (X-axis) and their analytical calculations (Y-axis). ME (first row), MSE (second  
 404 row), low quality QRF (first column), medium quality QRF (second column) and high-quality QRF (third  
 405 column). Dots : mean values of the calculated standard errors on ME and MSE over the 500 iterations,  
 406 bars: twice the standard deviations of the calculated standard errors on ME and MSE over the 500  
 407 iterations)

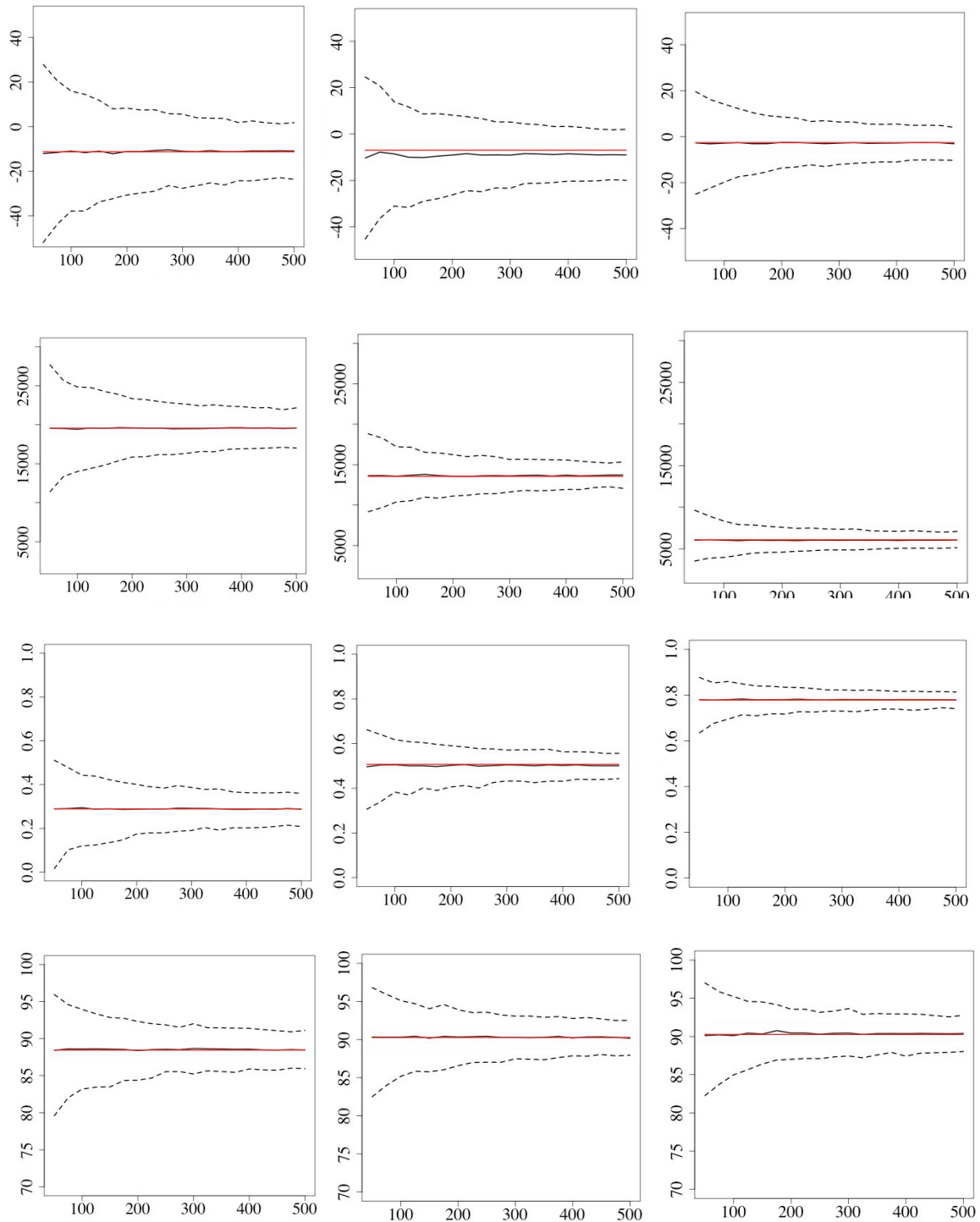
408

### 409 3.3. Comparisons of estimated standard errors of ME and MSE

410 For each DSM model, the standard deviations of the ME and MSE observed over the 500  
 411 iterations for the different sizes of evaluation sets were compared with the standard errors of  
 412 the ME and MSE calculated following equations 4, 5 and 6 (figure 3).

413 The dots of figure 3 representing the mean values over the 500 iterations of the calculated  
 414 standard errors of the ME and MSE were found to be very close to the 1:1 line. This denoted  
 415 good agreement between the analytical calculations of the ME and MSE and the simulation  
 416 outputs, which was expected. Interestingly, the error bars around the dots of figure 3

417 representing twice the standard deviation over the 500 iterations of the analytical calculations  
418 of the standard errors of the ME and MSE were great for the largest standard errors and  
419 decreased as the standard errors decreased, regardless of the indicator and the DSM model.  
420 This revealed the residual impacts of the evaluation sites' locations that were selected by the  
421 spatial sampling, the DSM model and the size of the evaluation set being fixed.



422

423 Figure 4: Values of uncertainty indicators obtained using validation sets of different sizes (from 50 to  
 424 500 sites): ME (first row), MSE (second row),  $SS_{MSE}$  (third row) and PICP (fourth row), Low quality QRF  
 425 (first column), medium quality QRF (second column) and high-quality QRF (third column). Black line:

426 mean value of the uncertainty indicators over 500 trials. Dotted black lines: 0.05 and 0.95 quantiles  
427 of the uncertainty indicators over 500 trials.. Red line: reference values of the uncertainty indicators

### 428 3.4. Numerical experiment results

429

430 Figure 3 shows the result of the numerical experiment on the three tested QRFs and the four  
431 uncertainty indicators. Each result consists of three graphical curves showing the evolution of  
432 the median (black line) and the 0.05 and 0.95 quantiles (dotted black lines) of the uncertainty  
433 indicator values, with the size of evaluation sets expressed by its number of sites (in abscissa).  
434 A red line shows the reference value of the indicator, as calculated from the master evaluation  
435 set.

436 All the results exhibited similar patterns. The mean values of the uncertainty indicators were  
437 all close to the reference values. The differences between the quantiles (or the confidence  
438 interval widths) were generally large, which reveals imprecise estimations of the uncertainty  
439 indicators.

440 An increase in precision was observed as the sizes of the evaluation sets increased. For small  
441 sets (fewer than 100 sites), the confidence interval widths were so great that the estimations  
442 of the uncertainty indicators were weakly informative. For the largest sets (500 sites), the  
443 confidence interval widths were much smaller but still conveyed a non-negligible imprecision.  
444 For example, the confidence interval widths for the ME,  $SS_{MSE}$  and PICP were, respectively, 17  
445 g/kg, 0.11 and 5% for the medium precision QRF.

446 Some differences in the results across uncertainty indicators and models had to be noted. The  
447 high precision QRF exhibited less difference between quantiles than did the two other QRFs  
448 for the ME, MSE and  $SS_{MSE}$ . A “better” model would also be a model whose performance can

449 be more easily assessed. However, this was not the case for PICP, which exhibited differences  
450 in quantile values as large as those observed for the two other models.

451

452

#### 453 **4. Discussion**

454

##### 455 4.1. Suitability of hyperspectral data for testing DSM models

456

457 The comparisons of the variograms showed that the pseudo values of the clay content  
458 obtained from hyperspectral data represented the spatial structure of a soil property well,  
459 apart from a smoothing of the small range variability revealed by a decrease in the nugget  
460 value. It must be noted that similar sill and range closeness and nugget decreases were  
461 observed in a previous study (Gomez et al, 2012a). This smoothing is attributable to several  
462 factors that may perturb the spectral signature of the topsoil clay content (Lagacherie et al.,  
463 2008): atmospheric conditions, changes in support (a square block of 5 m side) compared to  
464 soil sampling, or variations in the stoniness, vegetation and rugosity of the soil surface. This  
465 may be also the result of using the partial least square regression that as a linear model,  
466 smooths the variations of the predicted variable. This slight underestimation of the variability  
467 of clay content could lead to slightly underestimated uncertainty indicators. However, this  
468 artefact cannot compromise the results obtained from the variations across the sampling of  
469 these indicators.

470 The spatial pattern of the pseudo values of clay content (as obtained from hyperspectral data)  
471 can be considered a good approximation of a real pattern of soil properties while providing a  
472 quasi-unlimited set of possible sites with soil property measurements. In this paper, we exploit



473 these advantages for experimentally assessing the quality of estimations of the usual  
474 uncertainty indicators of DSM models, which, to the best of our knowledge, has not been done  
475 before. Furthermore, such data make it possible to accurately validate and compare DSM  
476 models thanks to the large size of the validation sets, from which the uncertainty indicators  
477 can be computed.

478 Although it cannot be envisaged that spatial sets of pseudo values of soil properties derived  
479 from airborne hyperspectral imagery could be collected for each DSM application, several  
480 study areas with such soil datasets across the world can be used to enlarge the range of tested  
481 soil properties and pedological contexts (Schwangart and Jarmer, 2011; Stevens et al., 2010;  
482 Ben Dor et al., 2002, Gomez et al., 2012a, Vaudour et al, 2016). Furthermore, other study  
483 areas could be added to this initial set with the aim of building national, regional or global  
484 benchmarks for DSM models, which exist in other disciplines (Rosensweig et al., 2013, Luo et  
485 al., 2012).

486

#### 487 4.2. Uncertainty of the evaluation process

488

489 Our results (figure 4) revealed that the uncertainty indicators can vary across evaluations sets,  
490 from different sample counts, and between evaluation sets of the same number of samples.  
491 This highlights that the uncertainty indicators calculated from statistical validations are  
492 themselves (depending on the evaluation setup) prone to non-negligible uncertainty.

493 It therefore can be claimed from these results that too low of a number of evaluation sites  
494 cannot accurately estimate the performance of DSM models since the values of the  
495 uncertainty indicators may vary a lot with the locations of the evaluation sites. Even with large  
496 numbers of evaluation sites and an unbiased probability sampling, there can still be an

497 imprecision that prevents the models from being ranked in terms of the calculated uncertainty  
498 indicators if the differences in their performances are too small. For example, two models with  
499 differences in  $SS_{MSE}$  below 0.05 may have confidence intervals that overlap each other by more  
500 than half their width. However, it must be noted that the ME, MSE and  $SS_{MSE}^2$  calculated for a  
501 “high quality” model could be less uncertain. The decrease in the uncertainty of predictions  
502 also correspond with a decrease in the spatial variability of the errors, which in turn may  
503 correspond with a decrease of the sensitivity of these uncertainty indicators to the evaluation  
504 dataset. The question then is how does one know for sure that a model is of high quality;  
505 because of the variability of the indicators for low and average quality models, some models  
506 might mistakenly be considered high quality models. This seems particularly true when  
507 considering indicators such as PICP and MSE. This emphasizes the usefulness of estimating  
508 models using a set of complementary indicators.

509

#### 510 4.3. Benefit of the numerical experiment

511 It must be noted that the impact of the sample size on the uncertainty of some uncertainty  
512 indicators (ME and MSE) is already well established in analytical expressions (equations 4, 5  
513 and 6) that have been applied in a few DSM studies (e.g. Kempen et al, 2011). As expected,  
514 our numerical experiment well reproduced the results obtained from the analytical  
515 expressions both for estimating the uncertainty of the ME and MSE (figure 3) and for  
516 reproducing the effect of the sample sizes (Figure 4). Beyond reproducing these results, the  
517 large number of evaluation procedures performed in a same study (27,000) provided a  
518 comprehensive understanding of the respective impacts of the DSM model quality, the  
519 number of evaluation sites and the location of these sites on the values of uncertainty  
520 indicators, including those for which no analytical expression still exist ( $SS_{MSE}$ , PICP). It also

521 revealed that the analytical calculations of the ME and MSE uncertainties were themselves  
522 prone to uncertainty that could be important for the evaluation sets having the largest  
523 variances (Figure 3), which corresponded to those with the smallest sized evaluation sets  
524 (Figure 4).

525 In the future, such a numerical experiment can also be used for obtaining references about  
526 the loss of precision in evaluating the uncertainty of DSM models when it is not possible to  
527 perform a probabilistic sampling and thus to calculate ME and MSE by the analytical  
528 expressions. This occurs in most of the current DSM applications that use legacy soil data  
529 whose locations have been selected by a soil surveyor following a non-probabilistic process  
530 ("free survey").

531

#### 532 4.4. Improving the evaluation process

533

534 Our results clearly showed that the uncertainty of Digital Soil Mapping products cannot be  
535 estimated with a great precision. This must be better taken into account in the practices of  
536 soil mapping evaluations. A first recommendation is to systematically assess the standard  
537 error of the uncertainty indicators using the available analytical formulations when possible  
538 or by bootstrapping the validation set. Furthermore, better attention should be paid to the  
539 sampling techniques used to select the evaluation sites. Indeed, stratified random sampling  
540 using compact geographical strata ensures an even distribution of sites across space but does  
541 not avoid the error on the uncertainty indicators. Reducing this error by using more  
542 sophisticated sampling techniques is a priority. In this perspective, our case study provides a  
543 quasi-infinite number of validation sets that exhibited differences with the master validation  
544 set regarding the values of the uncertainty indicators. Analysing the variability of these

545 differences would permit the sampling criteria to be found that would ensure more accurate  
546 estimations. The new sampling techniques could also be extended to the calibration datasets  
547 of DSM models that provide a priori estimations of their errors or that are used for evaluations  
548 with cross-validation techniques (Brus et al., 2011). In addition, using uncertainty indicators  
549 that are less sensitive to outliers (Nusbaum et al., 2014) would be a complementary way to  
550 reduce the uncertainty of the uncertainty revealed by this paper. Finally, although a  
551 quantitative assessment of uncertainty represents a great progress over the current  
552 evaluation practices of traditional soil surveys, it should be completed by an expert-based  
553 assessment that could check the plausibility of the predicted spatial patterns with regard to  
554 the available pedological knowledge.

555

556

## 557 **5. Conclusion**

558

559 Different evaluation sets obtained by probabilistic sampling were tested for their ability to  
560 assess the prediction uncertainty of DSM models using (as a case study) a spatial pattern of  
561 pseudo-values of topsoil clay content obtained from airborne hyperspectral imagery. The  
562 main lessons are summarized as follows:

- 563 • The spatial patterns of pseudo-values of some soil properties that could be available  
564 in some study areas across the world constitutes a relevant network for experimental  
565 assessments of the uncertainty of validation results. This is because i) it allows the DSM  
566 model to be evaluated by using many sites that could not be envisaged if only real soil  
567 data were used, and ii) it allows different numbers and locations of possible evaluation  
568 sets to be tested. Thus, it may provide an useful complement to the analytical

569 expressions (Brus et al, 2011) for the indicators and the many DSM applications for  
570 which these analytical expressions are not valid.

571 • Any evaluation from independent sets conveys a non-negligible error on the  
572 uncertainty indicators that is greater when the number of sites is low. Such evaluations  
573 should therefore be interpreted with care and the uncertainty on validation results  
574 must be systematically estimated.

575 • The sampling techniques used for the calibration and evaluation datasets should be  
576 improved to reduce this error.

577

## 578 **6. Acknowledgments**

579

580 This research was conducted within the “Centre d’Expertise Scientifique Cartographie  
581 Numérique des sols” granted by the CNES-TOSCA program. We thank David Rossiter and an  
582 anonymous referee for their useful comments for improving the paper.

583

## 584 **7. References**

585 Arrouays, D., McBratney, A.B., Minasny, B., Hempel, J.W., Heuvelink, G.B.M., MacMillan, R.A.,  
586 Hartemink, A.E., Lagacherie, P., McKenzie, N.J., 2014. The GlobalSoilMap project  
587 specifications, in: Arrouays, D., McKenzie, N.J., Hempel, J.W., Richer-de-Forges, A.C.,  
588 McBratney, A.B. (Eds.), GlobalSoilMap: Basis of the global Spatial soil information system.  
589 CRC press & Taylor & francis group, Boca Raton, USA, pp 1-12

590 Ben-Dor E., Patkin K., Banin A., Karnieli A., 2002. Mapping of several soil properties using DAIS-  
591 7915 hyperspectral scanner data – a case study over clayey soils in Israel. *International*  
592 *Journal of Remote Sensing*, 23, p. 1043–1062

593 Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. *Classification and regression trees.*  
594 CRC press.

595 Breiman, L., 2001. Random Forests. *Machine Learning* 45, 5–32.

596 Brus, D.J., Kempen, B., Heuvelink, G.B.M., 2011. Sampling for validation of digital soil maps.  
597 *Eur. J. Soil Sci.* 62, 394–407.

598 Cho, E, Cho, M.J., Eltinge, J., 2004. The variance of sample variance from a finite population.  
599 *Proceedings of Joint American Statistical Association and International Statistical Institute*  
600 *Conference Toronto Canada*

601 Cochran, W.G., Snedecor, G.W., 1989. *Statistical methods.* Eighth edition. Iowa State  
602 University Press, Ames (Iowa).

603 Cressie, N.J.A., 1993. *Statistics for spatial data.* Revised edition. Wiley interscience publication.  
604 New York.

605 Gomez, C., Coulouma, G., Lagacherie, P., 2012a. Regional predictions of eight common soil  
606 properties and their spatial structures from hyperspectral Vis–NIR data, *Geoderma*, 189–  
607 190, 176-185.

608 Gomez, C., Lagacherie P., Bacha, S. 2012b. Using an VNIR/SWIR hyperspectral image to map  
609 topsoil properties over bare soil surfaces in the Cap Bon region (Tunisia). In “*Digital Soil*  
610 *Assessments and Beyond*” Minasny B., Malone B.P., McBratney A.B. (Ed.).Springer, 387-  
611 392.

612 Gomez, C., Oltra Carrio, R., Lagacherie, P., Bacha, S., and Briottet, X. 2015. Sensitivity of soil  
613 property prediction obtained from Hyperspectral Vis-NIR imagery to atmospheric effects  
614 and degradation in image spatial resolutions. *Remote Sensing of Environment* 164, 1-15.

615 Goovaerts, P., 2001. Geostatistical modeling of uncertainty in soil science. *Geoderma* 103, 3–  
616 26.

617 Heuvelink, G.B.M., 2014. Uncertainty quantification of GlobalSoilMap products. in: Arrouays,  
618 D., McKenzie, N.J., Hempel, J.W., Richer-de-Forges, A.C., McBratney, A.B. (Eds.),  
619 GlobalSoilMap: Basis of the global Spatial soil information system. CRC press & Taylor &  
620 Francis group, Boca Raton, USA, pp 327–332.

621 ISSS, ISRIC, FAO, 1998. World Reference Base for Soil Resources. 84 World Soil Resources  
622 report. FAO Rome Italy.

623 IUSS (International Union of Soil Scientists) Working Group WRB , 2006. World Reference  
624 Base for Soil Resources 2006, World Soil Resources Report No. 103. Food and Agriculture  
625 Organization of the United Nations, Rome, Italy.

626

627 Kempen, B., Brus, D.J., Stoorvogel, J.J., 2011. Three-dimensional mapping of soil organic  
628 matter content using soil type–specific depth functions. *Geoderma* 162, 107–123.

629 Lagacherie, P., Baret, F., Feret, J.-B., Madeira Netto, J., Robbez-Masson, J.M., 2008. Estimation  
630 of soil clay and calcium carbonate using laboratory, field and airborne hyperspectral  
631 measurements. *Remote Sensing of Environment* 112, 825–835. DOI  
632 10.1016/j.rse.2007.06.014

633 Lagacherie, P., Gomez, C., in press. Vis-NIR-SWIR Remote Sensing Products as New Soil Data  
634 for Digital Soil Mapping. in pedometrics A.B. McBratney et al. (eds.), Chapter 13. Progress  
635 in Soil Science, Springer. DOI 10.1007/978-3-319-63439-5\_13

636 Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest. R News 2(3), 18--  
637 22.

638 Luo, Y.Q., Randerson, J.T., Abramowitz, G., Bacour, C., Blyth, E., Carvalhais, N., Ciais, P.,  
639 Dalmonech, D., Fisher, J.B., Fisher, R., Friedlingstein, P., Hibbard, K., Hoffman, F.,  
640 Huntzinger, D., Jones, C.D., Koven, C., Lawrence, D., Li, D.J., Mahecha, M., Niu, S.L., Norby,  
641 R., Piao, S.L., Qi, X., Peylin, P., Prentice, I.C., Riley, W., Reichstein, M., Schwalm, C., Wang,  
642 Y.P., Xia, J.Y., Zaehle, S., Zhou, X.H., 2012. A framework for benchmarking land models.  
643 Biogeosciences 9, 3857–3874.

644 McBratney, A.B., Mendonca Santos, M.L., Minasny, B., 2003. On digital soil mapping.  
645 Geoderma 117, 3–52.

646 Meinshausen, N., 2006. Quantile Regression Forests. J. of Machine Learning Res. 7, 983–999.

647 Meinshausen, N., Schiesser, L., 2015. quantregForest: Quantile Regression Forests. R package.  
648 <https://cran.r-project.org>.

649 Mevik, B.-H., Wehrens, R. (2007); The pls Package: Principal Component and Partial Least  
650 Squares Regression in R; Journal of Statistical Software 18(2), 1–24

651 Nussbaum, M., Papritz, A., Baltensweiler, A., Walthert, L., 2014. Estimating soil organic carbon  
652 stocks of Swiss forest soils by robust external-drift kriging. Geosci. Model Dev. 7, 1197–  
653 1210.



654 Nussbaum, M., Spiess, K., Baltensweiler, A., Grob, U., Keller, A., Greiner, L., Schaepman,  
655 M.E., Papritz, A., 2017. Evaluation of digital soil mapping approaches with large sets of  
656 environmental covariates. *SOIL Discuss.* 1–32.

657 Pebesma, E.J., 2004. Multivariable geostatistics in S: the *gstat* package. *Computers &*  
658 *Geosciences*, 30: 683-691.

659 R Development Core Team (2008). R: A language and environment for statistical computing.  
660 R Foundation for Statistical Computing, Vienna, Austria.

661 Rosenzweig, C., Jones, J.W., Hatfield, J.L., Ruane, A.C., Boote, K.J., Thorburn, P., Antle, J.M.,  
662 Nelson, G.C., Porter, C., Janssen, S., Asseng, S., Basso, B., Ewert, F., Wallach, D., Baigorria,  
663 G., Winter, J.M., 2013. The Agricultural Model Intercomparison and Improvement Project  
664 (AgMIP): Protocols and pilot studies. *Agric. For. Meteorol.* 170, 166–182.

665 Savitzky, A., & Golay, M.J.E. (1964). Smoothing and differentiation of data by simplified least  
666 squares procedures. *Analytical Chemistry*, 36(8), 1627–1639

667 Schwanghart, W., Jarmer, T., 2011. Linking spatial patterns of soil organic carbon to  
668 topography — A case study from south-eastern Spain. *Geomorphology* 126, 252–263.

669 Shrestha, D.L., Solomatine, D.P., 2006. Machine learning approaches for estimation of  
670 prediction interval for the model output. *Neural Networks* 19 (2), 225–235.

671 Stevens, A., Udelhoeven, T., Denis, A., Tychon, B., Liou, R., Hoffman, L., Van Wesemael, B.,  
672 2010. Measuring soil organic carbon in croplands at regional scale using airborne imaging  
673 spectroscopy. *Geoderma* 158, 32-45.

674 Tenenhaus, M. (1998). *La régression PLS*. Editions Technip, Paris. 254 pp

675 Vaudour, E., Gilliot, J.M., Bel, L., Lefevre, J., Chehdi, K., 2016. Regional prediction of soil organic  
676 carbon content over temperate croplands using visible near-infrared airborne  
677 hyperspectral imagery and synchronous field spectra. *Int. J. Appl. Earth Observ. Geoinf.*,  
678 49, 24-38. DOI: 10.1016/j.jag.2016.01.005

679 Walvoort, D.J.J., Brus, D.J., de Gruijter, J.J., 2010. An R package for spatial coverage sampling  
680 and random sampling from compact geographical strata by k-means. *Computers and*  
681 *Geosciences*, 36, 1261–1267.

682 Zante, P., Collinet, J., Pepin, Y., 2005. Caractéristiques pédologiques et hydrométéorologiques  
683 du bassin versant de Kamech, Cap Bon, Tunisie. *UMR LISAH IRD Tunis, DG ACTA Direction*  
684 *des Sols Tunis, INRGREF Tunis*. (21 p. + 6 annexes).

685

686