



# Regional prediction of soil organic carbon content over temperate croplands using different multiscale measurements by hierarchical modelling

**L. Bel** | UMR MIA-Paris, AgroParisTech, INRA, Université Paris-Saclay | Paris | France

**J. Tressou** | UMR MIA-Paris, AgroParisTech, INRA, Université Paris-Saclay | Paris | France

**M. Zaouche** | UMR MIA-Paris, AgroParisTech, INRA, Université Paris-Saclay | Paris | France

**G. Wahlstedt** | UMR ECOSYS, AgroParisTech, INRA, Université Paris-Saclay | Paris | France

DOI: 10.1481/icasVII.2016.b13

## ABSTRACT

Spatial monitoring of soil quality especially in terms of soil organic carbon (SOC) content for the purpose of evaluating the effects of C-stocking practices at different spatial scales, is critical for both stakeholders and public authorities. Various studies have recently demonstrated the usefulness of hyperspectral satellite images for faster and cheaper prediction of SOC compared to standard soil chemical analyses. In this study we combine both sources of information by means of hierarchical modelling accounting for different spatial scales to provide an accurate map of SOC over a broad regional scale in a peri-urban region.

**Keywords:** SOC content, Hierarchical modelling, Multi-scale, Spatial statistics

## PAPER

### 1. Introduction

The overall decrease of soil organic matter in peri-urban croplands has become a threat for soil sustainability in Europe (Ciais et al., 2010). The COP21 conference recently held in November 2015 in Paris resulted in adopting a “4p1000” initiative aiming at favouring C storage practices to mitigate greenhouse gas emissions. Among possible storage practices, the recycling of exogenous organic matter (EOM) issued from organic waste treatments originating from urban, industrial and agricultural activities provides a promising source of C stocking amendments which may also substitute synthetic fertilizers (Houot et al., 2014; Noiro-Cosson et al., 2015). This study, as a part of the French TOSCA-“PLEIADES-CO” project, basically aims at spatially simulating the effects of such practice at the regional scale of a 221 km<sup>2</sup> peri-urban area close to Paris, where animal breeding has declined. The spatial simulation of such effects of EOM use practice (Noiro-Cosson, 2016) requires spatially explicit and updated information about initial soil organic carbon (SOC) stocks. SOC stock, defined as the carbon mass per unit area for a given depth, is the product of SOC content by considered depth (for a given horizon), soil bulk density, and the percent volume of soil out of rock fragments. Goidts et al. (2009) identified SOC content variability as one of the main sources of predictions uncertainty of SOC stocks and whatever scale, from global to local (Minasny et al., 2013), most studies considered SOC content as the main target variable, and this mainly for the topsoil layer comprised between 8 and 30 cm, being most often the ploughed layer for cropland, i.e. the layer directly modifiable by C-stocking practices such as EOM use. At such detailed scale such as our study area, accurate predictors of SOC content are required: error uncertainties shall be specified to each potential end-user so that, jointly to the map of SOC contents, a map of prediction uncertainty should be provided.

Whatever scale, environmental covariates are the most commonly used to predict SOC content. Depending on the context, either terrain attributes derived from a DEM (Orton et al., 2012a, 2012b; Lacoste et al., 2014), land use/land cover (Lacoste et al., 2014), satellite images (Normalized Difference Vegetation Index (NDVI) derived from remotely sensed images, (Wu et al., 2009)), remotely and proximally sensed visible to near infrared (NIR) reflectance (Peng et al., 2015), geological data (Lacoste et al., 2013), or gamma radiometrics (Malone et al., 2009), may be considered in the modelling.

Numerous methods have been developed for mapping soil properties generally using one or a combination of environmental covariates. They mostly rely on statistical modelling and can be classified into three categories: (i) remote-sensing methods, predicting SOC content by means of a regression from hyper/multi-spectral image reflectance spectra (Selige et al., 2006; Stevens et al., 2010, 2012; Gomez et al., 2012; Vaudour et al., 2013, 2016); (ii) geostatistical

methods, based on a network of soil sampled sites and a set of environmental covariate data being mostly terrain attributes derived from a DEM (Marchetti et al., 2010; Hamiache et al., 2012; Conforti et al., 2015) and/or soil types (Kempen et al., 2011) and/or an airborne gamma radiometric image (Malone et al., 2009); a spectral or vegetation index band of a multispectral satellite image (Wu et al., 2009); (iii) machine learning methods, such as random forest (Grimm and Behrens, 2010), regression trees (Lacoste et al., 2014; Peng et al., 2015; Somarathna et al., 2015) or support vector regression (Somarathna et al., 2015).

In this study, we aimed at mapping SOC contents of agricultural topsoil over a 221 km<sup>2</sup> peri-urban area, the EOM sources of which need to be spatially managed across the 106 km<sup>2</sup> cropland area (Noirot-Cosson et al., 2014). The further use of this map into a soil plant model and for practical decision for end-users dictates the need for uncertainty assessment, as well globally as locally, at the scale of each cultivated field.

A previous study (Hamiache, 2012) provided maps of SOC content by means of multivariate geostatistical methods, from soil measurements and DEM variables. Here we aim at integrating all kind of available information, at different scales and from various sources : hyper/multi-spectral image reflectance spectra, remotely and proximally sensed visible to near infrared (NIR) reflectance, terrain attributes derived from a DEM and soil measurements, in a single hierarchical model to provide an as precise as possible map of SOC content together with its uncertainty assessment.

The core of the hierarchical model is a latent spatial random field representing the SOC content, which is linked to the available data through several relationships that take into account the scale change, and the measurement and modelling uncertainty.

### 2. Hierarchical modelling

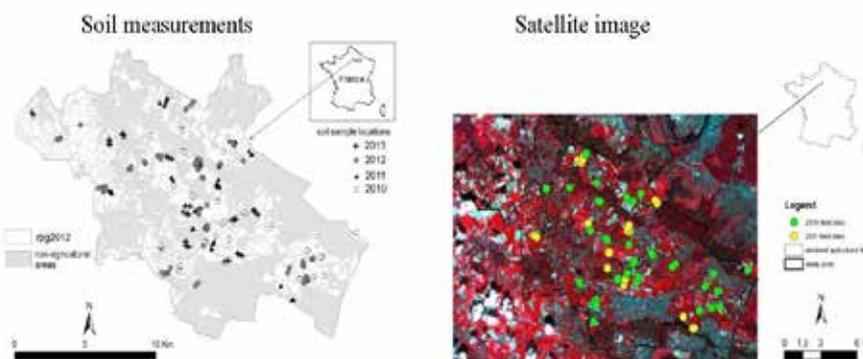
Let  $C(s)$  the SOC content at location  $s$ ,  $s$  in a domain  $S$ . We assume that  $C$  can be written as  $C(s) = X(s)\beta + \varepsilon(s)$ ,  $X(s)$  is a vector of covariates at position  $s$  and  $\varepsilon(\cdot)$  is a stationary Gaussian random field with covariance function  $\rho(\cdot)$ . Covariates  $X$  stem from a DEM (elevation, slope, ...) available at different resolution grids.

We have at our disposal two observation families that can be linked to the SOC content.

- $(Y_1^1, Y_2^1, \dots, Y_{n_1}^1)$  are the values of a satellite image acquired on March 2013, at pixels of size  $\delta_x \times \delta_y$ , centred on sites  $(s_{11}, s_{12}, \dots, s_{1n_1})$  selected because the corresponding areas were bare soils at the moment the image was taken. Usually the  $Y^1$ s are vectors with dimension  $nb$ , the number of spectral bands, but here we consider only the first one (infra-red). We assume that a value  $Y_i^1$  is a noisy affine transformation of the SOC average on the pixel centred on  $s_{1i}$ .
- $(Y_1^2, Y_2^2, \dots, Y_{n_2}^2)$  are values of soil chemical characteristics, including Organic Carbon, acquired from 2010 to 2013 at locations  $(s_{21}, s_{22}, \dots, s_{2n_2})$ .

Figure 1 shows the Plaine de Versailles region together with the locations of the soil measurements and a satellite image of the region acquired on 27 April 2010

Figure 1 - Plaine de Versailles: locations of the soil measurement by year and 2.5 m-SPOT5 image of 27 April 2010. SPOT/ISIS programme. Copyright CNES.



We consider a grid of mesh  $\delta_x \times \delta_y$ , so that we can write  $Y_i^1 = \frac{1}{4} \sum_{s_j \text{ neighbour } s_i} C(s_j) + \varepsilon_i^1$ ,  $(\varepsilon_i^1)$  are i.i.d zero-mean Gaussian random variables with variance  $\sigma_1^2$ .

If we assume that the measurement noise for SOC is normally distributed with variance  $\sigma_2^2$  and considering an affine transformation of the DEM covariate elevation  $E$ , then the hierarchical model is written :

$$C_g \square N(a + bE_g, \sigma_0 \Sigma_g)$$

$$Y^1 | C^1, c, d, \sigma_1 \square N(c + d\bar{C}^1, \sigma_1^2)$$

$$Y^2 | C^2, \sigma_2 \square N(C^2, \sigma_2^2)$$

$C_g$  and  $E_g$  are the vectors of the latent variable and the elevation on the grid,  $\Sigma_g$  the covariance matrix for  $C_g$ ,  $C^1$  is the vector of the  $C_g$  involved in the averaged calculation  $\bar{C}^1$  linking the image and the SOC content,  $C^2$  is the latent variable at locations  $(s_{21}, s_{22}, \dots, s_{2n_2})$  are not on the grid). The covariance function  $\rho(\cdot)$  is given by an exponential model and the matrix  $\Sigma_g$  can be written as  $\Sigma = \sigma_0^2 \Sigma_0$  with  $\Sigma_0(s, s') = \exp(-\|s - s'\| / \alpha)$ . The range  $\alpha$  parameter is estimated previously by usual geostatistical methods from the SOC measurements.

After straightforward calculations the two last equations of the model can be resumed as a linear model  $Y = MC_o + \varepsilon$ ,  $Y$  gathers vectors  $Y^1$  and  $Y^2$ , and  $C_o$  is the vector of the latent variable  $C$  at all the locations involved in the observational layer.

Inferring the whole model consists then to infer the latent variable  $C$  at all the grid points and the parameter vector  $\theta = (a, b, c, d, \sigma_0, \sigma_1, \sigma_2)$ .

The inference is driven in a Bayesian framework, by means of a Gibbs algorithm from classical prior distributions to provide good conjugate distributions. Updating the parameters  $\theta$  and the  $C^1$  is given by the standard formulas for linear models, and the component of  $C$  on the grid that are not included in  $C^1$  are obtained by Gaussian conditional simulation.

## 2. Simulations

To evaluate the algorithm performances and to evaluate to what extent the chains convergence hold, a simulation study is driven on a fine small size grid.

The size of the grid is  $100 \times 100$ , the scale parameter is  $\alpha = 10$ , and the  $\theta$  parameters are  $a = -4$ ,  $b = 2$ ,  $c = 10$ ,  $d = 5$ ,  $\sigma_0 = 5$ ,  $\sigma_1 = 2$ ,  $\sigma_2 = 2$ . We pick at random  $n_1 = 400$  and  $n_2 = 300$  values of the corresponding simulated  $Y^1$  and  $Y^2$ .

The inference algorithm is performed on several chains in order to enable a control of convergence of the distribution with several values for hyper-parameters to investigate the sensitivity to the priors.

Figure 2 - Posterior distribution for  $\theta = (a, b, c, d, \sigma_0, \sigma_1, \sigma_2)$

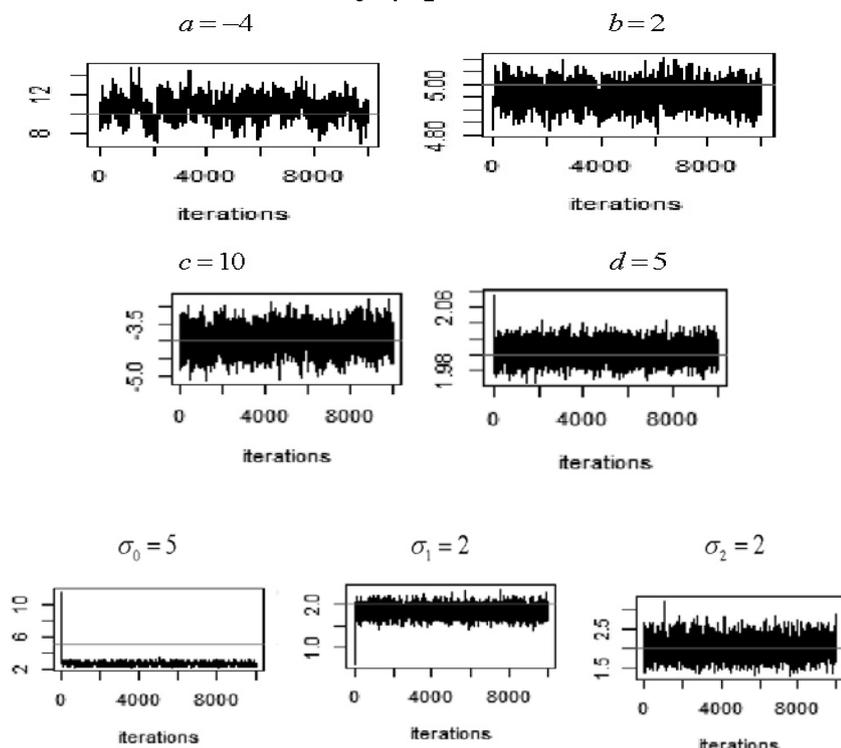


Figure 2 displays the posterior distribution for parameters  $\theta_0$ . For all the parameters except  $\theta_0$  the convergence toward the real value is reached quite rapidly. The lack of convergence of  $\theta_0$  may be due to an identifiability issue, common for spatial fields although the scale is not inferred, and must be investigated.

**Figure 3 - Latent field and averaged posterior fields**

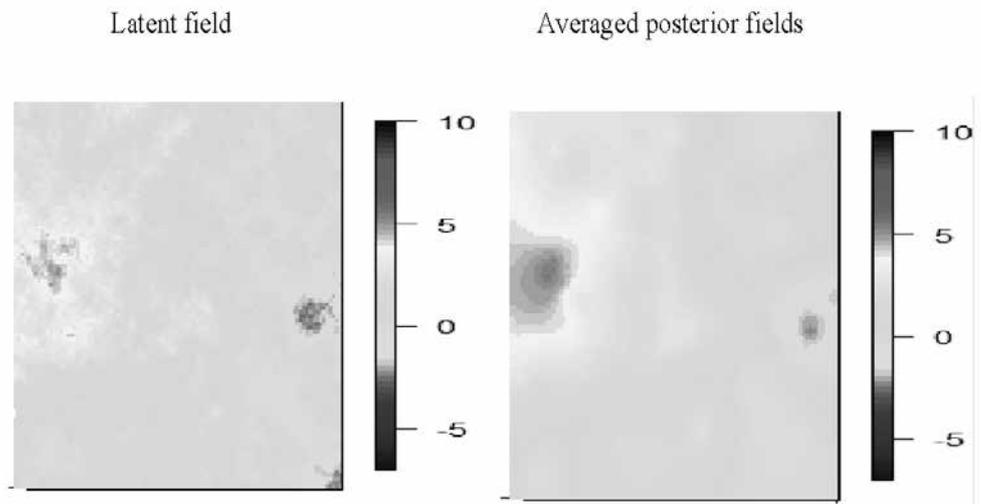


Figure 3 displays the simulated latent field  $C$ , together with some draws of the posterior distribution and the averaged posterior fields showing that the main features of the latent field are recovered in average.

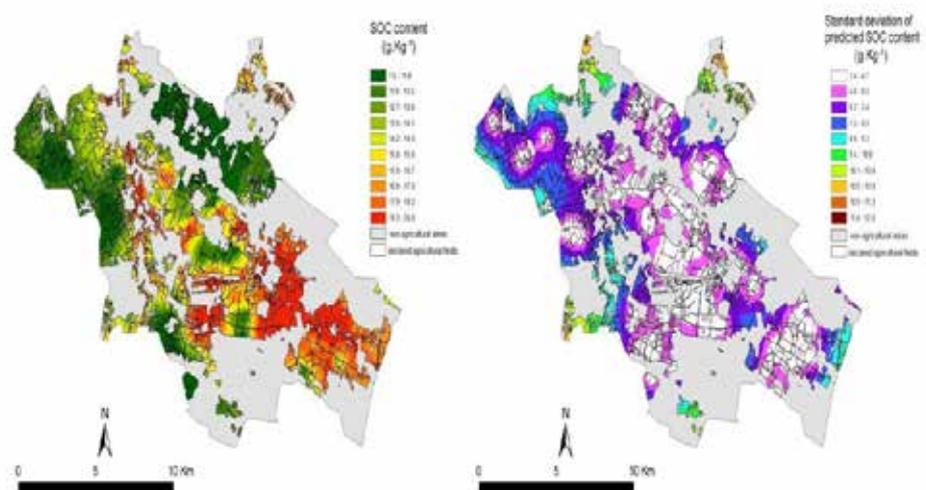
### 3. Real data

The simulation study shows encouraging results, but it is driven for a small dataset size. The dataset we are faced with in an operational framework is an image on a grid of  $243 \times 407$  cells,  $n_1$  10192 cells representing bare soils and  $n_2$  253 top soil measurements.

This leads to severe computational issues, as the conditional simulation step of the Bayesian algorithm needs the inversion of a matrix of size  $n_1 \times n_1$ . Using specific packages for very big matrices and storage utilities could help to carry out the inference, still it remains very time consuming. Making recourse to approximation methods seems unavoidable.

Figure 3 shows a map for the SOC content and the map of its standard deviation obtained with the same dataset except the image, but using geostatistical methods. This map will serve as a reference, to investigate to what extent data from satellite images may improve the mapping of SOC content and what would be the degraded performance using images instead of soil measurements.

**Figure 4 - Mapping for the SOC content and its standard deviation by geostatistical methods**



- Bartholomeus H., Kooistra L., Stevens A., Van Leeuwen M., Van Wesemael B., Ben-Dor E., Tychon B., 2011. Soil organic carbon mapping of partially vegetated agricultural fields with imaging spectroscopy. *Int. J. Appl. Earth Observ. Geoinf.* 13, 81-88.
- Bui, E., Henderson, B., Viergever, K., 2009. Using knowledge discovery with data mining from the Australian Soil Resource Information System database to inform soil carbon mapping in Australia. *Global Biogeochem. Cycles* 23.
- Ciais, P., Wattenbach, M., Vuichard, N., Smith, P., Piao, S. L., Don, A., ... & Leip, A. (2010). The European carbon balance. Part 2: croplands. *Global Change Biology*, 16(5), 1409-1428.
- Conforti, M., Castrignanò, A., Robustelli, G., Scarciglia, F., Stelluti, M., Buttafuoco, G., 2015. Laboratory-based Vis-NIR spectroscopy and partial least square regression with spatially correlated errors for predicting spatial variation of soil organic matter content. *Catena* 124, 60-67.
- Goidts, E., Van Wesemael, B., Crucifix, M., 2009. Magnitude and sources of uncertainties in soil organic carbon (SOC) stock assessments at various scales. *Eur. J. Soil Sci.* 60, 723-739.
- Gomez, C., Rossel, R.A.V., McBratney, A.B., 2008. Soil organic carbon prediction by hyperspectral remote sensing and field vis-NIR spectroscopy: an Australian case study. *Geoderma* 146, 403-411.
- Grimm, R. and Behrens, T. 2010. Uncertainty analysis of sample locations within digital soil mapping approaches. *Geoderma* 155 (3-4): 154-163.
- Hamiache, J., Bel, L., Vaudour, E., Gilliot, J.M., 2012. Spatial stochastic modeling of topsoil organic carbon content over a cultivated peri-urban region, using soil properties, soil types and a digital elevation model. *Digit. Soil Assess. Beyond* 161-164.
- Houot, S., Pons, M.-N., Pradel, M., Caillaud, M.-A., Savini, I., Tibi, A., 2014. Valorisation des matières fertilisantes d'origine résiduaire sur les sols à usage agricole ou forestier. Impacts agronomiques, environnementaux, socio-economiques. Expertise scientifique collective. INRA-CNRS-Irstea. Synthèse. 113 pp. <http://institut.inra.fr/Missions/Eclairer-les-decisions/Expertises/Toutes-lesactualites/Expertise-Mafor-effluents-boues-et-dechets-organiques#>.
- Kempen, B., Brus, D. J., & Stoorvogel, J. J. (2011). Three-dimensional mapping of soil organic matter content using soil type-specific depth functions. *Geoderma*, 162(1), 107-123.
- Malone, B.P., McBratney, A.B., Minasny, B., Laslett, G.M., 2009. Mapping continuous depth functions of soil carbon storage and available water capacity. *Geoderma* 154, 138-152.
- McKenzie, N.J., Ryan, P.J., 1999. Spatial prediction of soil properties using environmental correlation. *Geoderma* 89, 67-94.
- Lacoste, M., Minasny, B., McBratney, A., Michot, D., Viaud, V., & Walter, C. (2014). High resolution 3D mapping of soil organic carbon in a heterogeneous agricultural landscape. *Geoderma*, 213, 296-311.
- Marchetti, A., Piccini, C., Francaviglia, R., Santucci, S., & Chiuchiarelli, I. (2010). Estimating soil organic matter content by regression kriging. In *Digital Soil Mapping* (pp. 241-254). Springer Netherlands.
- Minasny, B., McBratney, A. B., Malone, B. P., & Wheeler, I. 2013. Digital Mapping of Soil Carbon. In D. L. Sparks (Ed.), *Advances in Agronomy* (pp.1-47). Elsevier
- Noirot-Cosson, P.E., Vaudour, E., Aubry, C., Gilliot, J.M., Gabrielle, B., Houot, S., 2014. Scenarios of organic amendment use to increase soil carbon stocks and nitrogen availability in cropped soils at the territory scale: spatial and temporal simulations with the NCSOIL/CERES-EGC crop model. *Geophys. Res. Abstr.*, EGU2014-2656, EGU General Assembly 2014, Vienna, Austria, 2014.
- Orton, T.G., Saby, N.P.A., Arrouays, D., Walter, C., Lemerrier, B., Schwartz, Lark, R.M., 2012a. Spatial prediction of soil organic carbon from data on large and variable spatial supports. I. Inventory and mapping. *Environmetrics*, 23 : 129-147
- Orton, T.G., Saby, N.P.A., Arrouays, D., Walter, C., Lemerrier, B., Schwartz, Lark, R.M., 2012b. Spatial prediction of soil organic carbon from data on large and variable spatial supports. II. Mapping temporal change. *Environmetrics*, 23 : 148-161
- Selige, T., Böhner, J., Schmidhalter, U., 2006. High resolution topsoil mapping using hyperspectral image and field data in multivariate regression modeling procedures. *Geoderma* 136, 235-244.
- Somarathna, P. D. S. N., Malone, B. P., & Minasny, B. (2015). Mapping soil organic carbon content over New South Wales, Australia using local regression kriging. *Geoderma Regional*.

- Stevens, A., Miralles, I., Van Wesemael, B., 2012. Soil organic carbon predictions by airborne imaging spectroscopy: comparing cross-validation and validation. *Soil Sci. Soc. Am. J.* 76, 2174-2183.
- Vaudour, E., Bel, L., Gilliot, J.M., Coquet, Y., Hadjar, D., Cambier, P., Michelin, J., Houot, S., 2013. Potential of SPOT multispectral satellite images for mapping topsoil organic carbon content over peri-urban croplands. *Soil Sci. Soc. Am. J.* 77, 2122-2139.
- Vaudour, E., Gilliot, J.M., Bel, L., Lefevre, J., Chehdi, K., 2016. Regional prediction of soil organic carbon content over temperate croplands using visible near-infrared airborne hyperspectral imagery and synchronous field spectra. *Int. J. Appl. Earth Observ. Geoinf.*, in revision.
- Vaudour, E., Noirot-Cosson, P.E., Membrive, O., 2015. Early-season mapping of crops and cultural operations using very high spatial resolution Pléiades images. *Int. J. Appl. Earth Observ. Geoinf.* 42, 128-141.
- Wu, C., Wu, J., Luo, Y., Zhang, L., DeGloria, S.D., 2009. Spatial prediction of soil organic matter content using cokriging with remotely sensed data. *Soil Sci. Soc. Am. J.* 73, 4, 1202-1208.