



HAL
open science

Combining measured sites, soilsclapes map and soil sensing for mapping soil properties of a region

Emily Walker, Pascal P. Monestiez, Cécile Gomez, Philippe Lagacherie

► To cite this version:

Emily Walker, Pascal P. Monestiez, Cécile Gomez, Philippe Lagacherie. Combining measured sites, soilsclapes map and soil sensing for mapping soil properties of a region. *Geoderma*, 2017, 300, pp.64-73. 10.1016/j.geoderma.2016.12.011 . hal-01510132

HAL Id: hal-01510132

<https://hal.science/hal-01510132>

Submitted on 26 Sep 2017

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 Combining measured sites, soilscares map and soil
2 sensing for mapping soil properties of a region

3 WALKER^a E., MONESTIEZ^a P., GOMEZ^c C., LAGACHERIE^b P.

4 *a BioSP, INRA, 84000, Avignon, France*

5 *b INRA Laboratoire d'étude des Interactions Sol Agrosystème Hydrosystème (LISAH),*
6 *Campus de la Gaillarde, 2 place Viala, 34060, Montpellier, France*

7 *c IRD Laboratoire d'étude des Interactions Sol Agrosystème Hydrosystème (LISAH),*
8 *Campus de la Gaillarde, 2 place Viala, 34060, Montpellier, France*

9 **Abstract**

10 The limited availability of soil information has been recognized as a main
11 limiting factor in Digital Soil mapping (DSM) studies. It is therefore impor-
12 tant to optimize the joint use of the three sources of soil data that can be
13 used as inputs of DSM models, namely spatial sets of measured sites, soil
14 maps and soil sensing products.

15 In this paper, we propose to combine these three inputs, through a cok-
16 riging with a categorical external drift (CKCED). This new interpolation
17 technique was applied for mapping seven soil properties over a 24.6 km²
18 area located in the vineyard plain of Languedoc (Southern France), using
19 an hyperspectral imagery product as example of a soil sensing data. Cross-
20 validation results of CKCED were compared with those of five spatial and
21 non-spatial techniques using one of these inputs or a combination of two of
22 them.

23 The results obtained in the La Payne Catchment showed i) the utility of
24 soil map and hyperspectral imagery products as auxiliary data for improving
25 soil property predictions ii) the greater added-value of the latter against the
26 former in most situations and iii) the feasibility and the interest of CKCED in
27 a limited number of soil properties and data configurations. Testing CKCED
28 in case study with soil maps of better quality and soil sensing techniques
29 covering more area and depths should be necessary to better evaluate the
30 benefits of this new technique.

31 *Keywords:*

32 Digital Soil Mapping, remote sensing, hyperspectral data, kriging, cross

Preprint submitted to Geoderma

December 8, 2016

34 **1. Introduction**

35 Given the relative lack of, and the huge demand for, quantitative spatial
36 soil information to be used in environmental managing and modelling, digital
37 soil mapping (DSM) has been proposed as an alternative to the classical soil
38 surveys for the quantitative mapping of soil properties over regions at inter-
39 mediate (20-200m) spatial resolutions (McBratney et al., 2003). McBratney
40 et al. (2003) proposed the equation $S = f(s,c,o,r,p,a,n)$ for summarizing the
41 general principle of DSM. According to this equation, a soil property (S) can
42 be predicted by a spatial inference function (f) using, as input, the existing
43 soil information (s), the spatial covariates that map the different factors of
44 soil formation early defined by Jenny (1941) (c,o,r,p,a,) and the geograph-
45 ical location (n) that can highlight any spatial trends missed by the other
46 covariates.

47 It has been early stressed that the limited availability of the soil infor-
48 mation (the s component) was a severe limiting factor in DSM applications
49 (Lagacherie, 2008). Up to now, most of the soil information used as input in
50 DSM for mapping soil properties has been either soil maps or spatial sam-
51 pling of sites with measured soil properties. When available under the form
52 of soil databases (Rossiter, 2004), the former may provide estimates of soil
53 properties over larger areas with however limited spatial resolutions and ac-
54 curacy (Marsman and de Gruijter, 1986, Leenhardt et al., 1995, Odgers et
55 al., 2012). Pedometricians have developed a large range of algorithms for ex-
56 ploiting spatial sampling of sites for mapping soil properties, using sites with
57 measured soil properties combined with spatial covariates (Oliver and Web-
58 ster, 1989). Recent operational applications of DSM are converging toward
59 the use of regression kriging (Malone et al., 2009; Hengl et al., 2015) in which
60 the two sources of soil data are used together, soil map as a soil covariate
61 among others and spatial sampling with measured soil properties as input
62 data for calibration of the regression model and for spatial interpolation of
63 the regression residuals. However, in situations of sparse spatial sampling
64 that often occurs in operational DSM, the performances of the regression
65 kriging remain severely limited (Vaysse and Lagacherie, 2015).

66 The spatial estimations of soil properties produced by Soil Sensing are a
67 third type of soil information that may be considered also as a DSM input

68 that may mitigate the dearth in soil data. A growing number of sensors is
69 now available for producing very high resolution (< 5 m) images of estimated
70 soil properties, either by field-based (or proximal) soil sensing techniques
71 (Adamchuk and Rossel, 2011, Mouazen et al, 2007) or by airborne sensing
72 techniques (Selige, 2006; Stevens et al, 2008; Gomez et al, 2008). However,
73 these soil sensing products are most often available over uncompleted and
74 scattered areas because of their high costs and of their limited conditions of
75 application. This prevents from using them as soil covariates in a classical
76 regression kriging approach. As an alternative for mapping soil properties
77 over a region with soil sensing products, we proposed a co-kriging approach
78 (Lagacherie et al, 2012) that combined such input with a spatial sampling of
79 measured sites. By taking hyperspectral-based estimations of clay content
80 over a limited set of fields with bare surfaces as an example of soil sensing in-
81 put, we showed that soil sensing could bring a significant increase of accuracy
82 of clay content predictions over a whole region.

83 In this paper, we went a step further by developing and testing a new krig-
84 ing approach, namely cokriging with a categorical external drift (CKCED),
85 which combines the three possible soil inputs - soil map, spatial sampling
86 of measured sites and soil sensing products -. This approach was compared
87 with spatial and non-spatial techniques using one of these inputs or a com-
88 bination of two of them. The comparisons were performed for seven soil
89 properties (Clay, silt, sand, Calcium Carbonate, pH, Total Iron and CEC)
90 mapped over a 24.6 km² area located in the vineyard plain of Languedoc
91 (Southern France).

92 **2. Case study**

93 *2.1. Study area*

94 The study was carried out in the La Peyne catchment (Figure 1) in the
95 South of France 43°9'0"N and 3°2'0" E. Vineyards form the primary land
96 use in the area. Marl, limestone and calcareous sandstones from Miocene
97 marine and lacustrine sediments formed the parent material of several soil
98 types observed in this area, including Lithic Leptosols, Calcaric Regosols and
99 Calcaric Cambisols (WRB soil classification, ISSS-ISRIC-FAO, 1998). These
100 sediments were partly covered by successive alluvial deposits ranging from the
101 Pliocene to Holocene and differed in their initial nature and in the duration
102 of weathering conditions. These sediments have produced an intricate soil
103 pattern that includes a large range of soil types, such as Calcaric, Chromic

104 and Eutric Cambisols, Chromic and Eutric Luvisols and Eutric Fluvisols
105 (Coulouma et al 2008). The local transport of colluvial material along the
106 slopes has added to the complexity of the soil patterns. An earlier ground
107 sampling made in the study region (Lagacherie et al., 2008) showed that these
108 complex soil patterns correspond to a great variability of clay content at the
109 soil surface (from 65 g.kg^{-1} to 452 g.kg^{-1}). A study area of 24.6 km^2 (Figure
110 1) was defined by intersecting this region of interest with the hyperspectral
111 image used in this study.

112 2.2. Data

113 2.2.1. Spatial sampling of measured sites

114 143 sites (average sampling density of 1 site / 17 ha) were sampled in the
115 study area for measurements of soil properties. All of these samples were
116 composed of five sub-samples collected to a depth of 5 cm for representing a
117 5 meters x 5 meters square. The geographical position at the centre of this
118 square was recorded by a decimetric GPS instrument. After homogenization
119 of the sample, and removal of plant debris and stones, sieving and air dry-
120 ing, about 20 g was devoted to soil properties laboratory analysis. Seven
121 soil properties for which previous estimations from hyperspectral data were
122 attempted (Gomez et al, 2012a) were determined using classical physico-
123 chemical soil analysis (Baize, 1988): calcium carbonate content (CaCO_3),
124 clay content (granulometric fraction $< 2 \mu\text{m}$), silt content (granulometric
125 fraction between 2 to $50 \mu\text{m}$), sand content (granulometric fraction between
126 0,05 and 2mm), free iron content, cation-exchange capacity (CEC) and pH.

127 Two subsets of sites can be distinguished among the set of 143 sites. 95
128 sampled sites were located in the bare soil fields. Both soil properties mea-
129 surements and hyperspectral data suitable for estimation of soil properties
130 were available for these 95 sites (Figure 1 left). The remaining 48 sites had
131 soil content measurements but unsuitable hyperspectral data because they
132 were located in vineyard fields covered by vegetation. Both subsets were
133 sampled for obtaining an even spatial distribution of sites while respecting
134 the relative importance of the soil mapping units delineated by Coulouma et
135 al (2008). It must be noted that the criteria of selection of the two subsets of
136 sites (bare soil vs vegetated fields) was totally independent from the spatial
137 distribution of soils, which therefore did not generate any sampling bias.

138 *2.2.2. Soil map*

139 The soil map was derived from a very detailed soil map of the study
140 area (Coulouma et al, 2008) by an expert-based grouping of the initial soil
141 units into seven soilscales as homogeneous as possible regarding the topsoil
142 properties focused in this study. These soilscales were described in details
143 in Gomez et al. (2012a). The grouping into soilscales was necessary for
144 obtaining soil mapping units that included a number of sites large enough
145 for applying the tested geostatistical procedures.

146 *2.2.3. Airborne HYMAP image and its derivative*

147 The HYMAP airborne imaging spectrometer measured reflected radiance
148 in 126 non-contiguous bands covering the 400 – 2500 nm spectral range with
149 around 19 nm bandwidths and average sampling intervals of 17 nm in the
150 400 – 2500 nm domain (<http://www.intspec.com/>). The HYMAP image
151 was acquired on 13 July 2003 from a 3000 m altitude, providing a 5 x 5 m
152 spatial resolution. Radiometric calibration was performed in flight (Richter,
153 1996) using nadir ground measurements (Beisl, 2001). The ATCOR4 code
154 for airborne sensors was used for atmospheric corrections (Richter and Schl
155 äpfer, 2000). Topographic corrections were performed with a high-resolution
156 digital elevation model from the Institut Géographique National (www.ign.fr)
157 and DGPS ground control points.

158 The image was masked by using NDVI to remove living vegetation (es-
159 sentially vineyards). The cellulose absorption band (2100 nm) was used to
160 remove dry vegetation. Small areas of bare soils located at the parcel margins
161 or along roads and pathway were also removed since they were not judged as
162 representative of the neighbouring soil surfaces. Finally, the image provided
163 usable data over 33 690 pixels covering 3.5% of the total area only, that is
164 the 192 bare soil fields that were randomly scattered over the region at the
165 date of measurement.

166 **3. Methods**

167 *3.1. Experimental set-up*

168 We present hereafter the general workflow of our testing (Figure 2). The
169 details on methods are presented further.

170 The new algorithm combining the three possible types of soil informa-
171 tion (CKCED) was compared with five non spatial and spatial methods
172 that involved less types of soil information (Figure 2). Ordinary Kriging

173 (OK) and Partial-Least-square-Regression (PLSR) were applied for provid-
174 ing estimations of soil properties (denoted products in figure 2) from the
175 spatial sampling of measured sites and from hyperspectral data respectively.
176 Soil Map and spatial sampling of measured sites were combined twice, first
177 by a baseline method that consists in computing a mean per soil mapping
178 units (SMM), second by a more sophisticated Kriging with Categorical Drift
179 (KCED, Monestiez et al, 2001). Finally the product derived from Hyperspec-
180 tral (PLSR on figure 2) was combined with the spatial sampling of measured
181 sites using a previously developed co-kriging procedure (CK, Lagacherie et
182 al, 2012)

183 3.2. Non spatial methods

184 Two non spatial methods were applied, namely 'soil mapping unit mean'
185 (SMM) and Partial least Square Regression (PLSR). The former is a trivial
186 method for combining a soil map and a spatial sampling of measured sites.
187 The latter is a well-known regression technique that is widely used in imaging
188 spectrometry (Ben-Dor et al, 2008). We provide a brief description of this
189 method and its application on our case study hereafter. More details can be
190 found in Gomez et al, (2012a).

191 Partial Least Square Regression (PLSR)(Tenenhaus, 1998) is a regres-
192 sion method that allows the management of 1) co-linearity between the re-
193 flectance values at different wavelengths and 2) a number of predictors (here
194 wavelengths) that is larger than the number of samples used for calibration
195 (here measured sites). The principle of PLSR is to project the variables in an
196 area of reduced size defined by a set of orthogonal vectors, called latent vari-
197 ables, that maximize the covariance between the descriptive variables (here
198 the reflectance values at different wavelengths) and the dependent variables
199 (here the soil properties).

200 PLSR was applied to estimate the seven topsoil properties from the 126
201 reflectance bands provided by the Hymap image for all pixels covered with
202 hyperspectral data. The PLSRs were calibrated using data from the above-
203 evoked 95 sites located in the bare soil fields and then applied to the bare soil
204 pixels for estimating the soil properties, including the 95 pixels with measured
205 sites. At this stage the spatial dependences between locations were ignored.
206 It must be also noted that this approach can only be applied for bare soil
207 fields with collocated hyperspectral data.

208 *3.3. Spatial methods*

209 The spatial method applied in this study was a bivariate Cokriging with
210 categorical external drift (CKCED). It combines data of soil properties mea-
211 sured on sampling sites (primary variable), hyperspectral data from soil
212 data predicted from hyperspectral imagery with PLSR (the secondary vari-
213 able) and the soilscares map (categorical external drift known everywhere).
214 CKCED was compared with other spatial methods that only use one -Ordinary
215 Kriging (OK)- or two - Kriging with a Categorical External Drift (KCED),
216 cokriging (CK)- inputs. CKCED, KCED and CK are presented hereafter.

217 *3.3.1. Variographic analyses*

218 For each soil property, a linear co-regionalization model (Wackernagel
219 1995) was built for the pair "measured value of soil property" and "PLSR
220 HYMAP estimated value of soil property". A difficulty was to take into
221 account the huge difference between the number of these two data. So the
222 cross-variograms were calculated and fitted on the set of 95 bare-soil field sites
223 at which the two variables were available. The two direct semi-variograms
224 were first modelled as linear combinations of two graphically selected basic
225 structures (spherical 300 m and spherical 2300 m) that were found suitable
226 for all the properties. The same basic structures were then fitted to the
227 cross-semi-variograms under the positive semi-definite constraint (Goovaerts,
228 1997). The fits were checked on simple variograms computed on full hymap
229 dataset (see Figure 3).

230 *3.3.2. Neighbourhood selection*

231 To limit the size of the cokriging system and its unbalanced block struc-
232 ture (33690 vs 95), it was necessary to sample the hymap sites in a neighbour-
233 hood of the kriged site x_0 . To preserve short and longer range effects, and
234 due to patchy structure of hymap data, a trial-and-error approach produced
235 the following trade off: all hymap sites were kept within a distance of 50 m
236 from x_0 (grid lag = 5 m), one over four within a distance of 500 m (grid lag
237 = 10 m) and finally, one over sixteen within a distance of 1500 m (grid lag =
238 20m). The resulting number of selected neighbours was in most case lower
239 than one thousand and at least greater than two hundred. Considering soil
240 sample sites (95), all sites were kept for cokriging in a unique neighbourhood
241 mode (see Figure 4).

242 3.3.3. Statistical modelling for kriging

243 The variable of interest, i.e. one of the above soil properties, is modelled
244 by a random function $Z(x)$ where x denotes the location index (vector of
245 coordinates). $Z(x)$ is decomposed into a deterministic unknown drift $m(x)$
246 and a stationary zero-mean random function $Z_R(x)$ assumed to be Gaussian
247 distributed. In the kriging with external drift approach, $m(x)$ is modelled as
248 a linear function of a deterministic external variable. In the kriging with cate-
249 gorical external drift (KCED) proposed by Monestiez et al. (1999; 2001) and
250 used here, $m(x)$ is modelled as a set of values $e_k, k = 1, \dots, p$, correspond-
251 ing to the five soilscape classes ($p = 5$). The values e_k may be unknown,
252 but the spatial partition of the domain in soilscape classes must be known
253 everywhere. The model can be written as

$$Z(x) = \sum_{k=1}^p \mathbb{1}_{\{k\}}(x) e_k + Z_R(x) \quad (1)$$

254 where e_k is a mean effect for class k to be estimated and $\mathbb{1}_{\{k\}}(x)$ is the
255 indicator function of the class k : it is equal to one if x is in class k , and it
256 is equal to zero otherwise. The variable Z was sampled at n_i sites x_i , for
257 $i = 1, \dots, n_i$. ($n_i = 95$). The second variable $Y(x)$, i.e. the covariate of the
258 bivariate cokriging, denoted further CK, which is here the predicted property
259 by PLSR, is modelled on the same way.

$$Y(x) = \sum_{k=1}^p \mathbb{1}_{\{k\}}(x) e_k + Y_R(x) \quad (2)$$

260 By construction of the PLSR estimates, the mean e_k is the same for Y
261 and Z . The variable Y was sampled at n_j sites x_j , for $j = 1, \dots, n_j$ and
262 where n_j is the number of neighbours selected among the 33690 HYMAP
263 pixels.

264 To simplify notation in the following, the covariance function of Z for a
265 pair of points $C_{ZZ}(x_i - x_{i'})$ is noted $C_{i,i'}^{(ZZ)}$ and the cross-covariance between
266 Z and Y , $C_{ZY}(x_i - x_j)$ is noted $C_{i,j}^{(ZY)}$.

267 Covariances and cross-covariances are directly derived from fitted vari-
268 ograms and co-variograms. Similarly, $Z(x_i)$ and $Y(x_j)$ are respectively noted
269 Z_i and Y_j .

270 3.3.4. *Kriging with external drift*

271 Following Monestiez et al. (1999), the KCED predictor is given by :

$$Z^*(x_0) = \sum_{i=1}^{n_i} \lambda_i Z_i \quad (3)$$

272 where the λ_i 's solve the following kriging system with $n_i + p$ equations to
273 ensure unbiasedness and minimisation of the MSE:

$$\begin{cases} \sum_{i'=1}^{n_i} \lambda_{i'} C_{i,i'}^{(ZZ)} - \sum_{k=1}^p \mu_k \mathbb{1}_{\{k\}}(x_i) = C_{i,0}^{(ZZ)} & \text{for } i = 1, \dots, n_i \\ \sum_{i=1}^{n_i} \lambda_i \mathbb{1}_{\{k\}}(x_i) = \mathbb{1}_{\{k\}}(x_0) & \text{for } k = 1, \dots, p \end{cases} \quad (4)$$

274 3.3.5. *Cokriging*

275 The cokriging CK

$$Z^*(x_0) = \sum_{i=1}^{n_i} \lambda_i Z_i + \sum_{j=1}^{n_j} \lambda'_j Y_j, \quad (5)$$

276 where the λ_i 's and λ'_j 's solve the following cokriging system with $n_i + n_j + 2$
277 equations to ensure unbiasedness and minimisation of the MSE:

$$\begin{cases} \sum_{i'=1}^{n_i} \lambda_{i'} C_{i,i'}^{(ZZ)} + \sum_{j=1}^{n_j} \lambda'_j C_{i,j}^{(ZY)} - \sum_{k=1}^p \mu_k = C_{i,0}^{(ZZ)} & \text{for } i = 1, \dots, n_i \\ \sum_{j'=1}^{n_j} \lambda'_{j'} C_{j,j'}^{(YY)} + \sum_{i=1}^{n_i} \lambda_i C_{i,j}^{(ZY)} - \sum_{k=1}^p \mu_k = C_{j,0}^{(ZY)} & \text{for } j = 1, \dots, n_j \\ \sum_{i=1}^{n_i} \lambda_i = 1 \quad \text{and} \quad \sum_{j=1}^{n_j} \lambda'_j = 0 \end{cases} \quad (6)$$

278 3.3.6. *Cokriging with categorical external drift*

279 The cokriging with categorical external drift (CKCED) predictor is for-
280 mally the same as an Universal Cokriging, and the has the same $Z^*(x_0)$

281 expression where the λ_i 's and λ'_j 's solve the following cokriging system with
 282 $n_i + n_j + p$ equations to ensure unbiasedness and minimisation of the MSE:

$$\left\{ \begin{array}{l} \sum_{i'=1}^{n_i} \lambda_{i'} C_{i,i'}^{(ZZ)} + \sum_{j=1}^{n_j} \lambda'_j C_{i,j}^{(ZY)} - \sum_{k=1}^p \mu_k \mathbb{1}_{\{k\}}(x_i) = C_{i,0}^{(ZZ)} \quad \text{for } i = 1, \dots, n_i \\ \sum_{j'=1}^{n_j} \lambda'_{j'} C_{j,j'}^{(YY)} + \sum_{i=1}^{n_i} \lambda_i C_{i,j}^{(ZY)} - \sum_{k=1}^p \mu_k \mathbb{1}_{\{k\}}(x_j) = C_{j,0}^{(ZY)} \quad \text{for } j = 1, \dots, n_j \\ \sum_{i=1}^{n_i} \lambda_i \mathbb{1}_{\{k\}}(x_i) + \sum_{j=1}^{n_j} \lambda'_j \mathbb{1}_{\{k\}}(x_j) = \mathbb{1}_{\{k\}}(x_0) \quad \text{for } k = 1, \dots, p \end{array} \right. \quad (7)$$

283 Compared to the previous bivariate cokriging system, the constraints on
 284 λ 's and λ' 's are summed up considering Z and Y have same theoretical mean
 285 e_k for each class k . To get a kriging prediction free from class effects e_k , p
 286 constraints are necessary so that the sum of weights for the class to whom x_0
 287 belongs must be one, and the sum of weights in all other classes must be 0.
 288 As a consequence, the unit sum on all λ 's : $\sum_{i=1}^{n_i} \lambda_i + \sum_{j=1}^{n_j} \lambda'_j = 1$ is directly
 289 obtained by summing the p constraints.

290 There are p Lagrange parameters μ_1 to μ_p . Only one term μ , the one
 291 corresponding to the class at x_0 , remains in the kriging variance whose ex-
 292 pression is :

$$\sigma_K^2(x_0) = C_{0,0}^{(ZZ)} - \sum_{i=1}^{n_i} \lambda_i C_{i,0}^{(ZZ)} - \sum_{j=1}^{n_j} \lambda'_j C_{j,0}^{(ZY)} + \sum_{k=1}^p \mu_k \mathbb{1}_{\{k\}}(x_0). \quad (8)$$

293 3.4. Validation

294 To assess the performance of spatial predictions, a leave-one-out cross
 295 validation R_{CV}^2 was calculated. Two distinct data configurations were con-
 296 sidered for the comparisons of these methods, whether the predicted site was
 297 located in a bare soil field with collocated hyperspectral data or not. In the
 298 available data set of measured sites, these two configurations corresponded
 299 to 95 and 48 sites respectively. Because the aim of this paper was to compare
 300 DSM models that used different combinations of input data it was however
 301 preferable to validate each model with the same dataset. Furthermore, be-
 302 cause of the low number of the latter, the specific locations of the sites could
 303 have hampered the comparisons between methods and data configurations.

304 which would have made comparisons less effective. Therefore we tested the
305 methods in the two data configurations from the same set of 95 sites. For
306 these sites we obtained the absence of collocated hyperspectral data by re-
307 moving all hymap data of the bare soil plot to whom belongs the prediction
308 point. We however kept the whole set of sites (143) for testing the Ordinary
309 kriging.

310 4. Results

311 4.1. Co-regionalization models

312 The fitted models are composed of two spherical models for ranges of
313 300 m and 2300 m. The sills for both models were estimated for simple
314 variograms and crossed variograms, as described in the table ??.

315 As shown by the examples of fitted variograms for three representative
316 soil properties (Figure 3) acceptable fits were obtained. As expected, smaller
317 sills were obtained from PLSR HYMAP data than from measured values, the
318 former being unable to capture the whole soil variability. Table ?? exhibited
319 also contrasted 300 m sill / 2300 m sill ratio across soil properties. The
320 largest ones, i.e. the largest proportions of "local" variability, were observed
321 for CaCO₃ and Iron whereas textural properties and CEC had the smallest
322 ones. pH represented an intermediate situation.

323 4.2. Performance of estimation techniques

324 Table 2 shows the performances of the six estimation techniques using
325 various number of soil inputs, for the seven soil properties of interest and for
326 two data configurations, namely collocated HYMAP data vs no collocated
327 HYMAP data but with hymap data in the neighbourhood. All the results
328 are expressed in R² calculated by cross-validation over the subset of 95 sites
329 for which all the estimation techniques can be tested (see section ??).

330 Spatial estimation techniques that combined soil inputs (KCED, CK or
331 CKCED) generally outperformed estimation techniques using a single in-
332 put (OK, PLSR) or non-spatial combination of measured sites with a soil
333 map (SMM). However, in the case of collocated hymap data, the improve-
334 ment was only moderate for iron, which had already good performances with
335 PLSR. Moreover, in the case of no collocated hymap data, combining mea-
336 sured sites and hymap outputs (CK) even produced a decrease in prediction
337 performances for Clay and CEC.

338 Combining measured sites with either the soil map (KCED) or the hymap
339 data (CK) had contrasted interests across soil properties and data configura-
340 tions. In the case of collocated hymap data, CK clearly outperformed KCED
341 whatever the soil properties, with however greater differences for soil prop-
342 erties having already good results with the Hymap data alone (PLSR). In
343 the case of no collocated hymap data, KCED and CK gave similar results
344 for most of the soil properties (iron, silt, sand and pH). However KCED
345 outperformed CK for CEC and Clay whereas CK outperformed KCED for
346 CaCO₃. It must be noted that neither the individual performances of the
347 added inputs (PLSR and SMM, table ??) nor the spatial structures of the
348 soil properties (table ??) could explain these differences.

349 The newly developed estimation technique that combined the three soil
350 inputs (CKCED) provided an improvement for only three properties (Iron,
351 silt and sand) in the case of no-collocated hymap data. In all other cases,
352 the performances of CKCED was similar to those of CK. Here again, it was
353 not possible to relate the differences of results across soil properties with the
354 individual performances of the added inputs and the spatial structures of the
355 soil properties.

356 4.3. Mapping

357 Figure 4 shows images of clay, sand and iron obtained from the cokriging
358 with categorical external drift (CKCED) interpolation . The image of clay
359 showed a global increase of clay content from the north to the south of the
360 area. This is probably the effect of the parent materials, the old (Pliocene)
361 fluvial deposits located in the southern part of the area, being more clayey
362 than any other parent materials. The image of sand showed the converse
363 spatial distribution, apart from the south West of the study area where soils
364 formed on limestone out crops had both low clay and low sand contents. The
365 Iron image exhibited a significantly different soil pattern from the previous
366 ones with two distinct iron-rich areas that corresponded to soil formed on
367 Wurm (North) and Pliocene (south) fluvial deposits. This last image was
368 also the one in which the delineations of the soil map were the most visible.

369 5. Discussion

370 5.1. Case study representativeness

371 Bivariate cokriging and the other interpolation techniques were tested
372 in a Mediterranean area that has been used as a case study for digital soil

373 mapping and remote sensing for a long time (e.g. Leenhardt et al, 1994
374 ; Lagacherie and Voltz, 2000, Lagacherie et al, 2008, Gomez et al, 2012).
375 In spite of its moderate size, it includes a great variety of parent materials
376 and landscape positions that yield complex patterns of soil variations. This
377 was confirmed by the study of variograms of seven soil properties that all
378 exhibited bi-scaled spatial structures and contrasted ratio of short and large-
379 scale variations with properties.

380 In this study, seven soil properties were considered. This allowed ob-
381 serving contrasted situations with regard to the quality of the auxiliary spa-
382 tial data used as input of the interpolation techniques. The proportion of
383 variances captured by the hyperspectral-based estimations of soil properties
384 ranged between $R^2 = 0.20$ for sand to $R^2 = 0.78$ for iron, which corresponds
385 to the range of performances shown in the literature (e.g. Selige et al., 2006;
386 Gomez et al., 2008; Ben-Dor et al, 2008, Stevens et al., 2010). As already
387 observed by Ben Dor et al (2002), the soil properties that corresponded to
388 a chromophore (here Clay, Iron, CEC and Calcium Carbonate) were pre-
389 dicted with more accuracy than the other soil properties (sand, silt and pH).
390 The range of proportions of variances captured by the soil map was smaller
391 ($R^2 < 0.31$). From the soilmap assessments performed in the same pedolog-
392 ical area (Lennhardt et al, 1994; Vaysse and Lagacherie, 2015), this results
393 correspond to a medium to short scale soil map, that cover substantial pro-
394 portions of land, e.g. 39% in Europe (King and Montanarella, 2012) and
395 11% in Africa (Nachtergaele and van Ranst, 2002).

396 In conclusion, the case study can be considered as matching well the
397 level of availability and quality of DSM soil inputs that can be currently
398 encountered nowadays. However, many regions in the world may include
399 hyperspectral data that cover a larger proportion of the study area and more
400 accurate soil maps. For these regions, better and more contrasted results
401 than those presented in this paper could certainly be expected.

402 *5.2. Interest of hyperspectral products as DSM soil input*

403 Up to now, the use in DSM of hyperspectral products that may provide
404 soil property estimations at both high resolutions and large extents has been
405 rarely experimented (Schwangart and Jammer, 2011, Lagacherie et al, 2012,
406 Gomez et al, 2012b,) and have never been compared with the more common
407 use of a soil map as a DSM input combined with measured sites (Mc Bratney
408 et al, 2003, Kempen et al, 2011[1]).

409 The results we obtained showed that hyperspectral products used as an
410 auxiliary input in cokriging generally provided better improvements of soil
411 property predictions than a soil map used as an auxiliary input in Krig-
412 ing with a Categorical external Drift. The only exceptions were for Clay and
413 CEC in locations with no collocated hyperspectral data, for which the combi-
414 nations with the hyperspectral products surprisingly decreased the precisions
415 obtained by simply interpolating the measured sites by Ordinary Kriging.

416 However, the seemingly greater interest of hyperspectral products must
417 be nuanced since we did not have in this case study examples of very well-
418 predicted soil properties by a soil map ($R^2 < 0.31$). Furthermore, one may
419 remember that hyperspectral products can only deliver estimations of surface
420 soil properties because the effective penetration depths of optical sensors do
421 not exceed several millimetres (Liang, 1997[2]), which limits, at best (i.e.
422 cultivated areas), the soil property predictions to the topsoil horizons only.

423 *5.3. Interest of combining three DSM soil inputs*

424 We proposed a cokriging with a categorical external drift that allowed
425 combining the two available auxiliary variables - the soil map and the hyper-
426 spectral estimations of soil properties- with the set of measured sites. This
427 new interpolation technique was found interesting in situations with no collo-
428 cated hyperspectral-based estimations and for a limited number of properties
429 (Table 2). These properties were characterized either by the worst perfor-
430 mances of the soilscapes map (silt and sand) or by the best performances
431 of the hyperspectral based predictions (iron). It must be noted that the
432 amount of local variation of the soil properties (table 1) that was expected
433 to decrease the interest of using non-collocated hyperspectral-based soil esti-
434 mations as auxiliary variable did not explain any difference in performances
435 between soil properties. Here again, we did not explore enough variability
436 of soil map precisions and distances to neighbouring hyperspectral situations
437 for identifying clearly the area of interest of CKCED.

438 *5.4. Future work*

439 The performances of the interpolation techniques tested in this paper
440 could be improved either by better auxiliary spatial variables or by better
441 spatial models.

442 Concerning the former, two ways could be explored. A better accuracy of
443 the soil map can be obtained by increasing its spatial resolution for obtaining
444 a more detailed soil map. However the number of sampled sites can become

445 a limiting factor since KCED requires a good estimate of the mean value
446 of the property within each soil mapping units, which cannot be obtained
447 without a denser spatial sampling of sites than the one used in this study.
448 Beside, since we observed that much better results were obtained within the
449 bare soil area where hyperspectral estimates of soil property were available
450 without interpolation, it would be worth extending this area. This can be
451 straightforwardly done by a better selection of the date of the fly (Gomez et
452 al, 2012b). Furthermore, the remaining vegetated area can be processed with
453 spectral unmixing (Bartholomeus et al, 2010) or source separation algorithms
454 (Ouerghemmi et al, 2016) for filtering the vegetation signal that may perturb
455 the estimations of soil properties. Finally, other soil sensing techniques than
456 hyperspectral imagery can be used as soil input to enlarge both the area and
457 the exploration depth of the targeted soil properties.

458 The spatial models underlying the interpolations could be improved first
459 by taking into account additional soil covariables like e.g. Digital Eleva-
460 tion Model and its derivatives e.g. slope, aspect, curvature, that have been
461 largely used in Digital Soil Mapping (McBratney et al, 2003). Another way
462 of improvement is to take into account the non stationarity of soil prop-
463 erty variations by applying interpolations based on local (Sun, 2012) and/or
464 anisotropic spatial models (Schwangart and Jammer, 2011).

465 6. Conclusion

466 This study tested the use of the three possible soil inputs for DSM models
467 – spatial set of measured sites, soil map and soil sensing products. A new
468 spatial interpolation technique – cokriging with a categorical external drift –
469 was developed for combining these three inputs. The results obtained in the
470 La Payne Catchment demonstrated the utility of auxiliary variables such as
471 soil map or hyperspectral imagery products for predicting soil properties and
472 the greater added-value of the latter against the former in most situations.

473 The combination of soilmap and hyperspectral-based estimations of soil
474 property allowed by the novel cokriging with categorical external drift pro-
475 cedure (CKCED) brought improvements for a limited number of soil proper-
476 ties and data configurations. However, to better evaluate its utility, this new
477 combination needs to be tested in other case study with soil maps of better
478 quality and soil sensing techniques covering more area and depths.

479 7. Acknowledgements

480 This research was granted by INRA, IRD and the French National re-
481 search agency (ANR) (ANR-08-BLAN-0284-01). We are indebted to Dr.
482 Steven M. de Jong, Utrecht University in The Netherlands and to Dr. An-
483 dreas Mueller of the German Aerospace Establishment (DLR) in Wessling,
484 Germany for providing the 2003 HyMap images for this study. We warmly
485 thank the two anonymous reviewers for their constructive and useful com-
486 ments.

487 8. References

- 488 Adamchuk, V.I., Viscarra Rossel, R.A. 2010. Development of on-the-go
489 Proximal Sensing
490 Systems. In: Proximal Soil Sensing , eds: R.A. Viscarra.Rossel, A.B.,
491 McBratney, B. Minasny, pp 15–28. Progress in Soil Science 1, Springer
492 Dordrecht, Heidelberg, London New York.
- 493 Bartholomeus, H., Kooistra, L., Stevens, A., Van Leeuwen, M., Van Wese-
494 mael, B., Ben-Dor, E. 2011. Soil organic carbon mapping of partially
495 vegetated agricultural fields with imaging spectroscopy. *International*
496 *Journal of Applied Earth Observation and Geoinformation*, 13, 81–88.
- 497 Beisl U. 2001. Correction of bidirectional effects in imaging spectrometer
498 data. Zurich University, Zurich (Switzerland).
- 499 Ben-Dor E., Patkin K., Banin A. and Karnieli A. (2002). Mapping of sev-
500 eral soil properties using DAIS-7915 hyperspectral scanner data-a case
501 study over clayey soils in Israel. *International Journal of Remote Sens-*
502 *ing*, 23 (6), p. 1043-1062.
- 503 Ben-Dor, E., Taylor, R.G., Hill, J., Dematte, J.A.M., Whiting, M.L., Chabril-
504 lat, S. 2008. Imaging spectrometry for soil applications. *Advances in*
505 *Agronomy*, 97, 321–392.
- 506 Ciampalini, R., Lagacherie, P., Monestiez, P., Walker, E., Gomez, C. 2012.
507 Cokriging of soil properties with VisNIR hyperspectral covariates in the
508 Cap Bon region (Tunisia). In: Minasny, B., Malone, B., McBratney,
509 A.B. (Eds.), *Digital Soil Assessments and Beyond* (). CRC Press. pp.
510 393–398

- 511 Coulouma, G., Barthes, J. P., Robbez-Masson, J. M. 2008. Carte des sols
512 de la Basse Vallée de la Peyne. Report and map UMR LISAH (INRA).
- 513 Gomez, C., Lagacherie, P., Coulouma, G. 2008. Continuum removal versus
514 PLSR method for clay and calcium carbonate content estimation from
515 laboratory and airborne hyperspectral measurements. *Geoderma*, 148
516 (2), 141–148.
- 517 Gomez, C., Lagacherie, P., Coulouma, G., 2012a. Regional predictions of
518 eight common soil properties and their spatial structures from hyper-
519 spectral Vis-NIR data. *Geoderma* 189-190:176–185.
- 520 Gomez, C., Lagacherie, P., Bacha, S., 2012b. Using Vis-NIR hyperspectral
521 data to map topsoil properties over bare soils in the Cap-Bon Region,
522 Tunisia, in: Minasny, B., Malone, B., McBratney, A.B. (Eds.), *Digital
523 Soil Assessment and Beyond*. CRC Press, pp. 387–392.
- 524 Goovaerts, P. 1997. *Geostatistics for Natural Resources Evaluation*. Oxford
525 University Press.
- 526 Grunwald, S. 2009. Multi-criteria characterization of recent digital soil
527 mapping and modeling approaches. *Geoderma* 152:195–207.
- 528 Henderson, B. L., Bui, E. N., Moran, C. J., Simon, D. A. P. 2005. Australia-
529 wide predictions of soil properties using decision trees. *Geoderma*,
530 124(3–4), 383–398.
- 531 Hengl, T., de Jesus, J.M., MacMillan, R.A., Batjes, N.H., Heuvelink, G.B.M.,
532 Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J.G.B., Walsh,
533 M.G., Gonzalez, M.R., 2014. SoilGrids1km \tilde{N} Global Soil Information
534 Based on Automated Mapping. *PLoS One* 9, e105992.
- 535 Jenny, H. 1941. *Factors of soil formation* (p. 281). New York, NY: McGraw-
536 Hill Book Company.
- 537 Kempen, B., Brus, D.J., Stoorvogel, J.J., 2011. Three-dimensional mapping
538 of soil organic matter content using soil type-specific depth functions.
539 *Geoderma* 162(1-2), 107-123.
- 540 King D. and Montanarella L. 2002. Inventaire et surveillance des sols en
541 Europe. *Etude et Gestion des Sols*, 9: 137–148.

- 542 Lagacherie, P., Voltz, M. 2000. Predicting soil properties over a region using
543 sample information from a mapped reference area and digital elevation
544 data: a conditional probability approach. *Geoderma*, 187–208.
- 545 Lagacherie, P. 2008. Digital soil mapping: a state of the art. In A. E.
546 Hartemink, A. B. McBratney, M. L. Mendonca Santos (Eds.), *Digital
547 soil mapping with limited data* (pp. 3–14). Springer science.
- 548 Lagacherie, P., Baret, F., Feret, J-B, Madeira Netto, J., Robbez-Masson,
549 J.M. 2008. Estimation of soil clay and calcium carbonate using labora-
550 tory, field and airborne hyperspectral measurements. *Remote Sensing
551 of Environment*, 112 (3), 825–835.
- 552 Lagacherie, P., Bailly, J.S., Monestiez, P., Gomez, C. 2012. Using scattered
553 hyperspectral imagery data to map the soil properties of a region. *Eur.J.
554 Soil Science*; 63:110–119.
- 555 Leenhardt, D., Voltz, M., Bornand, M., Webster, R. 1994. Evaluating soil
556 maps for prediction of soil water properties. *European Journal of Soil
557 Science*, 45(3), 293–301.
- 558 Liang, S., 1997. An investigation of remotely-sensed soil depth in the optical
559 region. *International Journal of Remote Sensing* 18, 3395–3408
- 560 Malone, B.P., McBratney, A.B., Minasny, B., 2011. Empirical estimates of
561 uncertainty for mapping continuous depth functions of soil attributes.
562 *Geoderma* 160, 614–626.
- 563 Marsman, B. A., Gruijter, J. J. de. 1986. Quality of soil maps: a comparison
564 of survey methods in a sandy area. *Soil Survey Papers*, Netherland Soil
565 Survey Institute, Wageningen.
- 566 McBratney, A. B., Mendonca Santos, M. L., Minasny, B. 2003. On digital
567 soil mapping. *Geoderma*, 117(1–2), 3–52.
- 568 Monestiez, P., Allard, D., Navarro Sanchez, I., Courault, D. 1999. Kriging
569 with categorical external drift: Use of thematic maps in spatial pre-
570 diction and application to local climate interpolation for agriculture,
571 in *geoENV II: Geostatistics for Environmental Applications*, Gomez-
572 Hernandez J., Soares A. and Froidevaux R. Eds, Kluwer Academic
573 Publishers, Dordrecht, 163–174.

- 574 Monestiez, P., Courault, D., Allard, D., Ruget, F., 2001. Spatial interpo-
575 lation of air temperature using environmental context: application to
576 crop model. *Environmental and Ecological Statistics*. 8: 297–309.
- 577 Mouazen, A. M., Maleki, M. R., De Baerdemaeker, J., Ramon, H. 2007.
578 Online measurement of some selected soil properties using a VIS–NIR
579 sensor. *Soil and Tillage Research*, 93(1), 13–27.
- 580 Mulder, V.L., De Bruin, S., Schaepman, M.E., and Mayr, T.R., 2011. The
581 use of remote sensing in soil and terrain mapping – A review. *Geoderma*
582 162, 1–19.
- 583 Nachtergaele, F. O., Van Ranst, E. 2002. Qualitative and Quantitative
584 Aspects of Soil Databases in Tropical Countries. In G. Stoops (Ed.),
585 *Evolution of Tropical Soil Science: Past and Future* (pp. 107–126).
586 Brussel: Koninklijke Academie voor Overzee Wetenschappen.
- 587 Odgers, N. P., Libohova, Z., Thompson, J. A. 2012. Equal–area spline
588 functions applied to a legacy soil database to create weighted-means
589 maps of soil organic carbon at a continental scale. *Geoderma*, 189,
590 153–163.
- 591 Oliver, M. A., Webster, R. 1989. A geostatistical basis for spatial weighting
592 in multivariate classification. *Mathematical Geology*, 21(1), 15–35.
- 593 Ouerghemmi, W., Gomez, C., Naceur, S., Lagacherie, P., 2016. Semi-blind
594 source separation for the estimation of the clay content over semi-
595 vegetated areas using VNIR/SWIR hyperspectral airborne data. *Re-
596 mote Sens. Environ.* 181, 251–263.
- 597 Richter, R. 1996. Atmospheric correction of DAIS hyperspectral image data.
598 *Computers and Geosciences* 22, 785–793.
- 599 Richter, R., Schlapfer, D.A., 2000. A unified approach to parametric geocod-
600 ing and atmospheric/topographic correction for wide FOV airborne
601 imagery. Part 2: atmospheric Correction. Proc. 2nd Intl. EARSeL
602 Workshop on Imaging Spectroscopy, Enschede, July 11–13, 2000.
- 603 Rossiter, D. G. 2004. Digital soil resource inventories: status and prospects.
604 *Soil use and management*, 20(3), 296–301.

- 605 Schwanghart, W., Jarmer, T. 2011. Linking spatial patterns of soil organic
606 carbon to topography – a case study from south-eastern Spain. *Geo-*
607 *morphology*, 126, 252-263. DOI: 10.1016/j.geomorph.2010.11.008.
- 608 Selige, T., Bohner, J., Schmidhalter, U. 2006. High resolution topsoil map-
609 ping using hyperspectral image and field data in multivariate regression
610 modeling procedures. *Geoderma*, 136, no1-2, pp. 235-244.
- 611 Stevens, A., Udelhoven, T., Denis, A., Tychon, B., Liroy, R., Hoffmann,
612 L., Wesemael, B. 2010. Measuring soil organic carbon in croplands
613 at regional scale using airborne imaging spectroscopy, *Geoderma*, 158,
614 1-2.
- 615 Sun, W., Minasny, B., McBratney, A. 2012. Analysis and prediction of soil
616 properties using local regression kriging. *Geoderma*, 171-172, 16-23.
617 doi:10.1016/j.geoderma.2011.02.010.
- 618 Tenenhaus M. 1998. *La regression PLS. Theorie et Pratique*. Editions T,
619 Paris.
- 620 Vaysse, K., Lagacherie, P. 2015. Evaluating Digital Soil Mapping ap-
621 proaches for mapping GlobalSoilMap soil properties from legacy data
622 in Languedoc Roussillon (France). *Geoderma Regional*, 4, 20-30.
623 doi:10.1016/j.geodrs.2014.11.003.
- 624 Wackernagel, H., 1995. *Multivariate geostatistics*. Springer Verlag Editions.
625 255 pp.

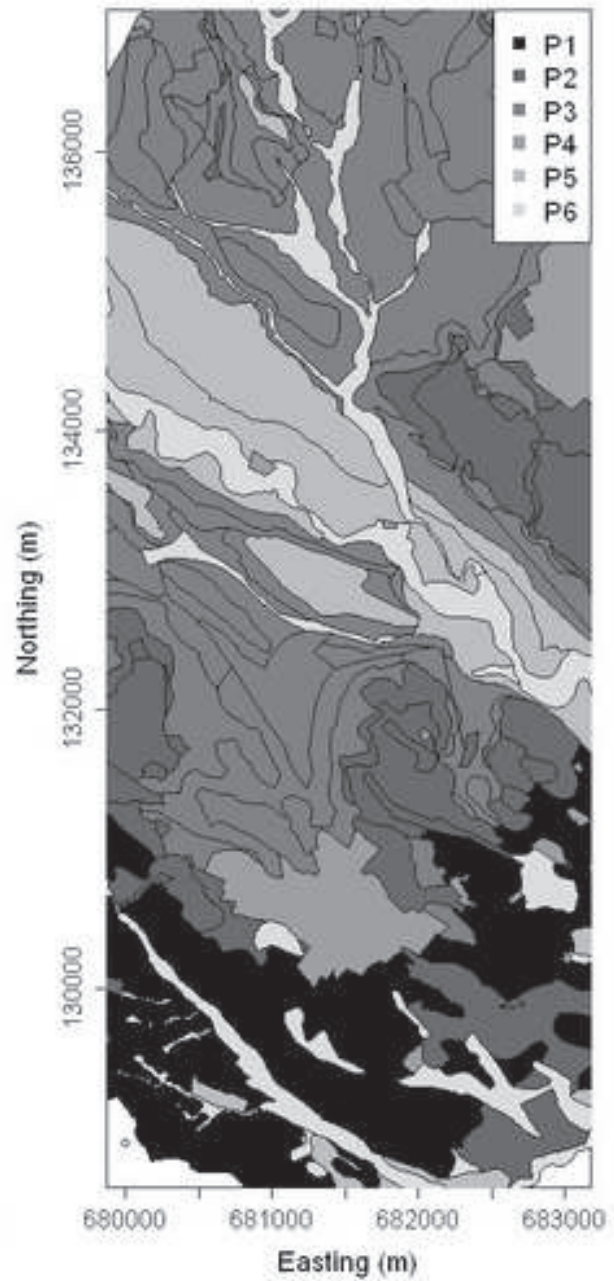
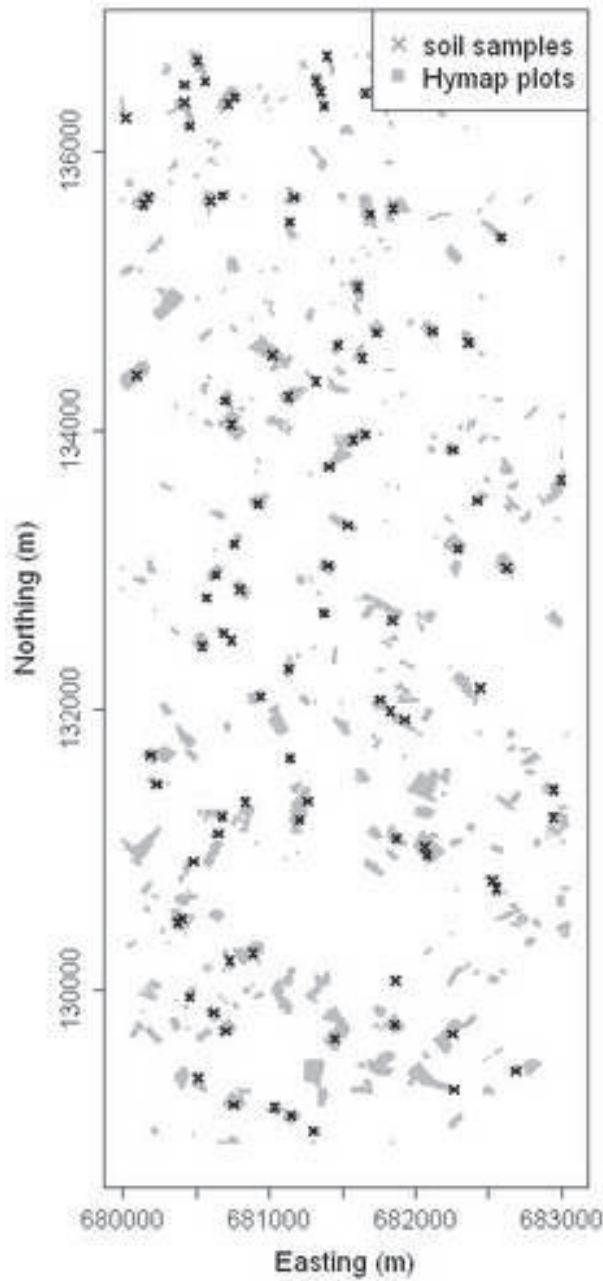
Table 1: Fitted sill and range parameters of direct (samples and hymap) and cross variograms. *in g^2/kg^2 for clay, $CaCO_3$, Iron, Sand and Silt; no unit for pH; $Meq^2/100g^2$ for CEC

Soil property	range (m)	samples sill*	crossed sill*	Hymap sill*
Clay	300	3578	1886	1600
	2300	1387	1691	2062
$CaCO_3$	300	7522	4819	4658
	2300	13412	12871	12352
CEC	300	6.94	4.59	4.51
	2300	1.79	1.56	2.03
Iron	300	0.169	0.129	0.141
	2300	0.314	0.274	0.245
pH	300	0.338	0.028	0.023
	2300	0.366	0.178	0.096
Sand	300	11146	1516	1270
	2300	3715	2969	2373
Silt	300	7910	1586	1081
	2300	249	504	1021

Table 2: Performances (cross validation R2) of the different methods for two data configurations: with collocated hymap data (Config. 1) and with no collocated hymap data **but with hymap data in the neighbourhood** (Config. 2). OK: Ordinary Kriging, PLSR: Partial least square Regression, SMM : mean per Soil mapping unit, KCED: Kriging with categorical external drift, CKCED: Cokriging with categorical external drift. *insensitive to data configuration (results are repeated for enabling comparisons). ** "-" means "not feasible with this data configuration"

Number of soil input	one		two			three
	OK*	PLSR	SMM*	KCED*	CK	CKCED
Config. 1						
Iron	0.45	0.78	0.31	0.48	0.80	0.79
CaCO ₃	0.45	0.76	0.20	0.46	0.84	0.84
CEC	0.30	0.62	0.23	0.36	0.71	0.71
Clay	0.29	0.67	0.26	0.35	0.71	0.70
Silt	0.26	0.17	0.07	0.30	0.37	0.37
Sand	0.12	0.20	0.02	0.18	0.35	0.35
pH	0.20	0.31	0.16	0.26	0.37	0.36
Config. 2						
Iron	0.45	- **	0.31	0.48	0.46	0.49
CaCO ₃	0.45	-	0.20	0.46	0.55	0.55
CEC	0.30	-	0.23	0.36	0.08	0.10
Clay	0.29	-	0.26	0.35	0.12	0.14
Silt	0.26	-	0.07	0.30	0.29	0.34
Sand	0.12	-	0.02	0.18	0.17	0.19
pH	0.20	-	0.16	0.26	0.26	0.26

Figure 1
[Click here to download high resolution image](#)



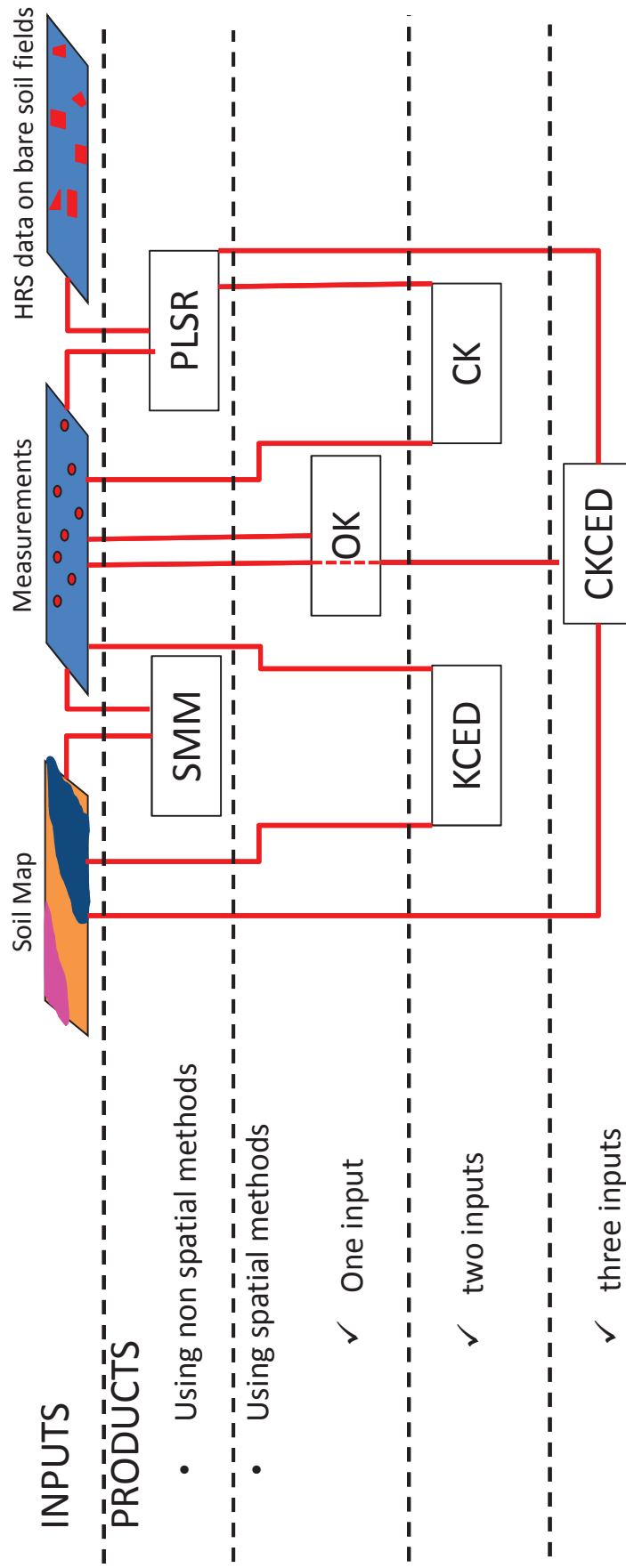
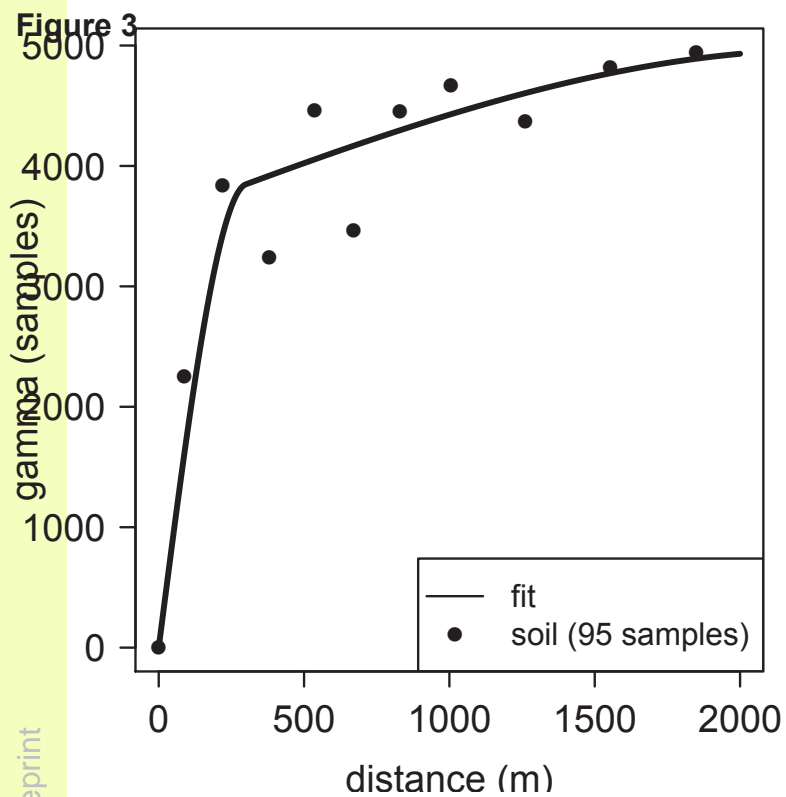
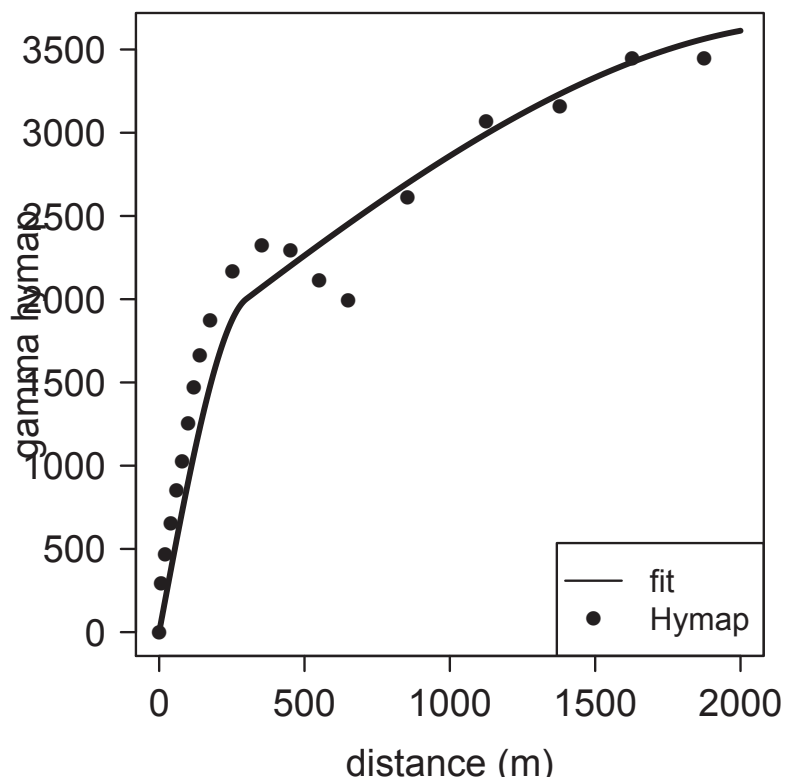
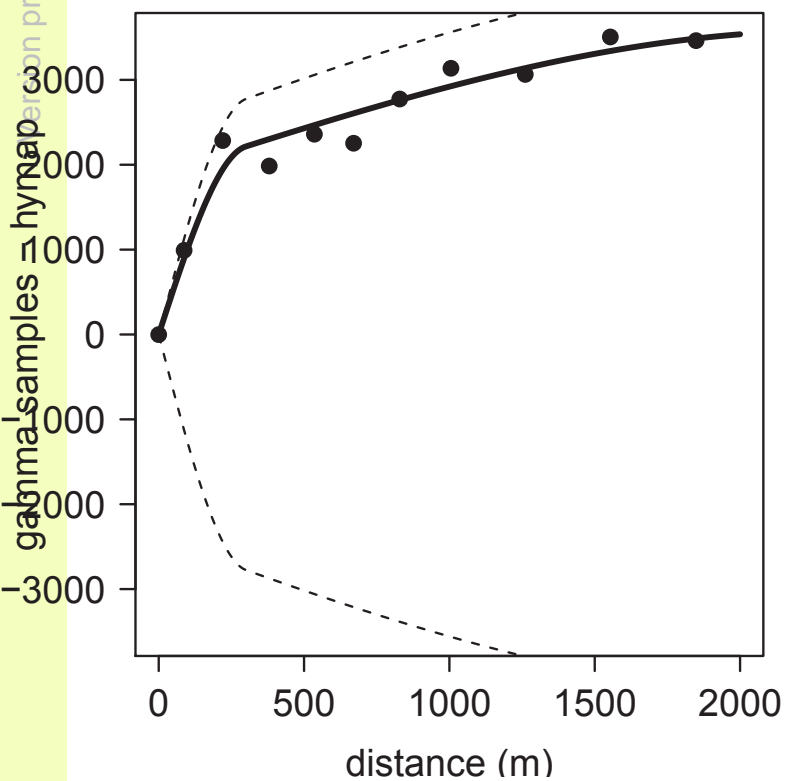


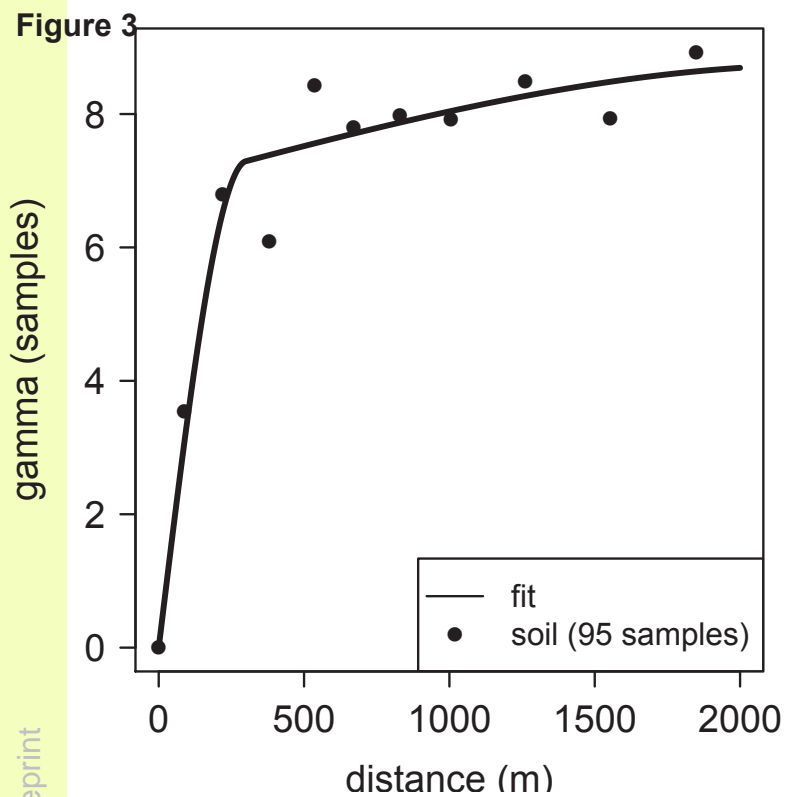
Figure 2

Comment citer ce document : Walker, E., Monestiez, P., Gomez, C., Lagacherie, P. (2017). Combining measured sites, soilscares map and soil sensing for mapping soil properties of a region. Geoderma, 300, 64-73. DOI : 10.1016/j.geoderma.2016.12.011

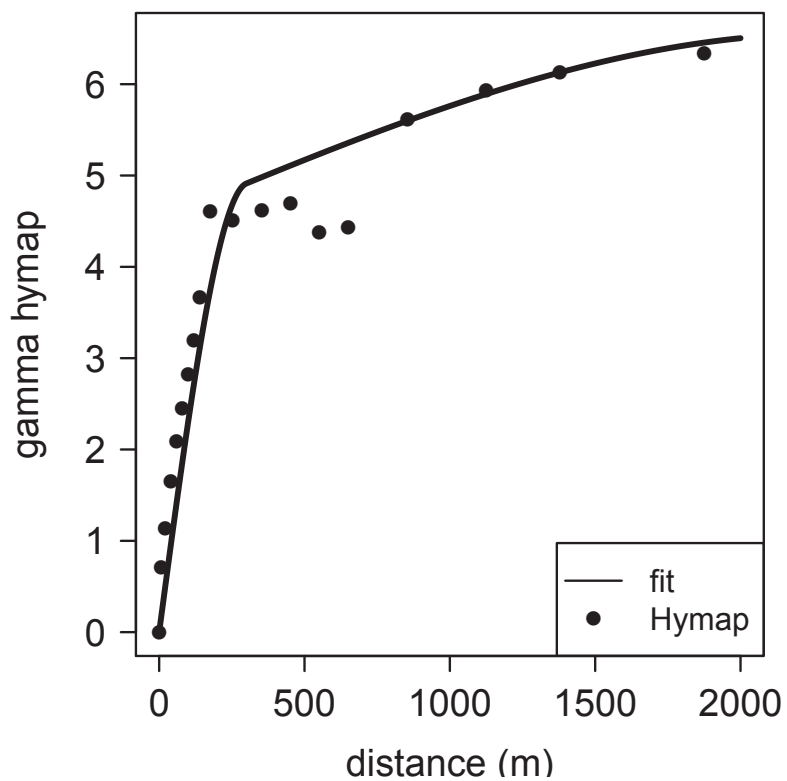
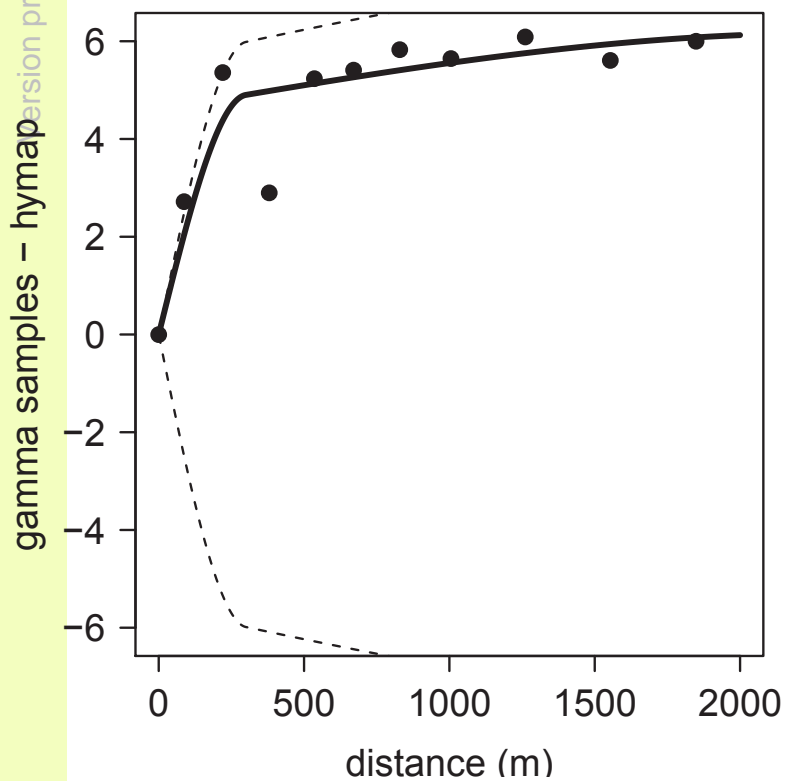


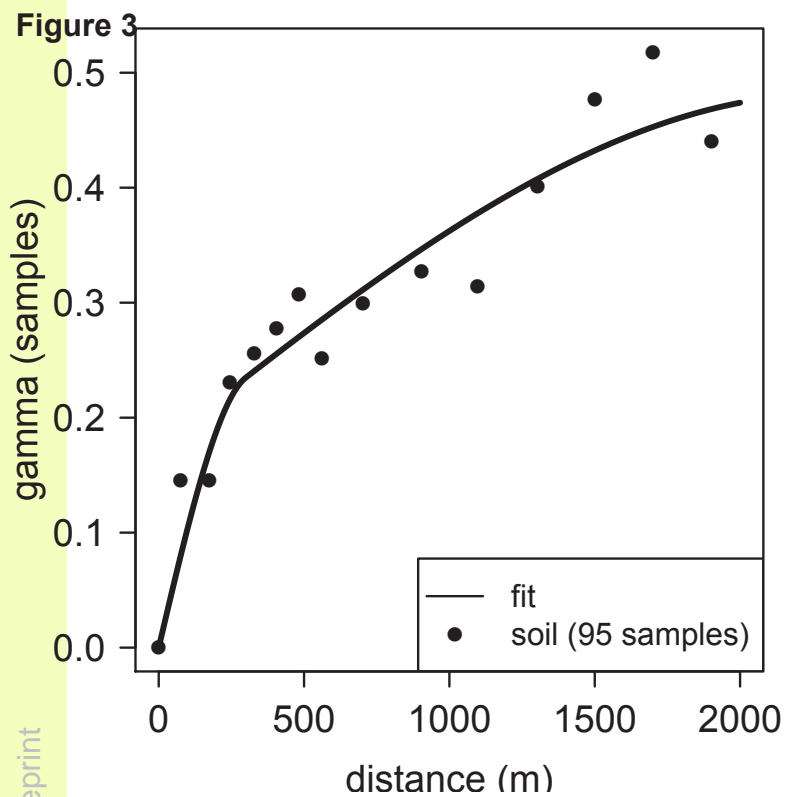
Clay





CEC





Iron

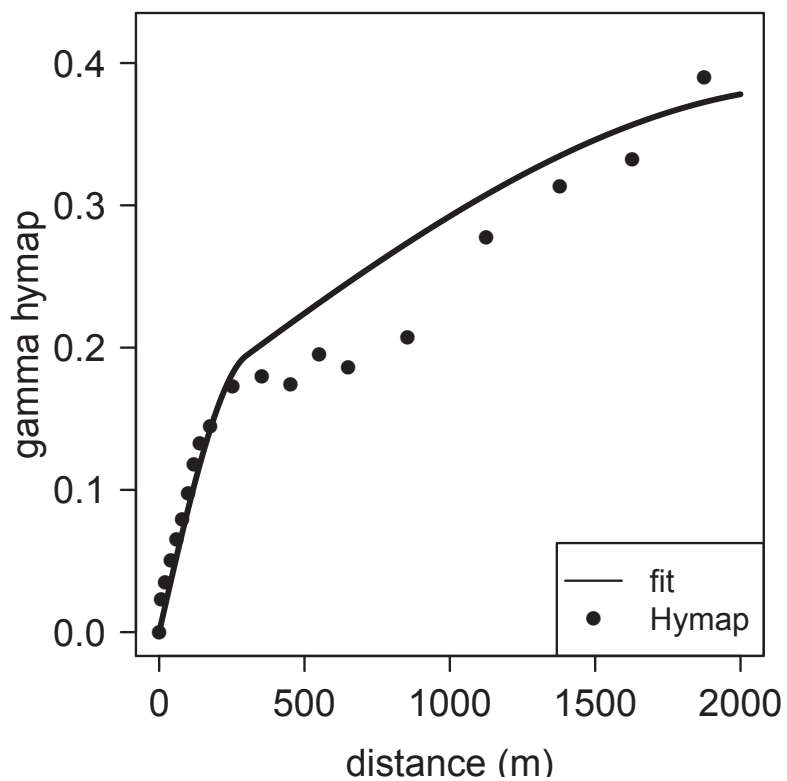
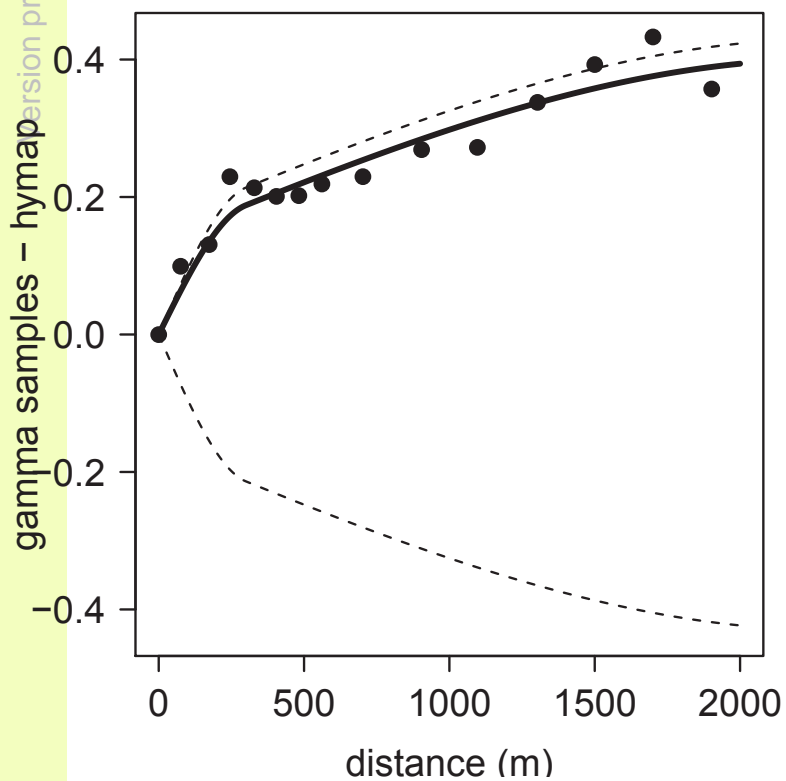
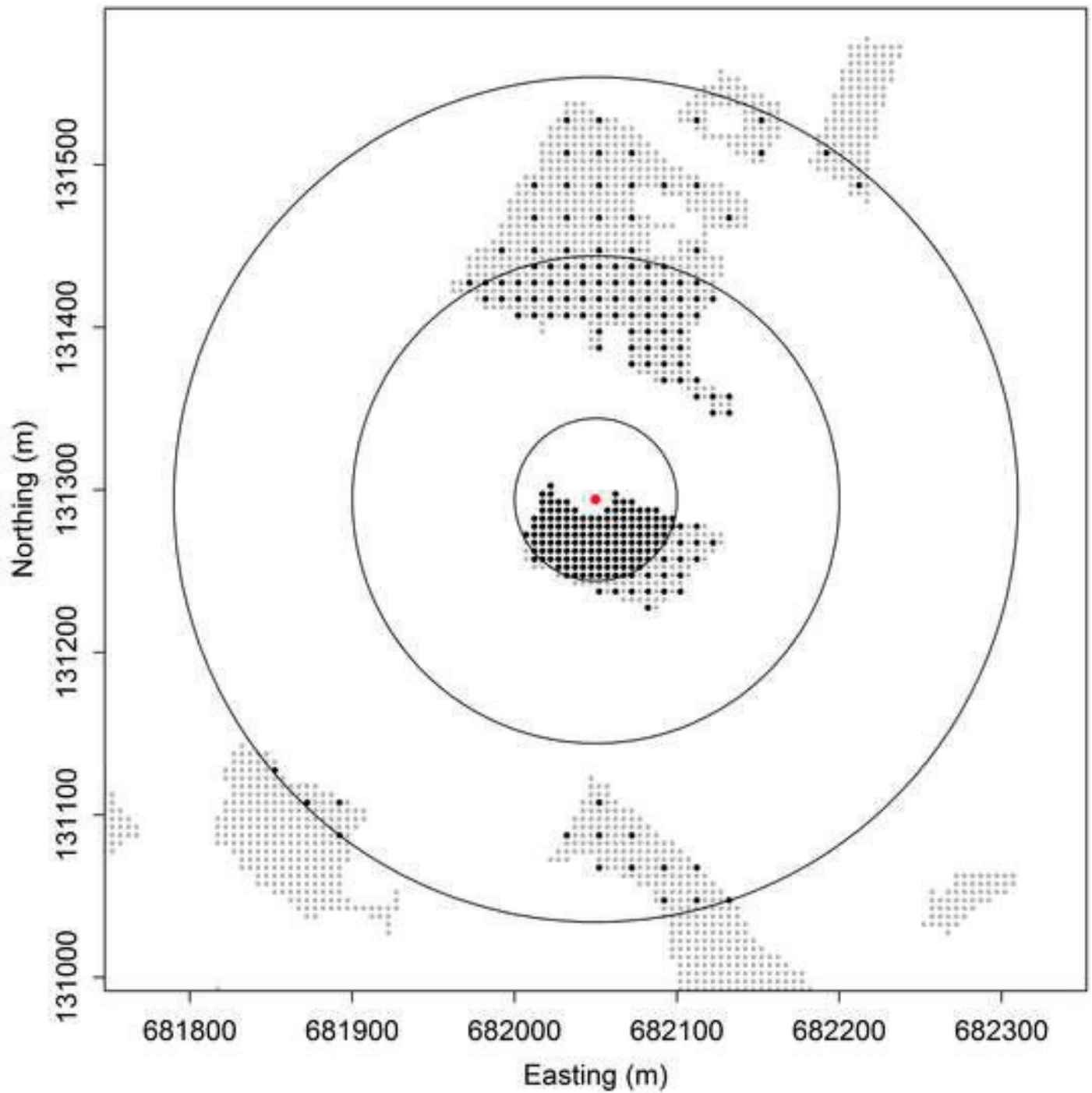


Figure 4
[Click here to download high resolution image](#)



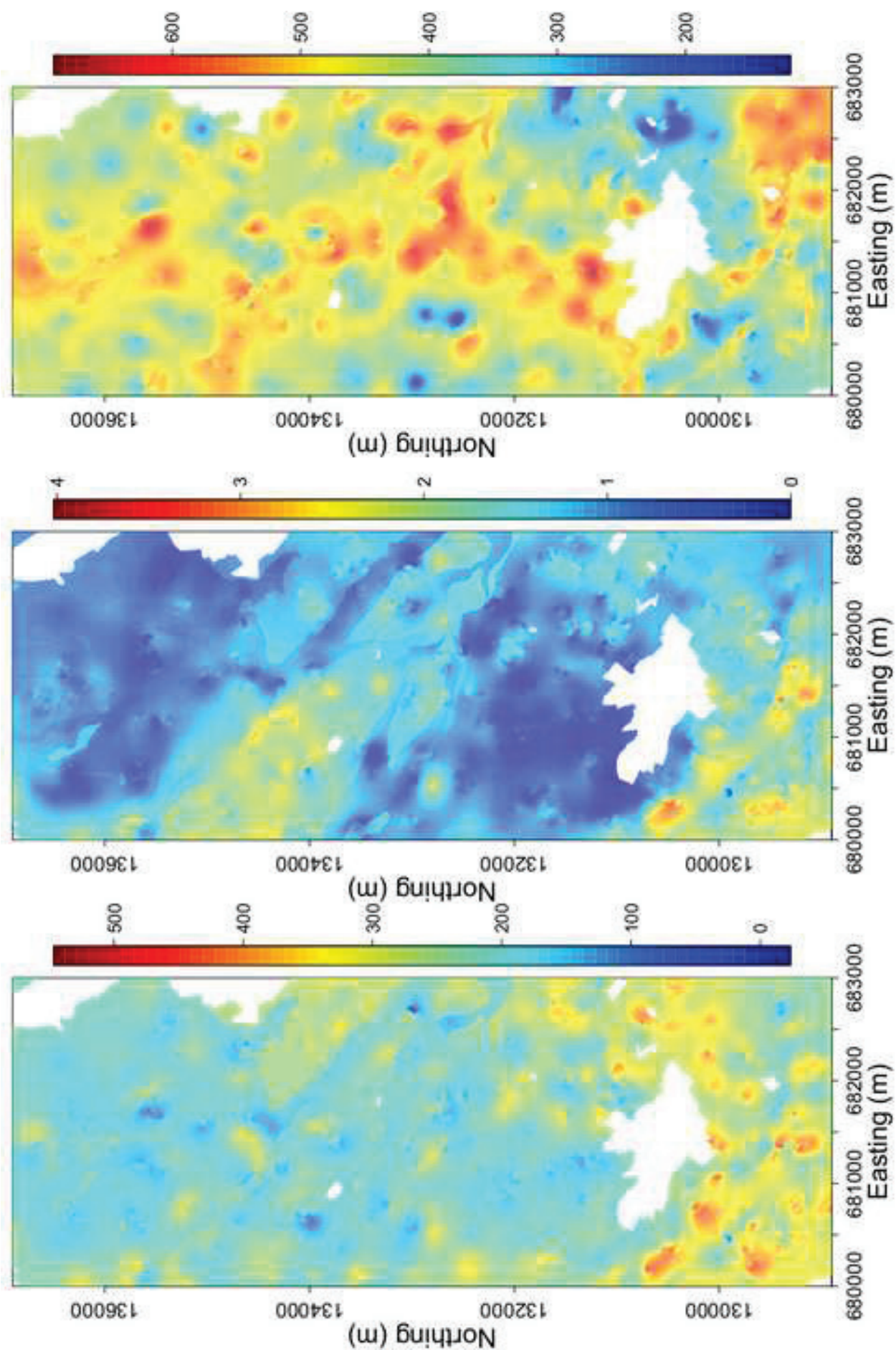


Figure 5
Click here to download high resolution image