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

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

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

WALKER[a] E., *MONESTIEZ*[a] P., *GOMEZ*[c] C., *LAGACHERIE*[b] P.

a BioSP, INRA, 84000, Avignon, France
b INRA Laboratoire d'étude des Interactions Sol Agrosystème Hydrosystème (LISAH), Campus de la Gaillarde, 2 place Viala, 34060, Montpellier, France
c IRD Laboratoire d'étude des Interactions Sol Agrosystème Hydrosystème (LISAH), Campus de la Gaillarde, 2 place Viala, 34060, Montpellier, France

---

Abstract

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

In this paper, we propose to combine these three inputs, through a cokriging with a categorical external drift (CKCED). This new interpolation technique was applied for mapping seven soil properties over a 24.6 km^2 area located in the vineyard plain of Languedoc (Southern France), using an hyperspectral imagery product as example of a soil sensing data. Cross-validation results of CKCED were compared with those of five spatial and non-spatial techniques using one of these inputs or a combination of two of them.

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

Keywords:
Digital Soil Mapping, remote sensing, hyperspectral data, kriging, cross

validation, soil map, soil properties

1. Introduction

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

It has been early stressed that the limited availability of the soil information (the s component) was a severe limiting factor in DSM applications (Lagacherie, 2008). Up to now, most of the soil information used as input in DSM for mapping soil properties has been either soil maps or spatial sampling of sites with measured soil properties. When available under the form of soil databases (Rossiter, 2004), the former may provide estimates of soil properties over larger areas with however limited spatial resolutions and accuracy (Marsman and de Gruijter, 1986, Leenhardt et al., 1995, Odgers et al., 2012). Pedometricians have developed a large range of algorithms for exploiting spatial sampling of sites for mapping soil properties, using sites with measured soil properties combined with spatial covariates (Oliver and Webster, 1989). Recent operational applications of DSM are converging toward the use of regression kriging (Malone et al., 2009; Hengl et al., 2015) in which the two sources of soil data are used together, soil map as a soil covariate among others and spatial sampling with measured soil properties as input data for calibration of the regression model and for spatial interpolation of the regression residuals. However, in situations of sparse spatial sampling that often occurs in operational DSM, the performances of the regression kriging remain severely limited (Vaysse and Lagacherie, 2015).

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

In this paper, we went a step further by developing and testing a new kriging approach, namely cokriging with a categorical external drift (CKCED), which combines the three possible soil inputs - soil map, spatial sampling of measured sites and soil sensing products -. This approach was compared with spatial and non-spatial techniques using one of these inputs or a combination of two of them. The comparisons were performed for seven soil properties (Clay, silt, sand, Calcium Carbonate, pH, Total Iron and CEC) mapped over a 24.6 km^2 area located in the vineyard plain of Languedoc (Southern France).

2. Case study

2.1. Study area

The study was carried out in the La Peyne catchment (Figure 1) in the South of France $43^o9'0''N$ and $3^o2'0''$ E. Vineyards form the primary land use in the area. Marl, limestone and calcareous sandstones from Miocene marine and lacustrine sediments formed the parent material of several soil types observed in this area, including Lithic Leptosols, Calcaric Regosols and Calcaric Cambisols (WRB soil classification, ISSS-ISRIC-FAO, 1998). These sediments were partly covered by successive alluvial deposits ranging from the Pliocene to Holocene and differed in their initial nature and in the duration of weathering conditions. These sediments have produced an intricate soil pattern that includes a large range of soil types, such as Calcaric, Chromic and Eutric Cambisols, Chromic and Eutric Luvisols and Eutric Fluvisols (Coulouma et al 2008). The local transport of colluvial material along the slopes has added to the complexity of the soil patterns. An earlier ground sampling made in the study region (Lagacherie et al., 2008) showed that these complex soil patterns correspond to a great variability of clay content at the soil surface (from 65 $g.kg^{-1}$ to 452 $g.kg^{-1}$). A study area of 24.6km^2 (Figure 1) was defined by intersecting this region of interest with the hyperspectral image used in this study.

2.2. Data

2.2.1. Spatial sampling of measured sites

143 sites (average sampling density of 1 site / 17 ha) were sampled in the study area for measurements of soil properties. All of these samples were composed of five sub-samples collected to a depth of 5 cm for representing a 5 meters x 5 meters square. The geographical position at the centre of this square was recorded by a decimetric GPS instrument. After homogenization of the sample, and removal of plant debris and stones, sieving and air drying, about 20 g was devoted to soil properties laboratory analysis. Seven soil properties for which previous estimations from hyperspectral data were attempted (Gomez et al, 2012a) were determined using classical physico-chemical soil analysis (Baize, 1988): calcium carbonate content (CaCO3), clay content (granulometric fraction $\prec 2\ \mu m$), silt content (granulometric fraction between 2 to 50 μm), sand content (granulometric fraction between 0,05 and 2mm), free iron content, cation-exchange capacity (CEC) and pH.

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

2.2.2. Soil map

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

2.2.3. Airborne HYMAP image and its derivative

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

The image was masked by using NDVI to remove living vegetation (essentially vineyards). The cellulose absorption band (2010 nm) was used to remove dry vegetation. Small areas of bare soils located at the parcel margins or along roads and pathway were also removed since they were not judged as representative of the neighbouring soil surfaces. Finally, the image provided usable data over 33 690 pixels covering 3.5% of the total area only, that is the 192 bare soil fields that were randomly scattered over the region at the date of measurement.

3. Methods

3.1. Experimental set-up

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

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

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

3.2. Non spatial methods

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

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

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

3.3. Spatial methods

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

3.3.1. Variographic analyses

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

3.3.2. Neighbourhood selection

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

3.3.3. Statistical modelling for kriging

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

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

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

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

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

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

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

3.3.4. Kriging with external drift

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

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

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

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

3.3.5. Cokriging

The cokriging CK

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

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

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

3.3.6. Cokriging with categorical external drift

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

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

$$\begin{cases} \displaystyle\sum_{i'=1}^{n_i} \lambda_{i'} C^{(\mathrm{ZZ})}_{i,i'} + \sum_{j=1}^{n_j} \lambda'_j C^{(\mathrm{ZY})}_{i,j} - \sum_{k=1}^{p} \mu_k \mathbb{1}_{\{k\}}(x_i) = C^{(\mathrm{ZZ})}_{i,0} & \text{for } i = 1, \dots, n_i \\ \displaystyle\sum_{j'=1}^{n_j} \lambda'_{j'} C^{(\mathrm{YY})}_{j,j'} + \sum_{i=1}^{n_i} \lambda_i C^{(\mathrm{ZY})}_{i,j} - \sum_{k=1}^{p} \mu_k \mathbb{1}_{\{k\}}(x_j) = C^{(\mathrm{ZY})}_{j,0} & \text{for } j = 1, \dots, n_j \\ \displaystyle\sum_{i=1}^{n_i} \lambda_i \mathbb{1}_{\{k\}}(x_i) + \sum_{j=1}^{n_j} \lambda'_j \mathbb{1}_{\{k\}}(x_j) = \mathbb{1}_{\{k\}}(x_0) & \text{for } k = 1, \dots, p \end{cases} \quad (7)$$

Compared to the previous bivariate cokriging system, the constraints on λ's and λ''s are summed up considering Z and Y have same theoritical mean e_k for each class k. To get a kriging prediction free from class effects e_k, p constraints are necessary so that the sum of weights for the class to whom x_0 belongs must be one, and the sum of weights in all other classes must be 0. 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 obtained by summing the p constraints.

There are p Lagrange parameters μ_1 to μ_p. Only one term μ, the one corresponding to the class at x_0, remains in the kriging variance whose expression is :

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

3.4. Validation

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

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

4. Results

4.1. Co-regionalization models

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

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

4.2. Performance of estimation techniques

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

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

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

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

4.3. Mapping

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

5. Discussion

5.1. Case study representativeness

Bivariate cokriging and the other interpolation techniques were tested in a Mediterranean area that has been used as a case study for digital soil mapping and remote sensing for a long time (e.g. Leenhardt et al, 1994 ; Lagacherie and Voltz, 2000, Lagacherie et al, 2008, Gomez et al, 2012). In spite of its moderate size, it includes a great variety of parent materials and landscape positions that yield complex patterns of soil variations. This was confirmed by the study of variograms of seven soil properties that all exhibited bi-scaled spatial structures and contrasted ratio of short and large-scale variations with properties.

In this study, seven soil properties were considered. This allowed observing contrasted situations with regard to the quality of the auxiliary spatial data used as input of the interpolation techniques. The proportion of variances captured by the hyperspectral-based estimations of soil properties ranged between $R^2 = 0.20$ for sand to $R^2 = 0.78$ for iron, which corresponds to the range of performances shown in the literature (e.g. Selige et al., 2006; Gomez et al., 2008; Ben-Dor et al, 2008, Stevens et al., 2010). As already observed by Ben Dor et al (2002),the soil properties that corresponded to a chromophore (here Clay, Iron, CEC and Calcium Carbonate) were predicted with more accuracy than the other soil properties (sand, silt and pH). The range of proportions of variances captured by the soil map was smaller ($R^2 < 0.31$). From the soilmap assessments performed in the same pedological area (Lennhardt et al, 1994; Vaysse and Lagacherie, 2015), this results correspond to a medium to short scale soil map, that cover substancial proportions of land, e.g. 39% in Europe (King and Montanarella, 2012) and 11% in Africa (Nachtergaele and van Ranst, 2002).

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

5.2. Interest of hyperspectral products as DSM soil input

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

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

However, the seemingly greater interest of hyperspectral products must be nuanced since we did not have in this case study examples of very well-predicted soil properties by a soil map ($R2 < 0.31$). Furthermore, one may remember that hyperspectral products can only deliver estimations of surface soil properties because the effective penetration depths of optical sensors do not exceed several millimetres (Liang, 1997[2]), which limits, at best (i.e. cultivated areas), the soil property predictions to the topsoil horizons only.

5.3. Interest of combining three DSM soil inputs

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

5.4. Future work

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

Concerning the former, two ways could be explored. A better accuracy of the soil map can be obtained by increasing its spatial resolution for obtaining a more detailed soil map. However the number of sampled sites can become a limiting factor since KCED requires a good estimate of the mean value of the property within each soil mapping units, which cannot be obtained without a denser spatial sampling of sites than the one used in this study. Beside, since we observed that much better results were obtained within the bare soil area where hyperspectral estimates of soil property were available without interpolation, it would be worth extending this area. This can be straightforwardly done by a better selection of the date of the fly (Gomez et al, 2012b). Furthermore, the remaining vegetated area can be processed with spectral unmixing (Bartholomeus et al, 2010) or source separation algorithms (Ouerghemmi et al, 2016) for filtering the vegetation signal that may perturb the estimations of soil properties. Finally, other soil sensing techniques than hyperspectral imagery can be used as soil input to enlarge both the area and the exploration depth of the targeted soil properties.

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

6. Conclusion

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

The combination of soilmap and hyperspectral–based estimations of soil property allowed by the novel cokriging with categorical external drift procedure (CKCED) brought improvements for a limited number of soil properties and data configurations. However, to better evaluate its utility, this new combination needs to be tested in other case study with soil maps of better quality and soil sensing techniques covering more area and depths.

7. Acknowledgements

This research was granted by INRA, IRD and the French National research agency (ANR) (ANR-08-BLAN-0284-01). We are indebted to Dr. Steven M. de Jong, Utrecht University in The Netherlands and to Dr. Andreas Mueller of the German Aerospace Establishment (DLR) in Wessling, Germany for providing the 2003 HyMap images for this study. We warmly thank the two anonymous reviewers for their constructive and useful comments.

8. References

Adamchuk, V.I., Viscarra Rossel, R.A. 2010. Development of on–the–go Proximal Sensing

Systems. In: Proximal Soil Sensing , eds: R.A. Viscarra.Rossel, A.B., McBratney, B. Minasny, pp 15–28. Progress in Soil Science 1, Springer Dordrecht, Heidelberg, London New York.

Bartholomeus, H., Kooistra, L., Stevens, A., Van Leeuwen, M., Van Wesemael, B., Ben-Dor, E. 2011. Soil organic carbon mapping of partially vegetated agricultural fields with imaging spectroscopy. International Journal of Applied Earth Observation and Geoinformation, 13, 81–88.

Beisl U. 2001. Correction of bidirectional effects in imaging spectrometer data. Zurich University, Zurich (Switzerland).

Ben-Dor E., Patkin K., Banin A. and Karnieli A. (2002). Mapping of several soil properties using DAIS-7915 hyperspectral scanner data-a case study over clayey soils in Israel. International Journal of Remote Sensing, 23 (6), p. 1043-1062.

Ben-Dor, E., Taylor, R.G., Hill, J., Dematte, J.A.M., Whiting, M.L., Chabrillat, S. 2008. Imaging spectrometry for soil applications. Advances in Agronomy, 97, 321–392.

Ciampalini, R., Lagacherie, P., Monestiez, P., Walker, E., Gomez, C. 2012. Cokriging of soil properties with VisNIR hyperspectral covariates in the Cap Bon region (Tunisia). In: Minasny, B., Malone, B., McBratney, A.B. (Eds.), Digital Soil Assessments and Beyond (). CRC Press. pp. 393–398

Coulouma, G., Barthes, J. P., Robbez-Masson, J. M. 2008. Carte des sols de la Basse Vallee de la Peyne. Report and map UMR LISAH (INRA).

Gomez, C., Lagacherie, P., Coulouma, G. 2008. Continuum removal versus PLSR method for clay and calcium carbonate content estimation from laboratory and airborne hyperspectral measurements. Geoderma, 148 (2), 141–148.

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

Gomez, C., Lagacherie, P., Bacha, S., 2012b. Using Vis-NIR hyperspectral data to map topsoil properties over bare soils in the Cap-Bon Region, Tunisia, in: Minasny, B., Malone, B., McBratney, A.B. (Eds.), Digital Soil Assessment and Beyond. CRC Press, pp. 387?392.

Goovaerts, P. 1997. Geostatistics for Natural ResourcesEvaluation. Oxford University Press.

Grunwald, S. 2009. Multi–criteria characterization of recent digital soil mapping and modeling approaches. Geoderma 152:195–207.

Henderson, B. L., Bui, E. N., Moran, C. J., Simon, D. A. P. 2005. Australia-wide predictions of soil properties using decision trees. Geoderma, 124(3–4), 383–398.

Hengl, T., de Jesus, J.M., MacMillan, R.A., Batjes, N.H., Heuvelink, G.B.M., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Gonzalez, M.R., 2014. SoilGrids1km Ñ Global Soil Information Based on Automated Mapping. PLoS One 9, e105992.

Jenny, H. 1941. Factors of soil formation (p. 281). New York, NY: McGraw-Hill Book Company.

Kempen, B., Brus, D.J., Stoorvogel, J.J., 2011. Three-dimensional mapping of soil organic matter content unsig soil type-specific depth functions. Geoderma 162(1-2), 107-123.

King D. and Montanarella L. 2002. Inventaire et surveillance des sols en Europe. Etude et Gestion des Sols, 9: 137–148.

Lagacherie, P., Voltz, M. 2000. Predicting soil properties over a region using sample information from a mapped reference area and digital elevation data: a conditional probability approach. Geoderma, 187–208.

Lagacherie, P. 2008. Digital soil mapping: a state of the art. In A. E. Hartemink, A. B. McBratney, M. L. Mendonca Santos (Eds.), Digital soil mapping with limited data (pp. 3–14). Springer science.

Lagacherie, P., Baret, F., Feret, J-B, Madeira Netto, J., Robbez-Masson, J.M. 2008. Estimation of soil clay and calcium carbonate using laboratory, field and airborne hyperspectral measurements. Remote Sensing of Environment, 112 (3), 825–835.

Lagacherie, P., Bailly, J.S., Monestiez, P., Gomez, C. 2012. Using scattered hyperspectral imagery data tomap the soil properties of a region. Eur.J. Soil Science; 63:110–119.

Leenhardt, D., Voltz, M., Bornand, M., Webster, R. 1994. Evaluating soil maps for prediction of soil water properties. European Journal of Soil Science, 45(3), 293–301.

Liang, S., 1997. An investigation of remotely-sensed soil depth in the optical region. International Journal of Remote Sensing 18, 3395Ð3408

Malone, B.P., McBratney, A.B., Minasny, B., 2011. Empirical estimates of uncertainty for mapping continuous depth functions of soil attributes. Geoderma 160, 614Ð626.

Marsman, B. A., Gruijter, J. J. de. 1986. Quality of soil maps: a comparison of survey methods in a sandy area. Soil Survey Papers, Netherland Soil Survey Institute, Wageningen.

McBratney, A. B., Mendonca Santos, M. L., Minasny, B. 2003. On digital soil mapping. Geoderma, 117(1–2), 3–52.

Monestiez, P., Allard, D., Navarro Sanchez, I., Courault, D. 1999. Kriging with categorical external drift: Use of thematic maps in spatial prediction and application to local climate interpolation for agriculture, in geoENV II: Geostatistics for Environmental Applications , Gomez-Hernandez J., Soares A. and Froidevaux R. Eds, Kluwer Academic Publishers, Dordrecht, 163–174.

Monestiez, P., Courault, D., Allard, D., Ruget, F., 2001. Spatial interpolation of air temperature using environmental context: application to crop model. Environmental and Ecological Statistics. 8: 297–309.

Mouazen, A. M., Maleki, M. R., De Baerdemaeker, J., Ramon, H. 2007. Online measurement of some selected soil properties using a VIS–NIR sensor. Soil and Tillage Research, 93(1), 13–27.

Mulder, V.L., De Bruin, S., Schaepman, M.E., and Mayr, T.R., 2011. The use of remote sensing in soil and terrain mapping – A review. Geoderma 162, 1–19.

Nachtergaele, F. O., Van Ranst, E. 2002. Qualitative and Quantitative Aspects of Soil Databases in Tropical Countries. In G. Stoops (Ed.), Evolution of Tropical Soil Science: Past and Future (pp. 107–126). Brussel: Koninklijke Academie voor Overzee Wetenschappen.

Odgers, N. P., Libohova, Z., Thompson, J. A. 2012. Equal–area spline functions applied to a legacy soil database to create weighted-means maps of soil organic carbon at a continental scale. Geoderma, 189, 153–163.

Oliver, M. A., Webster, R. 1989. A geostatistical basis for spatial weighting in multivariate classification. Mathematical Geology, 21(1), 15–35.

Ouerghemmi, W., Gomez, C., Naceur, S., Lagacherie, P., 2016. Semi-blind source separation for the estimation of the clay content over semi-vegetated areas using VNIR/SWIR hyperspectral airborne data. Remote Sens. Environ. 181, 251–263.

Richter, R. 1996. Atmospheric correction of DAIS hyperspectral image data. Computers and Geosciences 22, 785–793.

Richter, R., Schlapfer, D.A., 2000. A unified approach to parametric geocoding and atmospheric/topographic correction for wide FOV airborne imagery. Part 2: atmospheric Correction. Proc. 2nd Intl. EARSeL Workshop on Imaging Spectroscopy, Enschede, July 11–13, 2000.

Rossiter, D. G. 2004. Digital soil resource inventories: status and prospects. Soil use and management, 20(3), 296–301.

Schwanghart, W., Jarmer, T. 2011. Linking spatial patterns of soil organic carbon to topography – a case study from south–eastern Spain. Geomorphology, 126, 252-263. DOI: 10.1016/j.geomorph.2010.11.008.

Selige, T., Bohner, J., Schmidhalter, U. 2006. High resolution topsoil mapping using hyperspectral image and field data in multivariate regression modeling procedures. Geoderma, 136, no1–2, pp. 235–244.

Stevens, A., Udelhoven, T., Denis, A., Tychon, B., Lioy, R., Hoffmann, L., Wesemael, B. 2010. Measuring soil organic carbon in croplands at regional scale using airborne imaging spectroscopy, Geoderma, 158, 1–2.

Sun, W., Minasny, B., McBratney, A. 2012. Analysis and prediction of soil properties using local regression kriging. Geoderma, 171–172, 16–23. doi:10.1016/j.geoderma.2011.02.010.

Tenenhaus M. 1998. La regression PLS. Theorie et Pratique. Editions T, Paris.

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

Wackernagel, H., 1995. Multivariate geostatistics. Springer Verlag Editions. 255 pp.

Table 1: Fitted sill and range parameters of direct (samples and hymap) and cross variograms. *in g^2/kg^2 for clay, CaCO3, 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
CaCO3	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**
CaCO3	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**
CaCO3	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

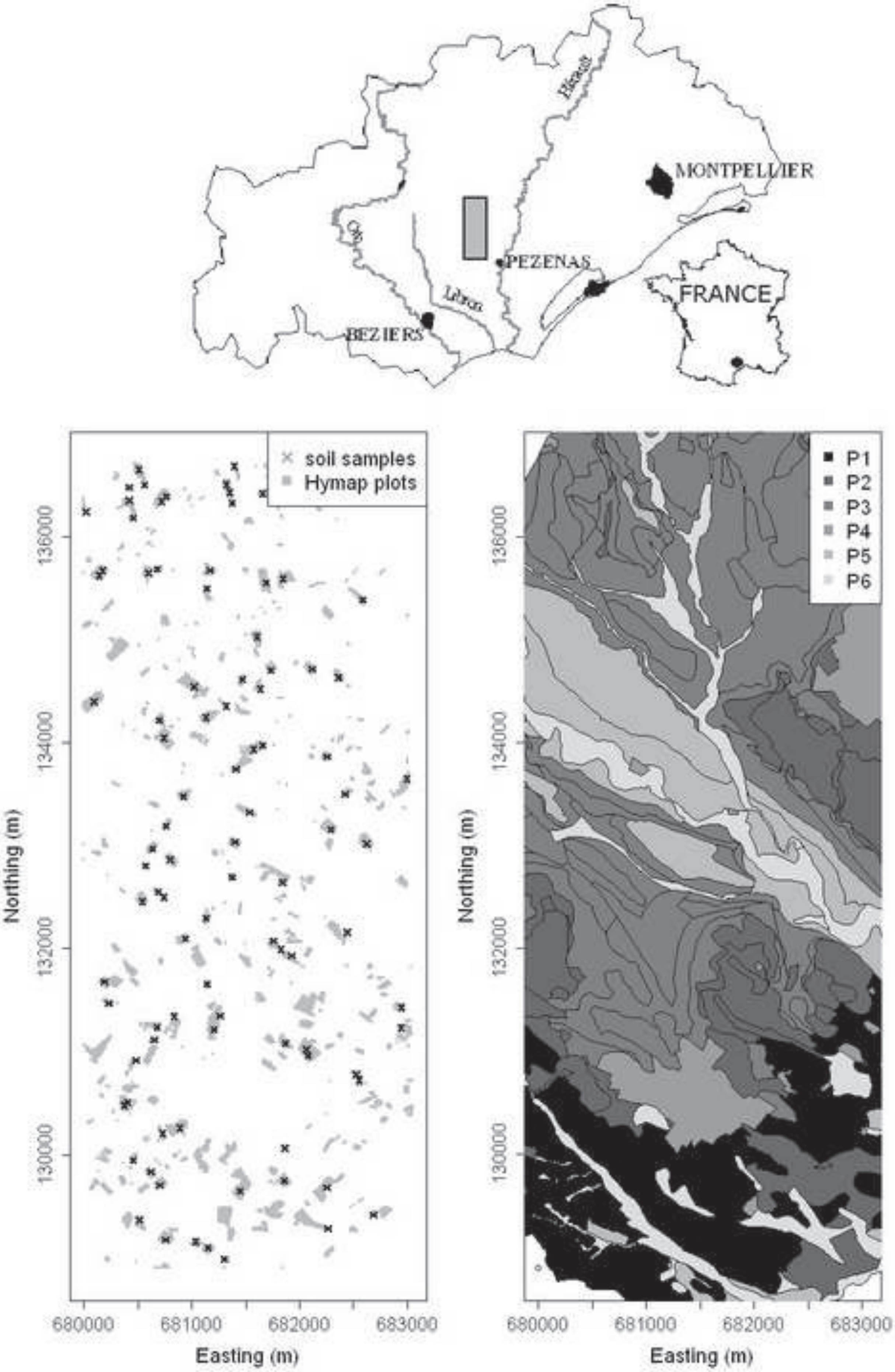


Figure 2

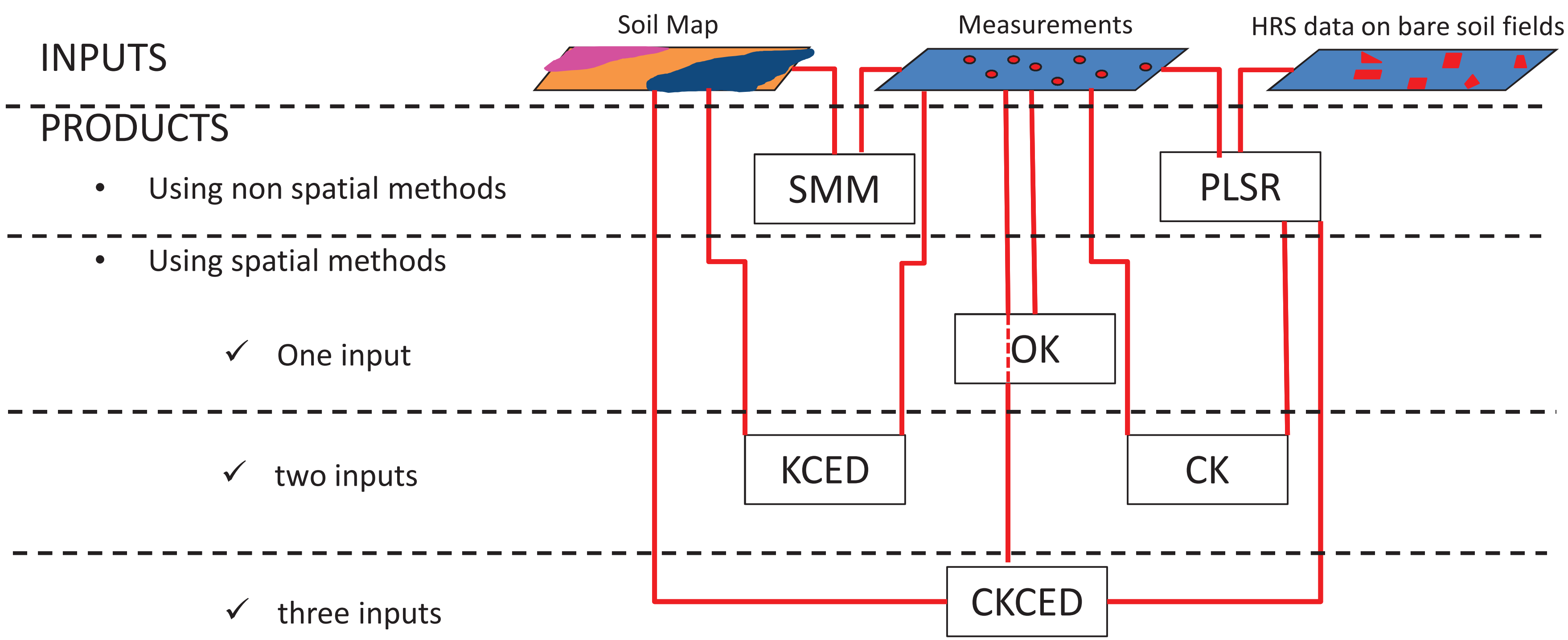

INPUTS
Soil Map
Measurements
HRS data on bare soil fields
PRODUCTS
• Using non spatial methods
• Using spatial methods
✓ One input
✓ two inputs
✓ three inputs
SMM
PLSR
OK
KCED
CK
CKCED

Figure 3

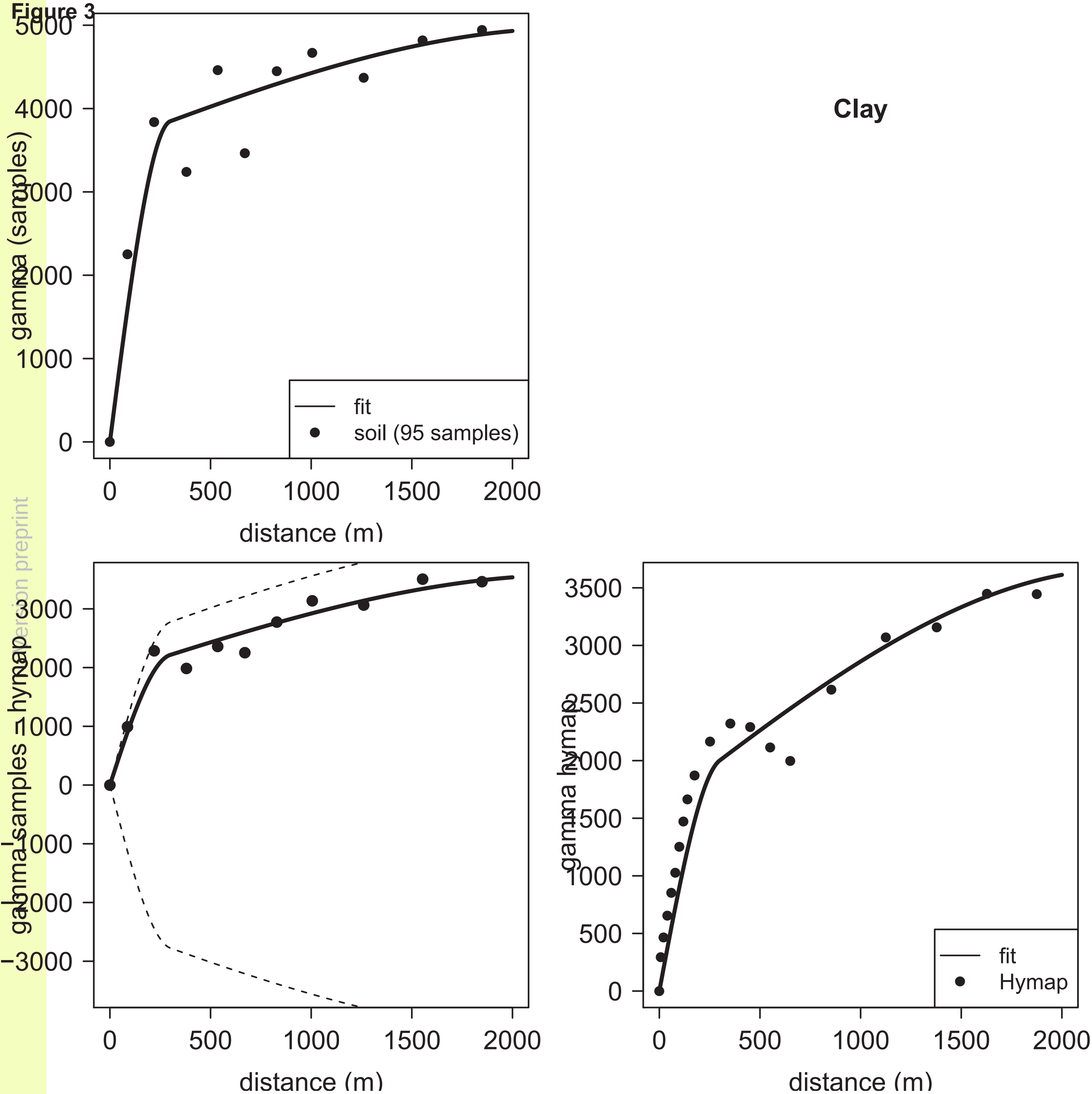
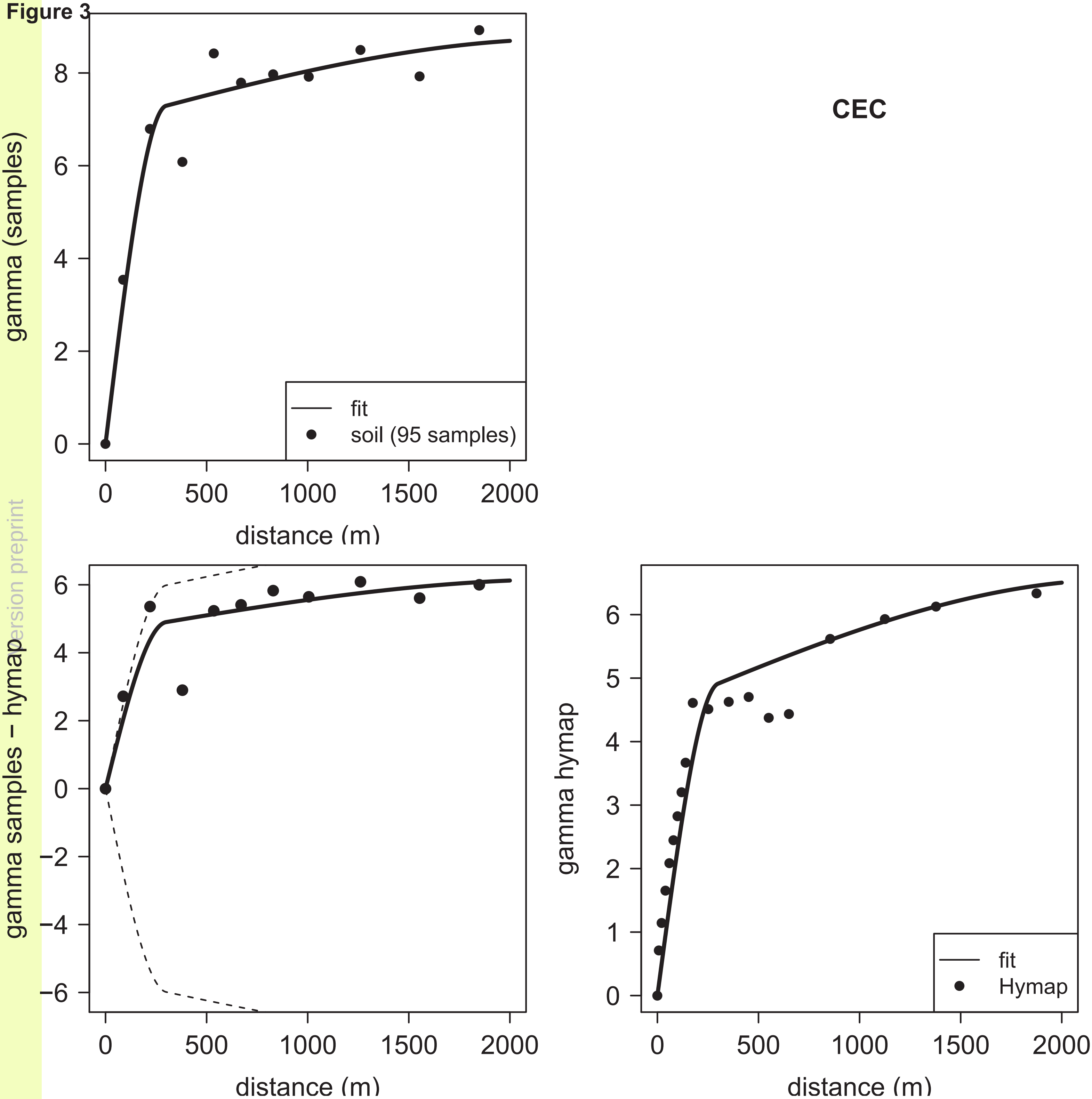

Clay
gamma (samples)
distance (m)
fit
soil (95 samples)
gamma_samples − hymap
distance (m)
gamma hymap
distance (m)
fit
Hymap

Figure 3



CEC
gamma (samples)
fit
soil (95 samples)
distance (m)
gamma samples – hymap
distance (m)
gamma hymap
fit
Hymap
distance (m)

Figure 3

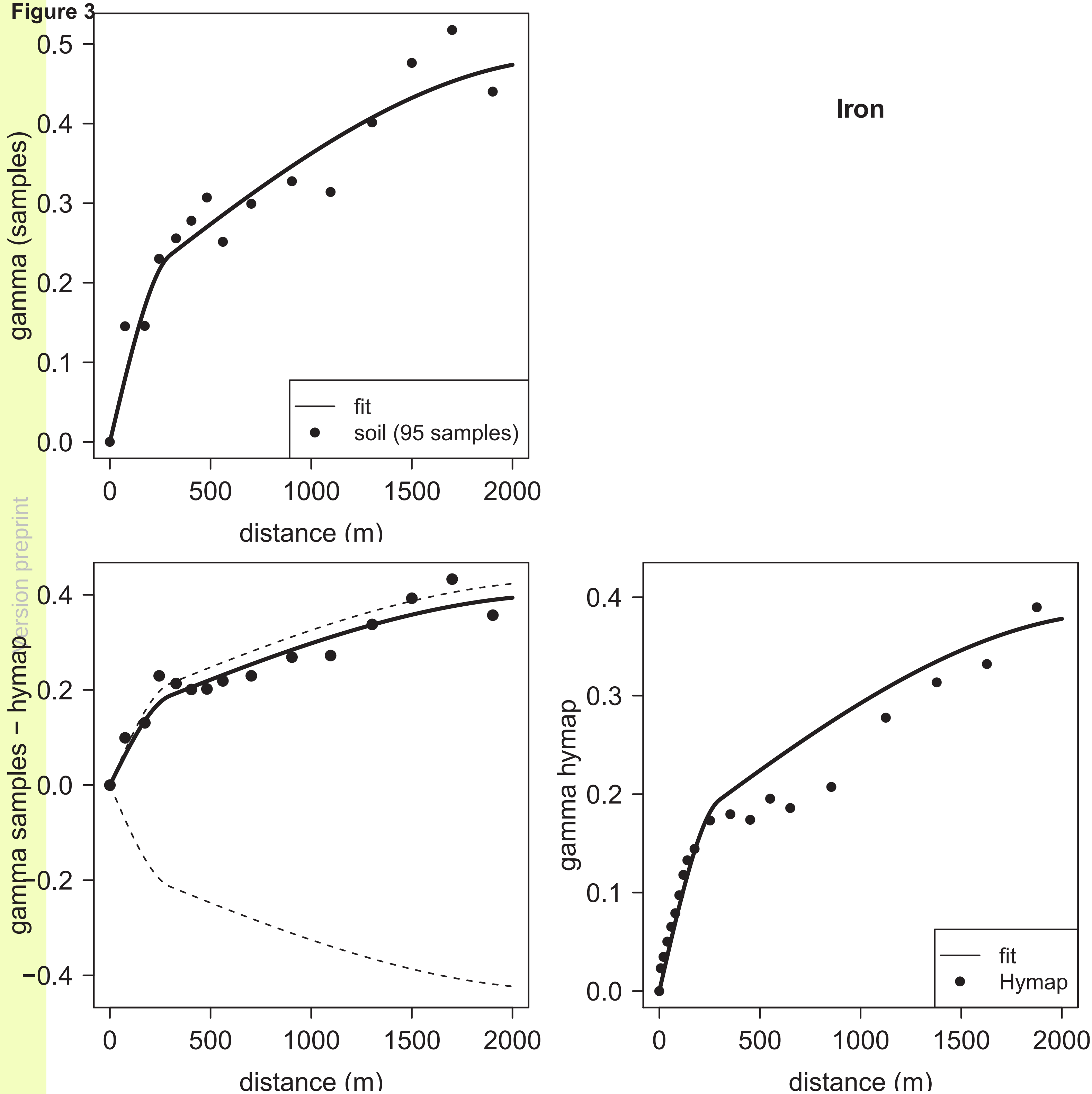

Iron
gamma (samples)
distance (m)
fit
soil (95 samples)
gamma samples − hymap
distance (m)
gamma hymap
distance (m)
fit
Hymap

Figure 4
Click here to download high resolution image

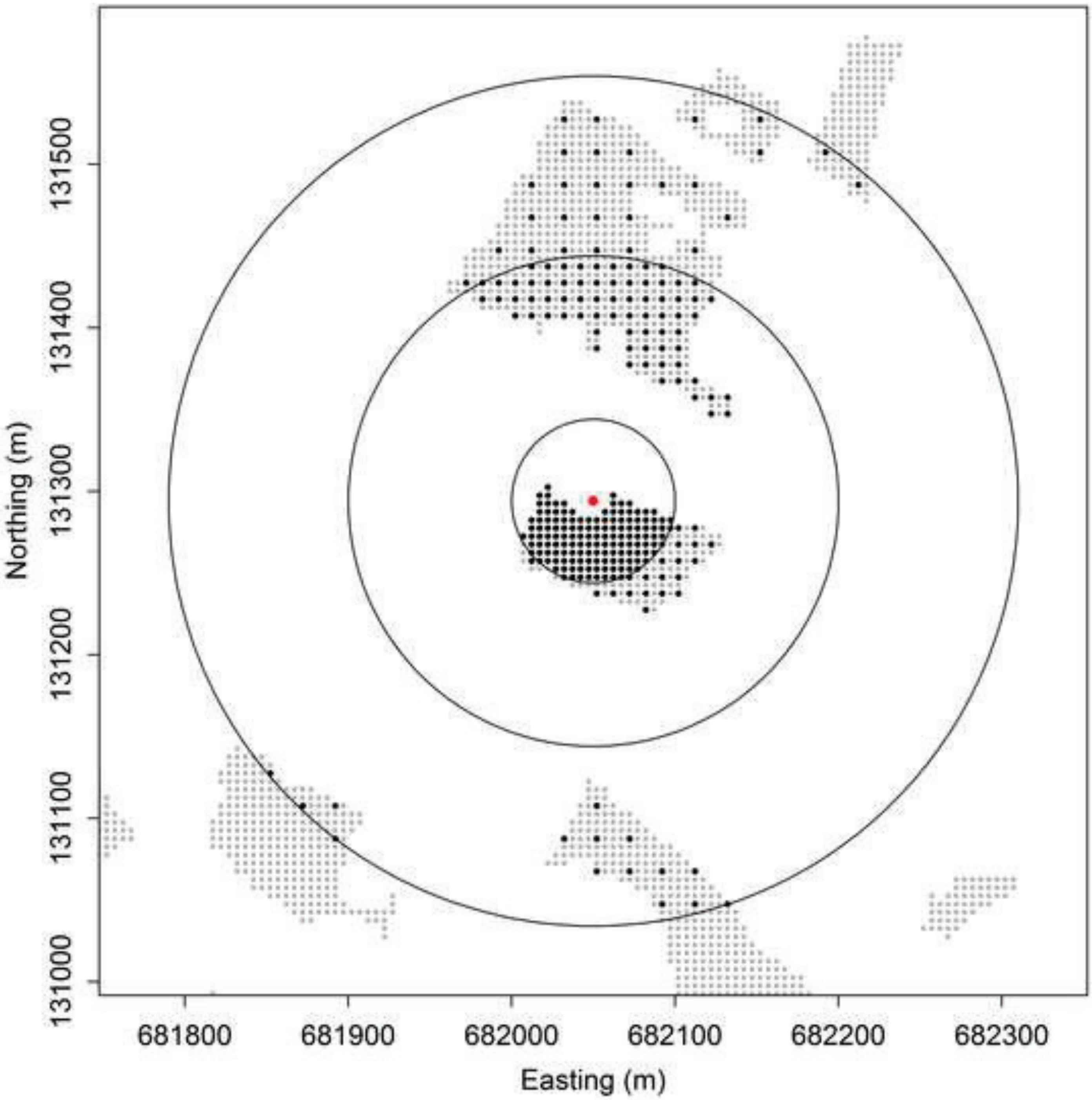


Figure 5
Click here to download high resolution image