Biogeosciences Discuss., 7, 8409–8443, 2010
www.biogeosciences-discuss.net/7/8409/2010/
doi:10.5194/bgd-7-8409-2010

Biogeosciences Discussions

Spatial distribution of soil organic carbon stocks in France

M. P. Martin[1], M. Wattenbach[2], P. Smith[3], J. Meersmans[1], C. Jolivet[1], L. Boulonne[1], and D. Arrouays[1]

[1]INRA Orléans, InfoSol Unit, US 1106, CS 40001, Ardon, 45075, Orléans cedex 2, France
[2]Freie Universität Berlin, Institute of Meteorology, Carl-Heinrich-Becker-Weg 6–10, 12165 Berlin, Germany
[3]School of Biological Sciences, University of Aberdeen, Cruickshank Building, St. Machar Drive, Aberdeen, AB24 3UU Scotland, UK

Received: 8 September 2010 – Accepted: 1 November 2010 – Published: 18 November 2010

Correspondence to: M. P. Martin (manuel.martin@orleans.inra.fr)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

Soil organic carbon plays a major role in the global carbon budget, and can act as a source or a sink of atmospheric carbon, whereby it can influence the course of climate change. Changes in soil organic soil stocks (SOCS) are now taken into account in international negotiations regarding climate change. Consequently, developing sampling schemes and models for estimating the spatial distribution of SOCS is a priority. The French soil monitoring network has been established on a 16 km × 16 km grid and the first sampling campaign has recently been completed, providing circa 2200 measurements of stocks of soil organic carbon, obtained through an in situ composite sampling, uniformly distributed over the French territory.

We calibrated a boosted regression tree model on the observed stocks, modelling SOCS as a function of other variables such as climatic parameters, vegetation net primary productivity, soil properties and land use. The calibrated model was evaluated through cross-validation and eventually used for estimating SOCS for the whole of metropolitan France. Two other models were calibrated on forest and agricultural soils separately, in order to assess more precisely the influence of pedo-climatic variables on soil organic carbon for such soils.

The boosted regression tree model showed good predictive ability, and enabled quantification of relationships between SOCS and pedo-climatic variables (plus their interactions) over the French territory. These relationship strongly depended on the land use, and more specifically differed between forest soils and cultivated soil. The total estimate of SOCS in France was 3.260 ± 0.872 PgC for the first 30 cm. It was compared to another estimate, based on the previously published European soil organic carbon and bulk density maps, of 5.303 PgC. We demonstrate that the present estimate might better represent the actual SOCS distributions of France, and consequently that the previously published approach at the European level greatly overestimates SOCS.

BGD

7, 8409–8443, 2010

1 Introduction

The increasing concentration of greenhouse gases in the atmosphere has led to the need for reliable estimates of the amounts of organic carbon that might be sequestered by soils (Batjes, 1996; Eswaran et al., 1993; Lal, 2004; Paustian et al., 1997; Post et al., 1982; Saby et al., 2008a; Schlesinger, 1991).

Indeed, the organic matter contained in the earth's soils is a large reservoir of carbon (C) that can act as a sink or source of atmospheric CO_2. The world's soils represent a large reservoir of C of about 1500 PgC (Batjes, 1996; Eswaran et al., 1993; Post et al., 1982). Needs of accurate estimates of this pool are of main importance, however their reliability depends upon suitable data in terms of organic carbon content and soil bulk density and on the methods used to upscale point data to exhaustive spatial estimates. Therefore, precise assessments of soil organic carbon stocks (SOCS) based on measurements over large areas are rather few because systematic sampling scheme including soil organic carbon (SOC), bulk density and coarse element content are quite rare (Morvan et al., 2008) and because large levels of SOC spatial variability require very high sampling density to get accurate estimates (Bellamy et al., 2005; Saby et al., 2008b). Jones et al. (2005) developed a methodology for estimating organic carbon concentrations (%) in topsoils (OCTOP) across Europe and recently published a map of SOCS by country. The information is available as a database which can be downloaded from the EU-soils web site (http://eusoils.jrc.it). This methodology, based on pedotransfer functions, gave results which were validated using data from England and Wales and Italy (Jones et al., 2005). However, the match between country level estimates of SOCS using this method and estimates based on national databases depends on the country. For instance, SOCS for the first 1 m in Denmark was estimated (Krogh et al., 2003) to vary from 0.563 to 0.598 PgC (among which 60% is found in the 0–28 cm layer) when the Joint Research Center(JRC)'s estimate is 0.6 PgC for the first 30 cm (Hiederer, 2010). The issue of accurately assessing SOCS, at the country level, is critical because they are used as input for studies about the impact of future land use changes or climate change on SOCS dynamics and potential GHG emissons (Chaplot et al., 2009). In this paper, we apply a new methodology named boosted regression trees (BRT), already successfully applied in India (Lo Seen et al., 2010), to predict the geographical distribution of SOCS in metropolitan France from a set of 1.974 paired observations of SOC and bulk densities. We examine the effects of the main controlling factors of SOCS distribution. We estimate the uncertainty of our national estimate and compare the results with those previously obtained by Arrouays et al. (2001) and Hiederer (2010) on the same territory.

2 Materials and method

2.1 Data

2.1.1 Site specific soil and agricultural data

Soil Organic Carbon Stocks (SOCS) were computed for a set of 1.974 sites from the French soil survey network (RMQS), covering a broad spectrum of climatic, soil and agricultural conditions (Fig. 1). In the near future, the RMQS will cover the entire metropolitan France. The network is based on a 16 km × 16 km square grid and the sites are selected at the centre of each grid cell resulting in about 2.200 soil sampling sites. In the case of soil being inaccessible at the centre of the cell (i.e. urban area, road, river, etc.), an alternative location with a natural (undisturbed or cultivated) soil is selected as close as possible, but within 1 km from the centre of the cell (for more information, see Arrouays et al., 2002).

At each site, 25 individual core samples were taken from the topsoil (0–30 cm) using a hand auger according to a stratified random sampling design within a 20 m × 20 m area. Individual samples were mixed to obtain a composite sample for each soil layer. Apart from composite sampling, at 5 m from the south border of the 20 m × 20 m area, a soil pit was dug, from which main soil characteristics were described and 6 bulk density

BGD
7, 8409–8443, 2010
Spatial distribution of soil organic carbon stocks in France
M. P. Martin et al.

measurements were done, as described previously (Martin et al., 2009). From these data, SOCS were computed for the 0–30 cm soil layer.

$$\mathrm{SOCS}_{30\mathrm{cm}} = \sum_{i=1}^{n} p_i \mathrm{BD}_i \mathrm{SOC}_i (1 - \mathrm{CE}_i) \tag{1}$$

Where n is the number of soil horizon present in the 0–30 cm layer, BD_i , CE_i and SOC_i the bulk density, percentage of coarse elements (relative to the mass of soil) and the SOC concentration (percent) in these horizons, and p_i the fraction of the horizons to take into account to reach the 30 centimetres.

Observational data regarding the land use were also used to assign land use categorical values. Land cover was described using a 3 levels classification, similarly to what is done for the Corine Land Cover maps (Feranec et al., 2010). The level 1 (L1) land covers include various crops (1), permanent grasslands (2), woodlands (3) orchards and vineyards, shrubby perennial crops (4), wasteland (5), specific natural systems (6) and parks and gardens (7). The levels 2 and 3 refine level one. For instance, for specific woody surfaces, one could find the following description: woody surface (L1), forest (L2) and coniferous forest (L3). The number of classes was 7, 22 and 41 for the L1, L2 and L3 levels, respectively.

Water budget of the soils was also described using two variables, *wlogging* and *wregime*, which were used as predictors for SOCS. *wlogging* indicates the possibility of water saturation occurrence going from not occurring or rarely occurring to permanently saturated up to the surface. *wregime* describes the water budget of the soil, from continuously dry to permanently saturated.

2.1.2 Net Primary Productivity data

The Moderate Resolution Imaging Spectroradiometer (Running et al., 2004) Net Primary Productivity (MODIS NPP, $\mathrm{gC\,m^{-2}\,y^{-1}}$) was used to get NPP estimates at each of the RMQS sites. MODIS NPP data are made of $926 \times 926\,\mathrm{m}^2$ resolution raster images that are to be used with corresponding MODIS Land Use raster images, since

BGD
7, 8409–8443, 2010

the MODIS algorithm for estimating NPP depends on the land cover type. Thus, each pixel of the MODIS raster images is given a NPP value as well as a land use value.

The method for estimating a NPP value at the RMQS sites consisted of a three steps procedure. For each RMQS site, first, pixels from the MODIS layer not matching the soil usage of the RMQS site where excluded. Second, mean and standard deviation of NPP values of pixels with matching land cover (i.e. not hidden in the previous step) and not distant of more than a limit distance (d_{lim}) were computed. Four d_{lim} were tested, in $\{5, 10, 20, 30\}$ km. Third, d_{lim} resulting in the highest mean/standard deviation of NPP values was selected. The estimate NPP at the RMQS site was the mean of MODIS NPP values for the selected d_{lim}. Prior to applying this procedure, MODIS land covers were reclassified to match the RMQS land cover classification (L1).

2.1.3 Climatic data

Available climatic data were monthly rain (mm/month), potential evapotranspiration (PET, mm/month), and temperature (°C) at each node of a $12 \times 12\,km^2$ grid (Fig. 4), averaged for the 1992–2004 period. These climatic data were obtained by interpolating observational data using the SAFRAN model (Quintana-Segui et al., 2008), which was initially designed for providing an analysis of the atmospheric forcing in mountainous areas for the avalanche forecasting. The RMQS site specific data were linked to the climatic data by finding for each RMQS site the closest node within the $12 \times 12\,km^2$ climatic grid. This grid was also used in turn as climatic data input when applying the BRT model to the whole territory. Elaborated agro-pedo-climatic data were also derived from the rough data: we used temperature (*a*) and water budget (*b* function of clay, land use and climatic data) mineralization modifiers, as modelled in the RothC model (Coleman et al., 1997), as predictors. The Roth-C mineralization modifiers were in turn compared to rough agro-pedo-climatic predictors (such as rain or land use). The RothC modelling of the influence of water content, *b*, onto the mineralization of soil organic carbon is applicable for soils that are both non-waterlogged soils (Coleman et al., 1997) and not organic organic (Yokozawa et al., 2010). We did not check for the

BGD

7, 8409–8443, 2010

first criteria since the use of other predictors such as *wlogging* and *wregime* gave the possibility to the statistical model to minimize the influence of *b* for specific values of *wlogging* or *wregime* where the RothC modeling would not have been relevant. Regarding the second criteria, following the World Reference Base system, organic soils (histosols) are characterized by organic matter contents above 30% for the first 30 cm (Isss-Isric-Fao, 1998). Our dataset contained only 1 such soil. Hence we did not make specific treatment for this single individual, taking into account the robustness, to the presence of outliers in the dataset, of the statistical models used in this study.

2.2 Data Used for interpolation

2.2.1 Surface data

Soil surface data was taken from the 1/1 000 000 European soil map. Land use data was taken from the TERUTI survey (Chakir and Parent, 2009) provided by the statistical center of the ministry of agriculture (SCEES). This survey comprises 150 000 observational locations where the land use is recorded. The same locations were surveyed yearly between 1992–2004 to determine the land cover and the land use. The survey provides with instant distribution of the land uses as well as temporal transitional data from one land use to another. The 2004 recordings of land use distribution were used for estimating the SOCS distribution. Prior to this, TERUTI data have been reclassified to match an adapted IPCC classification (see legend of Fig. 2).

2.3 Boosted Regression Trees (BRT) Modelling

Boosted regression trees belong to the Gradient Boosting Modelling (GBM) family. GBM is one among many methods to solve the predictive learning problem where the objective is to estimate the function F that maps the values of a set of predictor variables $x = \{x1,..,xp\}$ into the values of the output variable y, by minimizing a specified loss function L. It uses one particular approach to prediction, i.e. classification and regression trees (Breiman et al., 1984), that is extended using a powerful learning technique called boosting (Freund and Schapire, 1996). Boosting methods are generally applied to significantly improve the performance of a given estimation method, by generating instances of the method iteratively from a training data set and additively combining them in a forward "stagewise" procedure. BRT uses a specialized form (for regression trees) of the Stochastic Gradient Boosting (Friedman, 2001). A thorough description of the method is given in Friedman (2001) and a practical guide for using it in Elith et al. (2008).

BRT is known to have improved accuracy compared with simple regression trees, thanks to its stochastic gradient boosting procedure aiming at minimizing the risk of over-fitting and improving its predictive power (Lawrence et al., 2004). The fitting algorithm is an iterative process. At each iteration, individual regression trees, which will compose the final BRT model, are fitted on a fraction (namely the bag fraction) of the dataset sampled without replacement. The main parameters for fitting BRT (boosted regression trees) are the learning rate and the tree size, also known as interaction depth. The learning rate (*lr*), sometimes called shrinkage parameter, is the constant coefficient determining the influence of the individual trees combination that forms the final BRT model. The second important parameter is the tree size (*ts*). It gives the size of individual regression trees. When *ts* is one, each individual tree is made of a single node, thus modeling the effect of only one predictor variable. Then, the final additive model separately includes the effects of the predictor variables and the interactions between variables are not explicitly taken into account. When $ts=i$ and is strictly greater than one, each individual tree models the interaction of at least two predictor variables. This enables the use of models taking into account *i*th order interactions between predictor variables. The ability to represent interactions between predictor variables without a priori knowledge is one of the advantages of BRT and more generally of regression trees. Two other important parameters are the minimum number of observations in the terminal leaves of the trees (min.obs) and the bag fraction (bf).

The contribution of predictor variables are assessed using a variable importance

BGD
7, 8409–8443, 2010

index (VIM), based on the number of times a given variable is selected for splitting individual trees weighted by the square improvement to the model as a result of these trees, summed over all the individual trees (Friedman and Meulman, 2003).

The nature of the dependence between the predictors and the response variable can be assessed by using average or partial dependence plots (Hastie et al., 2001). Put it briefly, they represent the effect of a set of selected predictors (usually 1 to 3) on the modelled response variable after accounting for the effects or the remaining (not selected) predictors.

The BRT models were fitted and used for prediction using the "gbm" R package (Ridgeway, 2006). The stopping criterion for choosing the best iteration when fitting a BRT model was the cross validation method under "gbm" (with crossvalidation folds set to 5), since this method was shown to be the most efficient one (Ridgeway, 2006) amongst the ones available in the "gbm" package. In this study, whatever the BRT parameters' value, the maximum number of allowed iterations was set so that the choice of the model's best iteration did not depend on it. We undertook a tuning procedure for finding out the best combination of these parameters as in Martin et al. (2009).

2.3.1 Models of SOCS

Three models of SOCS were tested, for prediction on the 0–30 cm layer. Two models using all available predictors, among which one aimed at explaining SOC values on forest lands (*F* model), and the other one in cultivated areas (*Cult* model). The third model used only predictors available at the national scale and was applied to prediction at this scale. This model was fitted on the 0–30 cm stocks making up one additional model used for extrapolation (*Extra* model).

The *F* model was fitted on sites under forest (421 sites) and the *Cult* model on cultivated sites (1398 sites) only. This was done in order to facilitate models results interpretation and also because SOCS variability is known for being much more important in forest lands compared to cultivated land (Saby et al., 2008b).

BGD
7, 8409–8443, 2010

The predictors used for each model were:

- the *Cult* model: *lu1*, *lu2* and *lu3* (land use coded according to, respectively, the L1, L2 and L3 RMQS land cover classifications), *clay, silt* (%), *ce* (coarse elements, mass percentage), potential evapotranspiration (*pet*, mm/month), *rain* (mm/month), *ph, wregime* (water regime), *wlogging* (water-logging), the two RothC mineralization modifiers, *a* and *b* and the net primary productivity *npp* ($gC\,m^{-2}\,yr^{-1}$).

- the *F* model shared the same set of predictors except for *lu1* which was excluded since it exhibited only one level for forests.

- the *Extra* model: *lu_ipcc* (land use classification adapted from the IPCC guidelines, 2006), *clay, pet, rain, temp, a, b* and *npp.*

2.3.2 Validation procedure

The BRT models were validated in two ways. The first procedure involved fitting the models to the full dataset (with a restriction regarding the land use for the *C* and *F* models) and validating model predictions on this dataset. The second involved using cross-validation. The first procedure enabled to estimate the quality of the fit of the models of C prediction. Only the second validation procedure, which involved validation against independent data, enables to estimate the predictive power of the proposed models.

In both procedures, comparison between observed and predicted values of SOCS was carried out using several complementary indices as it is commonly suggested (Schnebelen et al., 2004): the mean prediction error (MPE), the standard deviation of the prediction error (SDPE), the root mean square prediction error (RMSPE) and the prediction coefficient of determination (R^2) measuring the strength of the linear relationship between predicted and observed values.

BGD
7, 8409–8443, 2010

The second validation procedure was done following principles similar to K-fold cross-validation, enabling us to perform what will be referred to in the following as external validation. 90% of the individuals was drawn randomly without replacement from the dataset and used as the training dataset. Validation was done on the remaining 10% of individuals (external validation). This procedure was repeated 1000 times, which provided robust results. External validation was used as a way to explore the predictive power of the resulting model for previously unseen data. In the following, the MPE, SDPE, RMSPE and R^2 indices, computed through this external validation, are adjoined the ext suffices (i.e. MPE_{ext}, $\text{RMSPE}_{\text{ext}}$ and so forth). Enclosing indices with the < and > signs indicates that the median value over the 1000 trials is given (for instance $<\text{MPE}^2_{\text{ext}}>$). $\text{RMSPE}_{\text{ext}}$ resulting from cross validation were also estimated as a function of SOCS values (Fig. 4). This enabled us to refine the estimation of uncertainty related to the estimation of the spatial SOCS. The error on the SOCS estimate for the whole territory was obtained by summing the errors on each elementary surface unit:

$$\Delta\text{SOCS} = \sum_{j=1}^{m} \text{S}_j \text{RMSPE}(\text{SOCS}_j) \quad (2)$$

where ΔSOCS is the global error, S_j is the surface of the elementary surface j, SOCS_j its estimated SOCS and RMSPE() the function relating the predicted SOCS to the model error.

2.3.3 Parameter settings for BRT models

Although some general recommendations exist for setting the values for tree size, learning rate, minimum number of observations in the terminal nodes values and bag fraction, a tuning procedure was run, because, in practice, as for single regression trees, optimum values may depend on the dataset (Lilly et al., 2008). *ts*, *lr*, min, obs and *bf* values were taken, according to recommendations found in the literature (Lilly

BGD
7, 8409–8443, 2010

et al., 2008; Ridgeway, 2006) from {5, 6, 7, 8, 9, 10}, {0.05, 0.075, 0.1, 0.125, 0.15}, {2, 4, 5, 10, 15} and {0.5, 0.75, 1}, respectively. This tuning procedure was carried out as in (Martin et al., 2009). The choice of the best parameter values combination was carried out, for the three models independently, looking for the best $<R^2_{ext}>$ values. This tuning procedure resulted in setting *lr*, *ts* and min.obs and *bf* to, {0.01,12,8}, {0.01,4,8}, and {0.01,8,8} respectively for the *C, F* and *Extra* models.

Selecting these parameter settings for each of the models was a preliminary step in the study. We then assumed that these settings could be applied to all subsequent fits. They were thus used in turn for producing all the results displayed in the paper, i.e. regarding (i) the BRT models' performance on the full dataset and (ii) the predictive performance tested against independent data.

3 Results

3.1 Observed SOCS

The SOCS depended greatly on the land cover type (Fig. 2). Highest values were observed for the forest, grasslands and wetlands (only two observations though). On the first 30 cm, the stock in forest (mean SOCS of 7.00 kg/m²) was less than under permanent grassland (mean SOCS of 7.57 kg/m²) with comparable standard deviation (3.42 and 3.51 kg/m², respectively). Dispersion of values on cultivated areas, excluding permanent grasslands was low (1.85 kg/m²) compared to permanent grasslands and forest lands. Lowest SOCS values were observed for vineyards (mean SOCS of 3.2 kg/m²) and some uncultivated coastal areas (mean SOCS of 2.42 kg/m²).

3.2 Goodness of fit and predictive performance

General indices of agreement of the models prediction and the observed data (MPE, SDPE, RMSPE, R^2), are given in Table 1. BRT models yielded good results when fitted

BGD
7, 8409–8443, 2010

on and validated against the full dataset. The fit was best for the Cult model, with R^2 value of 0.91 and RMSPE value of 0.934. The prediction was worse on forest soils, where the *F* model yielded 0.74 and 1.910 values for R^2 and RMSPE, respectively. For the three models, MPE was negligible indicating models with low precision and high accuracy. Ranking of the models performance using cross-validation was the same than according to validation on the dataset used for learning. The *Extra* model, developped for prediction on soils under any kind of landuse yielded $<R^2{}_{\mathrm{ext}}>$ value of 0.5 (with 95% confidence interval of [0.386, 0.613]) and $<\mathrm{RMSPE}_{\mathrm{ext}}>$ of 2.271 ($\mathrm{CI}_{95\%}$ of [1.862, 2.68]). $<\mathrm{MPE}_{\mathrm{ext}}>$ values, representing the bias, were in average low, if not negligible and reached −0.002 for the *Extra* model. For this model, the $\mathrm{CI}_{95\%}$ for $<\mathrm{MPE}_{\mathrm{ext}}>$ was large ([−0.348, 0.344]) indicating that some models, depending on the sub-dataset used for fitting produced significantly biased predictions on the sub-dataset used for validation. This model underestimated SOCS for low observed SOCS and overestimated SOCS for high observed values (Fig. 5). The best of the three models, when validated using cross-validation was the *Cult* model, with a $<R^2{}_{\mathrm{ext}}>$ value of 0.54 ([0.393, 0.688]) and $<\mathrm{RMSPE}_{\mathrm{ext}}>$ of 2.046 ([1.557, 2.536]).

The analysis of the *Extra* model's error (Fig. 3) indicates a positive correlation between the observed C stock value and the $<\mathrm{RMSPE}_{\mathrm{ext}}>$, estimated within C stock classes. Expected $<\mathrm{RMSPE}_{\mathrm{ext}}>$ lies between 1 and 3 kg/m² for SOCS belonging to the [2, 14] kg/m² range. Uncertainty on the error estimate itself can be computed and the results, as shown on Fig. 3, indicates $<\mathrm{RMSPE}_{\mathrm{ext}}>$ values under 8 kg/m² for SOCS below 15 kg/m². Above this threshold, mean $<\mathrm{RMSPE}_{\mathrm{ext}}>$ as well as the upper limit of the confidence interval rises indicating a very high uncertainty of the results in the model's prediction. CI95% could not be computed above 18 kg/m² because of the rarity of such high observed values.

BGD

7, 8409–8443, 2010

Spatial distribution of soil organic carbon stocks in France

M. P. Martin et al.

Title Page
Abstract | Introduction
Conclusions | References
Tables | Figures
◀| |▶
◀ ▶
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

3.3 Variable relative influence

The computation of the VIM values associated to the predictors for the three models (Table 2) indicates a strong influence of clay content. This predictor ranks second for the *Cult* model and first for the the *F* and *Extra* models. Rain is consistently ranked in the four most important predictors. For the Cult and *Extra* models, the land use appears to be important for predicting the SOCS. The fit of the *Cult* model showed that it is worth using a detailed description of the land use, since the *lu2* and *lu1* predictors had a negligible importance, whereas the *lu3* predictor had the most important VIM index. However, for the *F* model, the *lu3* variable, which in this case represent the kind of forest considered had a very low variable importance index. The VIM index value for coarse elements was more important for the *F* model than for the *Cult* model, and was ranked fourth. On the *F* model, the npp values computed on each RMQS site ranked fifth. On the *Cult* and *Extra* models, the temperature, best represented by the transformed a variable ranked 3 and 4, respectively. Temperature exhibited a limited importance for the *F* model, as *pet* did whatever the model.

3.4 Map of soil organic carbon stocks

The total stock for France (0–30 cm) computed on the 12 × 12 km² grid was 3.242 PgC for a surface of 541 060 km². The total surface represented by the grid is slightly smaller than the actual French metropolitan territory (543 965 km²). The total stock for the French metropolitan territory could thus be rescaled to 3.260 PgC. Estimated uncertainty was 0.872 PgC. Predicted SOCS ranged from 2.0 to 15.8 kg/m² over the French territory. The highest stocks were observed in mountaineous areas (Alps, Jura, Massif Central and Pyrénées), in Brittany and in parts of Lorraine regions (Fig. 4).

The comparison of empirical cumulative distribution function (ecdf) between the observed SOCS on RMQS sites, and the surface estimate from the *Extra* model reveals several aspects of the spatial prediction quality (Fig. 5). It shows that although the *Extra* model managed to reproduce the distribution of the observed values, when applied

BGD
7, 8409–8443, 2010

to the whole territory, the resulting distribution exhibits a sharp pattern by narrowing the range of predicted values. The variability on the predicted map was smaller than on observed or predicted SOCS values on RMQS sites, but the distributions were centered around close median values.

4 Discussion

4.1 Validity of the estimate

The total SOCS estimate was in good agreement with a previous estimate (3.1 PgC on a soil mass equivalent to 30 cm under forest, (Arrouays et al., 2001)). However, it disagreed with the estimate based on the organic carbon content layer available at the European level (Jones et al., 2005) of 5.0 PgC for the first 30 cm (Hiederer, 2010). We recalculated this estimate by combining JRC's octop layer (1 km × 1 km resolution, Jones et al., 2005) and a spatial layer of bulk density ($10' \times 10'$ grid, Smith et al., 2005) in topsoils (0–30 cm). Adjusting the resolution of the octop and bulk density layers to the resolution of our 12 km × 12 km grid was done using the ARCGIS zonal statistics algorithm for the SOC content and a weighted mean procedure for the bulk density layer. Our global estimate using these data layer was 5.303 PgC. This values lies outside the interval defined by taking into account the uncertainty associated to BRT model (±0.872 PgC). The magnitude of the overestimation related to the JRC's European SOC content layer matched the one found by Dendoncker et al. (2008) at a much smaller scale for a small area of southern Belgium. Assuming that because of it systematic sampling scheme, the RMQS dataset is representative of the French territory, its cumulative distribution of SOCS can be used as a reference of SOCS in France. Figure 5 showed that the distribution resulting from the processing of JRC data consistently overestimated the SOCS. On the other hand, the *Extra* model spatial estimate was unbiased but the occurrence of high SOC values (above 8 kg/m^2) was much lower than for the distribution on RMQS sites. This discrepancy was not observed

BGD

7, 8409–8443, 2010

Spatial distribution of soil organic carbon stocks in France

M. P. Martin et al.

for values under the SOC median value (circa 5 kg/m²). Thus the total estimated SOCS might underestimate the real SOCS for France but according to Fig. 5 the absolute error of the estimate provided here was less than the one observed with the JRC data. The comparison between the empirical cumulative distribution function of observed RMQS SOCS and the one provided by the *Extra* model suggests that the distribution tails are poorly represented, i.e. that the extreme SOCS values where not reached by the model. This is likely to result from the spatial distribution of the predictors, since the model managed to predict extreme values when applied to the RMQS sites. The fact that there was (not shown here) a similar difference for clay, the most important predictor in the *Extra* model, between the ecdf of the spatial layer and the one of RMQS sites, supports this statement.

It can be argued that here, the resolution of the native datasets (especially for the SOC content layer of the JRC, 1 km × 1 km) are very different from the one presented in this paper. The aggregation of the data up to the 12 km × 12 km² may explain locally some of the differences with the estimate provided by the BRT model. However, at the national scale, i.e. when summing the SOCS over the whole map, the aggregation itself is not expected to explain much of the difference observed here. More likely, the difference between the results resulting from both methodologies, come from SOC and bulk density estimates themselves. The JRC SOC content estimate results from pedotransfer fitted on the European soil database (at a scale of 1:1 000 000) and validated on England, Wales and Italy only. Bulk densities have been estimated using pedotransfer rules as well. On the contrary the present estimation relies on a model fitted and validated against a systematic sampling scheme (16 km × 16 km resolution) with both SOC content and bulk density measurements.

Modelling CO_2 emissions from soils is in many cases multiplicative regarding the current SOCS (as in the RothC model, taking mineralization related to temperature change as an example). As a result, a simulation of SOCS changes for France under diverse scenarios can potentially result in very different estimates of emissions, whether a 3.260 PgC or a 5.303 PgC is considered as being the reference SOCS value.

BGD
7, 8409–8443, 2010

M. P. Martin et al.

The performance of European soil monitoring networks (SMN) for detecting long term SOC change trend has recently been demonstrated (Saby et al., 2008b), based on the JRC's SOC content map (Hiederer et al., 2004). Such SMN's could as well be used to refine continental SOCS estimates and in turn refine the estimation of their own performance. Numerous studies regarding SOC changes at this specific scale (Smith et al., 2005; Zaehle et al., 2007) could as well benefit such improvement of European wide SOC content or stocks characterization.

The uncertainty estimated for the BRT model results from the application of the uncertainty function depending on SOCS values provided by the cross validation trials (Fig. 3). The fitted model is characterized by very high uncertainties for SOCS values above 15 kg/m². Uncertainty on this estimate itself starts to increase notably from 11 kg/m², making it difficult to draw any conclusion about the validity of the model for such SOCS values. On the contrary for values under 11 kg/m², the value of the uncertainty of predicted SOCS values can be accurately known. The model error ($<RMSPE_{ext}>$) is comparable to results of other study studies based on different statistical techniques but among the few providing an assessment of model predictions based on crossvalidation. Different geostatistical models yielded a global estimate of 4.54 ± 0.74 PgC for Laos (Phachomphon et al., 2010) and a elementary RMSE of 2.89 kg/m² when mapping 0–50 cm SOCS for the Indiana state (Mishra et al., 2009). The quality of the fit was better than for recent studies applying generalized linear models to the prediction of SOCS in Tibetan grasslands and explaining 73% of the variation of SOC densities (Yang et al., 2008), to be compared to the R^2 of 0.73, 0.74 and 0.91 of the *Extra*, *F* and *Cult* models presented here. On the RMQS dataset, the SOCS values above 15 kg/m², which could be considered outside the validity domain of the BRT model are rare (2% of the RMQS sites display SOCS values above 15 kg/m², Fig. 5). The predicted distribution of SOCS includes a negligible fraction of SOCS above 15 kg/m² (below 0.01%), and consequently such surfaces, where estimated uncertainty is high, have a negligible impact on the global uncertainty related to the national SOCS prediction (0.14%).

BGD
7, 8409–8443, 2010

4.2 Relative importance of the predictors

4.2.1 Effect of the land use

Discrepancies between the *Cult* and the *F* models might give an estimate on how agricultural practices, both in grassland and arable lands, determine the relationships between pedo-climatic variables and SOCS, compared to forest systems. For instance, the lesser importance of soil pH for the *Cult* model might have resulted from the influence of some agricultural practices onto this chemical parameter. Similarly, it was possible to show (not display here) that the effect of clay depended on the land use and was attenuated for croplands. This might be explained by the fact that farmers have, for crop cultivation for instance, the chance of mitigating the influence of an unfavourable water budget, related to low clay contents, by tuning the cultivation calendar or the irrigation timing. More generally, the *F* model performed much worse than the *Cult* model (R^2_{ext} are 0.36 and 0.58, respectively). This means that the SOCS under forest have a great amount of variability that remained unexplained by the set of variables that were included in the model. The VIMs of predictors related to the land use showed that if in some case a detailed land use description is relevant (predictor *lu3* in model *Cult*), a coarser description (i.e. *lu_ipcc* in model *Extra*) is still valuable for predicting SOCS and of the same importance as information about the clay content (Table 2).

4.2.2 Effect of the soil properties

The modelled effect of clay onto SOCS was monotonic increasing (Fig. 6, (a)). This expected effect may results from several processes. The most commonly cited is the physical interaction, mediated by various soil elements and biological activity, between the clay materials and organic compounds (Arrouays et al., 2006; Chaplot et al., 2009). It tends to protect OM from decomposition (Liao et al., 2009). The modelled response to clay content may include other processes such as the influence of clay onto the soil's moisture regimes *via* its influence on the water holding capacity (Wosten et al., 1999). The soil moisture regime itself influences the mineralization (Bauer et al., 2008) as well as the plant primary production, and in turn soil carbon inputs and outputs. From Table 2, the combination of *clay* with the climatic variables, as it is done within the RothC model (predictor *b*) was of much less importance than variables such as *rain* or *clay* alone. Thus the modeling of the relationship between clay, rain and PET on one side and mineralization on the other side, as it is done in the RothC model, was not, in average, relevant for our dataset.

The inclusion of water content variables (*wlogging* and *wregime*) did succeed poorly. This was surprising, since the soil water regime has been reported, as well in field experiment as on large scale statistical surveys (Meersmans et al., 2008) to influence the SOM decomposition and consequently the observed SOCS. There might be several reasons for this, mainly coming from the available dataset. In many cases (25%) this information was missing, which decreases the final VIM of this variable in the fitted models. Secondly, the water regime was available at the whole profile level only and might not have been representative of the first 30 cm. Thirdly, this water regime was informed based on the observation at the sampling time, and, again, might not have been representative of the water regime across the year.

4.2.3 Effect of the climatic variables

The relationship between climatic variables and amount of organic carbon in soil is also well known, and again, is linked to the effect of these variables onto plant productivity on one side, and soil carbon decomposition on the other. The effect of these variables, as they are measured here (rain, PET, above ground temperature), is mediated by soil properties and the vegetation cover. As such, the *rain* predictor was consistently one of the most important one. The effect of temperature (predictors *temp* and *a*), which may be dependent upon other variables such as physical protection, chemical protection, drought, flooding and freezing (Davidson et al., 2000), was important too, but less than the effect of the rainfall. Temperature increase enhance NPP and mineralization at the same time (Heimann and Reichstein, 2008), assuming that temperature remains

BGD
7, 8409–8443, 2010

below a given threshold. The trade-off between mineralization and NPP increase determines the sign of relationship between SOCS and temperature. Here, the relationship between SOCS and *a* was monotonic decreasing (not displayed here), which could indicate that the effect of temperature onto mineralisation is, in France, more important than the effect onto NPP.

4.3 Possible improvements of the models

From the current models of SOC dynamics, the influence of decomposition modifiers (here *a*, *b*) is expected to be of same magnitude as the estimated soil carbon inputs (Martin et al., 2009). Nevertheless, our estimate of carbon inputs, the *npp* variable, had a low VIM value. This demonstrated that our estimate was inaccurate. Both the resolution of the MODIS data and algorithms used for providing NPP, and our procedure for retrieving values at our sampling locations might have resulted in an irrelevant *NPP* predictor. Additional work would be necessary for estimating more accurately SOC inputs on the RMQS sites.

Topography was not taken into account in this study. Indeed it has been shown that it is relevant to SOCS prediction. Importance of Digital Elevation Models derived variables has been demonstrated at the national (Chaplot et al., 2009) and small region scale (Grimm et al., 2008). This might be related to the redistribution processed related to soil erosion for instance. In our case this information was not readily available both at the RMQS sites locations and at the national scale and thus was not included in the models.

Fe and Al oxydes, CEC are also known for being correlated to SOCS (Chaplot et al., 2009). Although this information was available along with SOCS measurements at RMQS sites, this information cannot be seen as an external variable which the SOCS is a function of, because these soil properties, and mainly CEC, are difficult to inform spatially, as SOCS are for that matter. Consequently their use for predicting SOCS spatial distributions is limited.

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

BGD

7, 8409–8443, 2010

Spatial distribution of soil organic carbon stocks in France

M. P. Martin et al.

In addition SOCS are dependent upon mineralogy. For instance, a positive relationship between the occurrence of non-crystalline minerals such as allophanes in volcanic soils (Torn et al., 1997) or 2:1 clays like smectite (Grimm et al., 2008) and organic carbon was observed. However, in France, the clay mineralogy is not very much contrasted, and including this predictor in the model is likely (but *in fine* this should be tested in future developments of this study) to have led to little improvement of the model. Moreover, Grimm et al. (2008) remind the fact that clay amount might be more important than clay mineralogy. This, in addition with the fact that this information (as well as soil type) was not readily available for the RMQS, resulted in not including it in our analysis.

The best next candidate among soil properties would be the soil pH. The spatial distribution at a national scale of this predictor, relevant for forest soils, will be accessible in a near future. Its omission in the *Extra* model led to some discrepancies between known SOCS distribution and the modelled one. For instance, the model predicted low SOCS in the Landes region (south west of France), most probably because of low clay contents, whereas acid forest soils in this region are known for exhibiting higher SOCS values, between 8 and 14 kg/m^2 (Jolivet et al., 2003).

Land management and agricultural practices influence on SOCS has been and still is currently widely studied and its role might be in some cases underestimated (Bell and Worrall, 2009). It is well established that some specific practices, such as organic matter addition (Lashermes et al., 2009), reduced tillage practices (Metay et al., 2009) or crop residues management and permanent cover crops (Rice, 2006), may influence the SOC inputs and its fate in agricultural soils. Not speaking about specific agricultural practices, including information about detailed land use showed to be valuable for explaining observed SOCS: the VIM value of the *lu3* variable greatly outperformed those of the *lu2* and *lu1* variables, which are less informative about the land use. The inclusion of the *lu3* variable in the model used for estimating SOCS at the national scale was out of concern simply because spatial information with this level of detail was out of reach. Obtaining such an information is needed in order to refine our estimate of the spatial distribution of SOCS in French soils. Similarly, it could support detailed implementation of future land use changes and its consequences in terms of SOCS dynamics.

5 Conclusions

We gave in this paper a new estimate for the spatial distribution of the first 30 cm SOCS for France, based on the French monitoring network (RMQS). The total estimate is 3.260 ± 0.872 PgC. It was compared to another estimate based on the previously published European *octop* bulk density maps. This later estimate was 5.303 PgC, consistent with the SOCS published by the JRC for European countries, was much higher than the estimate provided in this paper and based on RMQS data. Two elements advocate the preferential use of this later estimate, for instance for supporting future GHG emission studies. First, it relies on a dataset provided by a sampling scheme ensuring an efficient treatment of the spatial variability of SOC, both locally (through composite sampling) and of over a larger extend (through the use of a regular 16 × 16 km^2 grid). The RMQS sampling protocol is also one of the few, at the European level, providing bulk densities. This avoids the use of pedo-transfer function for estimating it and the resulting uncertainties associated to them (Liebens and VanMolle, 2003). Second, the proposed model relied on the use of BRT which has been confirmed here as being a robust tools for predicting SOCS. While offering a good predictive performance, it enabled quantification of relationships between SOCS and pedo-climatic variables (plus their interactions) over the French territory. These relationship strongly depended on the land use, and more specifically differed between forest soils and cultivated soil. Along with land use, the clay content of soils was the most driving variable of SOCS. Besides the improvement of the model by including more predictors, the refinement of spatial data layers, regarding soil and land use will be a critical step for improving the SOCS assessments at the country level.

BGD
7, 8409–8443, 2010

Acknowledgements. The sampling and soil analyses were supported by a French Scientific Group of Interest on soils: the GIS Sol, involving the French Ministry for Ecology, Energy, Sustainable Development and Land Use (MEEDDAT), the French Ministry of Agriculture (MAP), the French Agency for Environment and Energy Management (ADEME), the Institute for Research and Development (IRD), the National Forest Inventory (IFN) and the National Institute for Agronomic Research (INRA). The authors thank all the soil surveyors and technical assistants involved in sampling the sites. Special thanks are addressed to the technical assistants from the National French Soil Bank for sample handling and preparation.

References

Arrouays, D., Deslais, W., and Badeau, V.: The carbon content of topsoil and its geographical distribution in France, Soil Use Manage., 17, 7–11, 2001. 8412, 8423

Arrouays, D., Jolivet, C., Boulonne, L., Bodineau, G., Saby, N., and Grolleau, E.: A new initiative in France: a multi-institutional soil quality monitoring network, Comptes rendus de l'Academie d'Agriculture de France, 88, 93–105, 2002. 8412

Arrouays, D., Saby, N., Walter, C., Lemercier, B., and Schvartz, C.: Relationships between particle-size distribution and organic carbon in French arable topsoils, Soil Use Manage., 22, 48–51, 2006. 8426

Batjes, N. H.: Total carbon and nitrogen in the soils of the world, European J. Soil Sci., 47, 151–163, 1996. 8411

Bauer, J., Herbst, M., Huisman, J. A., Weihermuller, L., and Vereecken, H.: Sensitivity of simulated soil heterotrophic respiration to temperature and moisture reduction functions, Geoderma, 145, 17–27, 2008. 8427

Bell, M. J. and Worrall, F.: Estimating a region's soil organic carbon baseline: the undervalued role of land-management, Geoderma, 152, 2009. 8429

Bellamy, P. H., Loveland, P. J., Bradley, R. I., Lark, R. M., and Kirk, G. J. D.: Carbon losses from all soils across England and Wales 1978–2003, Nature, 437, 245–248, 2005. 8411

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J.: Classification and regression trees, Wadsworth, Inc. Monterey, Calif., U.S.A., 368 pp., 1984. 8416

Chakir, R. and Parent, O.: Determinants of land use changes: A spatial multinomial probit approach, Pap. Reg. Sci., 88, 327–344, 2009. 8415

Chaplot, V., Bouahom, B., and Valentin, C.: Soil organic carbon stocks in Laos: spatial variations and controlling factors, Glob. Change Biol., 16, 1380–1393, 2009. 8412, 8426, 8428

Coleman, K., Jenkinson, D. S., Crocker, G. J., Grace, P. R., Klir, J., Korschens, M., Poulton, P. R., and Richter, D. D.: Simulating trends in soil organic carbon in long-term experiments using RothC-26.3, Geoderma, 81, 29–44, 1997. 8414

Davidson, E. A., Trumbore, S. E., and Amundson, R.: Soil warming and organic carbon content, Nature, 408, 789–790, 2000. 8427

Dendoncker, N., van Wesemael, B., Smith, P., Lettens, S., Roelandt, C., and Rounsevell, M.: Assessing scale effects on modelled soil organic carbon contents as a result of land use change in Belgium, Soil Use Manage., 24, 8–18, 2008. 8423

Elith, J., Leathwick, J. R., and Hastie, T.: A working guide to boosted regression trees, J. Animal Ecol., 77, 802–813, 2008. 8416

Eswaran, H., Vandenberg, E., and Reich, P.: Organic-carbon in soils of the world, Soil Sci. Soc. Am. J., 57, 192–194, 1993. 8411

Feranec, J., Jaffrain, G., Soukup, T., and Hazeu, G.: Determining changes and flows in European landscapes 1990–2000 using Corine land cover data, Applied Geography, 30, 19–35, 2010. 8413

Freund, Y. and Schapire, R. E.: Experiments with a New Boosting Algorithm, in: Machine Learning: Proceedings of the Thirteenth International Conference, 148–156, Morgan Kauffman, San Francisco., 1996. 8416

Friedman, J. H.: Greedy function approximation: A gradient boosting machine, Ann. Stat., 29, 1189–1232, 2001. 8416

Friedman, J. H. and Meulman, J. J.: Multiple additive regression trees with application in epidemiology, Stat. Med., 22, 1365–1381, 2003. 8417

Grimm, R., Behrens, T., Marker, M., and Elsenbeer, H.: Soil organic carbon concentrations and stocks on Barro Colorado Island – Digital soil mapping using Random Forests analysis, Geoderma, 146, 102–113, 2008. 8428, 8429

Hastie, T., Tibshirani, R., and Friedman, J.: The Elements of Statistical Learning, Data Mining, Inference, and Prediction, Second Edition, Springer Series in Statistics, 746 pp., 2001. 8417

Heimann, M. and Reichstein, M.: Terrestrial ecosystem carbon dynamics and climate feedbacks, Nature, 451, 289–292, 2008. 8427

Hiederer, R.: Organic carbon per country, http://eusoils.jrc.ec.europa.eu/ESDB_Archive/octop/Resources/%OC_Per_Country.pdf, last accessed: 11/5/2010, 2010. 8411, 8412, 8423

Hiederer, R., Jones, R. J. A., and Montanarella, L.: Topsoil Organic Carbon Content in Europe, Special publication No. SP.I.04.72, map in ISO B1 format. European Communities, Joint Research Centre, Ispra, Italy, 2004. 8425

Isss-Isric-Fao: World Reference Base for Soil Resources, World Soil Resources Reports 84, Tech. rep., Fao, Rome, 91 pp., 1998. 8415

Jolivet, C., Arrouays, D., Leveque, J., Andreux, F., and Chenu, C.: Organic carbon dynamics in soil particle-size separates of sandy Spodosols when forest is cleared for maize cropping, European J. Soil Sci., 54, 257–268, 2003. 8429

Jones, R. J. A., Hiederer, R., Rusco, E., and Montanarella, L.: Estimating organic carbon in the soils of Europe for policy support, European J. Soil Sci., 56, 655–671, 2005. 8411, 8423

Krogh, L., Noergaard, A., Hermansen, M., Greve, M. H., Balstroem, T., and Breuning-Madsen, H.: Preliminary estimates of contemporary soil organic carbon stocks in Denmark using multiple datasets and four scaling-up methods, Agr. Ecosyst. Environ., 96, 19–28, 2003. 8411

Lal, R.: Soil carbon sequestration to mitigate climate change, Geoderma, 123, 1–22, 2004. 8411

Lashermes, G., Nicolardot, B., Parnaudeau, V., Thuries, L., Chaussod, R., Guillotin, M. L., Lineres, M., Mary, B., Metzger, L., Morvan, T., Tricaud, A., Villette, C., and Houot, S.: Indicator of potential residual carbon in soils after exogenous organic matter application, Eur. J. Soil. Sci., 60, 297–310, 2009. 8429

Lawrence, R., Bunn, A., Powell, S., and Zambon, M.: Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis, Remote Sens. Environ., 90, 331–336, 2004. 8416

Liao, Q. L., Zhang, X. H., Li, Z. P., Pan, G. X., Smith, P., Jin, Y., and Wu, X. M.: Increase in soil organic carbon stock over the last two decades in China's Jiangsu Province, Glob. Change Biol., 15, 861–875, 2009. 8426

Liebens, J. and VanMolle, M.: Influence of estimation procedure on soil organic carbon stock assessment in Flanders, Belgium, Soil Use Manage., 19, 364–371, 2003. 8430

Lilly, A., Nemes, A., Rawls, W. J., and Pachepsky, Y. A.: Probabilistic approach to the identification of input variables to estimate hydraulic conductivity, Soil Sci. Soc. Am. J., 72, 16–24, 2008. 8419

Lo Seen, D., Ramesh, B. R., Nair, K. M., Martin, M., Arrouays, D., and Bourgeon, G.: Soil carbon stocks, deforestation and land-cover changes in the Western Ghats biodiversity hotspot

BGD

7, 8409–8443, 2010

Spatial distribution of soil organic carbon stocks in France

M. P. Martin et al.

(India), Glob. Change Biol., 16, 1777–1792, 2010. 8412
Martin, M. P., Seen, D. I. o., Boulonne, L., Jolivet, C., Nair, K. M., Bourgeon, G., and Arrouays, D.: Optimizing pedotransfer functions for estimating soil bulk density using boosted regression trees, Soil Sci. Soc. Am. J., 73, 485–493, 2009. 8413, 8417, 8420, 8428
Meersmans, J., De Ridder, F., Canters, F., De Baets, S., and Van Molle, M.: A multiple regression approach to assess the spatial distribution of Soil Organic Carbon (Soc) at the regional scale (Flanders, Belgium), Geoderma, 143, 1–13, 2008. 8427
Metay, A., Mary, B., Arrouays, D., Labreuche, J., Martin, M., Nicolardot, B., and Germon, J. C.: Effects of reduced or no tillage practices on C sequestration in soils in temperate regions, Can. J. Soil. Sci., 89, 623–634, 2009. 8429
Mishra, U., Lal, R., Slater, B., Calhoun, F., Liu, D. S., and Van Meirvenne, M.: Predicting Soil Organic Carbon Stock Using Profile Depth Distribution Functions and Ordinary Kriging, Soil Sci. Soc. Am. J., 73, 614–621, 2009. 8425
Morvan, X., Saby, N. P. A., Arrouays, D., Le Bas, C., Jones, R. J. A., Verheijen, F. G. A., Bellamy, P. H., Stephens, M., and Kibblewhite, M. G.: Soil monitoring in Europe: A review of existing systems and requirements for harmonisation, Sci. Total Environ., 391, 1–12, 2008. 8411
Paustian, K., Andren, O., Janzen, H. H., Lal, R., Smith, P., Tian, G., Tiessen, H., Van Noordwijk, M., and Woomer, P. L.: Agricultural soils as a sink to mitigate CO2 emissions, Soil Use Manage., 13, 230–244, 1997. 8411
Phachomphon, K., Dlamini, P., and Chaplot, V.: Estimating carbon stocks at a regional level using soil information and easily accessible auxiliary variables, Geoderma, 155, 372–380, 2010. 8425
Post, W. M., Emanuel, W. R., Zinke, P. J., and Stangenberger, A. G.: Soil carbon pools and world life zones, Nature, 298, 156–159, 1982. 8411
Quintana-Segui, P., Le Moigne, P., Durand, Y., Martin, E., Habets, F., Baillon, M., Canellas, C., Franchisteguy, L., and Morel, S.: Analysis of near-surface atmospheric variables: Validation of the Safran analysis over France, J. Appl. Meteorol. Clim., 47, 92–107, 2008. 8414
Rice, C. W.: Introduction to special section on greenhouse gases and carbon sequestration in agriculture and forestry, J. Environ. Qual., 35, 1338–1340, 2006. 8429
Ridgeway, G.: gbm: Geberalized Boosted regression Models. R package version 1.5-7., 2006. 8417, 8420
Running, S. W., Nemani, R. R., Heinsch, F. A., Zhao, M. S., Reeves, M., and Hashimoto, H.: A continuous satellite-derived measure of global terrestrial primary production, Bioscience,

BGD
7, 8409–8443, 2010

54, 547–560, 2004. 8413
Saby, N. P. A., Arrouays, D., Antoni, V., Lemercier, B., Follain, S., Walter, C., and Schvartz, C.: Changes in soil organic carbon in a mountainous French region, 1990-2004, Soil Use Manage., 24, 254–262, 2008a. 8411
Saby, N. P. A., Bellamy, P. H., Morvan, X., Arrouays, D., Jones, R. J. A., Verheijen, F. G. A., Kibblewhite, M. G., Verdoodt, A., Uveges, J. B., Freudenschuss, A., and Simota, C.: Will European soil-monitoring networks be able to detect changes in topsoil organic carbon content?, Glob. Change Biol., 14, 2432–2442, 2008b. 8411, 8417, 8425
Schlesinger, W. H.: Biogeochemistry: an analysis of global change, Academic Press, San Diego Usa, 443 pp., 1991. 8411
Schnebelen, N., Nicoullaud, B., Bourennane, H., Couturier, A., Verbeque, B., Revalier, C., Bruand, A., and Ledoux, E.: The Stics model to predict nitrate leaching following agricultural practices, Agronomie, 24, 423–435, 2004. 8418
Smith, J., Smith, P., Wattenbach, M., Zaehle, S., Hiederer, R., Jones, R. J. A., Montanarella, L., Rounsevell, M. D. A., Reginster, I., and Ewert, F.: Projected changes in mineral soil carbon of European croplands and grasslands, 1990–2080, Glob. Change Biol., 11, 2141–2152, 2005. 8423, 8425
Torn, M. S., Trumbore, S. E., Chadwick, O. A., Vitousek, P. M., and Hendricks, D. M.: Mineral control of soil organic carbon storage and turnover, Nature, 389, 170–173, 1997. 8429
Wosten, J. H. M., Lilly, A., Nemes, A., and Le Bas, C.: Development and use of a database of hydraulic properties of European soils, Geoderma, 90, 169–185, 1999. 8426
Yang, Y. H., Fang, J. Y., Tang, Y. H., Ji, C. J., Zheng, C. Y., He, J. S., and Zhu, B. A.: Storage, patterns and controls of soil organic carbon in the Tibetan grasslands, Glob. Change Biol., 14, 1592–1599, 2008. 8425
Yokozawa, M., Shirato, Y., Sakamoto, T., Yonemura, S., Nakai, M., and Ohkura, T.: Use of the RothC model to estimate the carbon sequestration potential of organic matter application in Japanese arable soils, Soil Sci. Plant Nutr., 56, 168–176, 2010. 8414
Zaehle, S., Bondeau, A., Carter, T. R., Cramer, W., Erhard, M., Prentice, I. C., Reginster, I., Rounsevell, M. D. A., Sitch, S., Smith, B., Smith, P. C., and Sykes, M.: Projected changes in terrestrial carbon storage in Europe under climate and land-use change, 1990–2100, Ecosystems, 10, 380–401, 2007. 8425

BGD
7, 8409–8443, 2010

Table 1. Fit and cross validation results for a ratio of 0.9/0.1 training vs. validation datasets. Quality of the fit on the full data set is expressed using R^2, mean prediction error (MPE), standard deviation of the prediction error (SDPE), and root mean square prediction error (RMSPE). The cross-validation results are expressed using $<R^2{}_{ext}>$, $<MPE_{ext}>$, $<SDPE_{ext}>$ and $<RMSPE_{ext}>$ estimated using the validation datasets. The 95% confidence intervals obtained for the corresponding normal distributions using the standard percentile method are given in brackets.

Model	R^2	MPE	SDPE	RMSPE	$<R^2{}_{ext}>$		$<MPE_{ext}>$		$<SDPE_{ext}>$		$<RMSPE_{ext}>$	
Cult	0.91	−0.001	0.935	0.934	0.58	[0.445, 0.723]	−0.041	[−0.379, 0.297]	1.94	[1.481, 2.397]	1.94	[1.486, 2.395]
F	0.74	2e-04	1.912	1.910	0.36	[0.141, 0.57]	−0.009	[−0.845, 0.827]	2.75	[2.036, 3.467]	2.76	[2.053, 3.459]
Extra	0.73	−0.001	1.727	1.727	0.5	[0.386, 0.613]	−0.002	[−0.348, 0.344]	2.27	[1.86, 2.68]	2.27	[1.862, 2.68]

BGD
7, 8409–8443, 2010

Spatial distribution of soil organic carbon stocks in France

M. P. Martin et al.

Title Page
Abstract Introduction
Conclusions References
Tables Figures
Back Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

Table 2. Relative influences of the predictors for each model, expressed as variable importance indexes (VIM), and rank according to the VIM values. The predictors are grouped, starting with the variables related to land use, then related to the climatic or pedo-climatic factors, then to plant productivity and finally related to the soil properties only.

	Cult model		*F* model		*Extra* model	
Predictor	VIM	rank	VIM	rank	VIM	rank
lu3	33.66	1	0.77	11	–	–
lu2	1.26	13	0.00	14	–	–
lu1	0.11	15	–	–	–	–
lu_ipcc	0	16	–	–	26.83	2
a	7.1	3	1.47	10	8.76	4
b	3.72	7	4.83	7	6.53	6
rain	6.6	4	13.27	3	10.66	3
pet	3.3	8	4.4	8	5.73	7
temp	3.03	9	1.83	9	6.77	5
npp	2.89	10	6.54	5	5.33	8
wlogging	1.34	12	0.06	12	–	–
wregime	1.14	14	0.03	13	–	–
ce	6.08	5	8	4	–	–
clay	22.55	2	29.55	1	29.4	1
silt	1.96	11	5.91	6	–	–
ph	5.26	6	23.35	2	–	–

BGD
7, 8409–8443, 2010

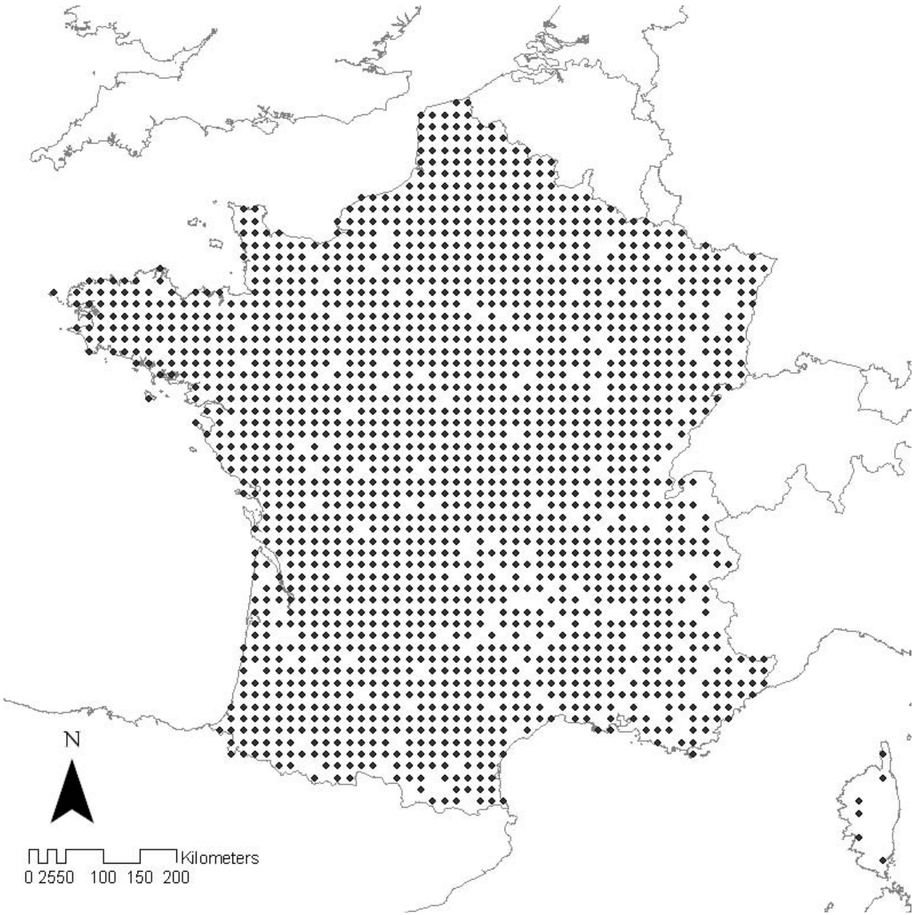

Fig. 1. Distribution of the 1974 sites within the French monitoring network which were used in the present study.

BGD
7, 8409–8443, 2010

Spatial distribution of soil organic carbon stocks in France

M. P. Martin et al.

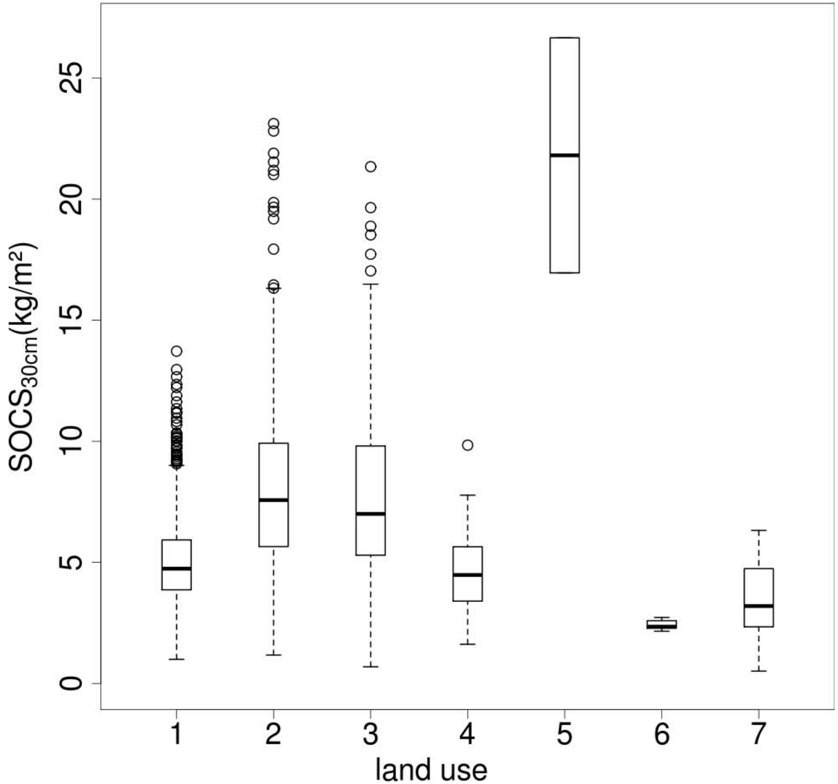

Fig. 2. SOCS for the first 30 cm as a function of land cover type according to the adapted IPCC land use classification (various crops (1, $n = 817$), permanent grasslands (2, $n = 463$), woodlands (3, $n = 468$) orchards and shrubby perennial crops (4, $n = 18$), wetlands (5, $n = 2$), others (6, $n = 5$), vineyards (7, $n = 32$)).

BGD
7, 8409–8443, 2010

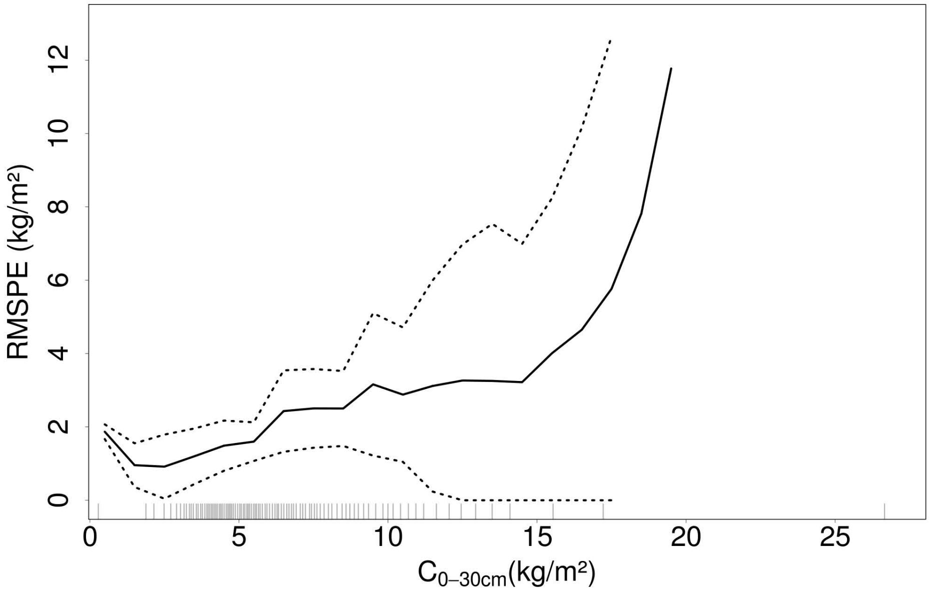

Fig. 3. Uncertainty of the *Extra* model, as a function of the organic carbon stock (30 cm). Uncertainty values are calculated as <$RMSPE_{ext}$> resulting from cross-validations trials as a function of predicted SOCS, grouped within intervals of 1 kg/m² width, from 0 to 30 kg/m². The solid line represents the mean of uncertainty within each interval of SOCS values, and the upper and lower dashed lines represent the bounds of the $CI_{95\%}$ assuming a normal distribution within each interval. Tick marks at the lower border of the diagram give the 1% quantiles for the RMQS dataset.

BGD
7, 8409–8443, 2010

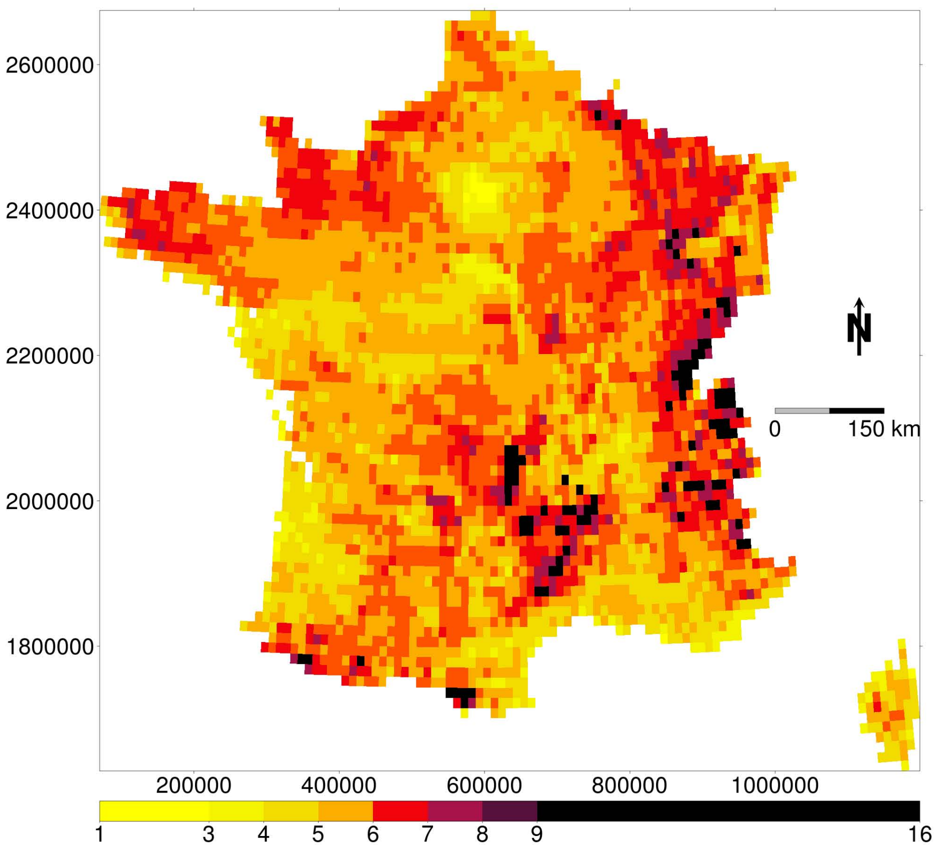

Fig. 4. Map of the soil organic carbon for the first 30 cm (kg/m²).

BGD
7, 8409–8443, 2010

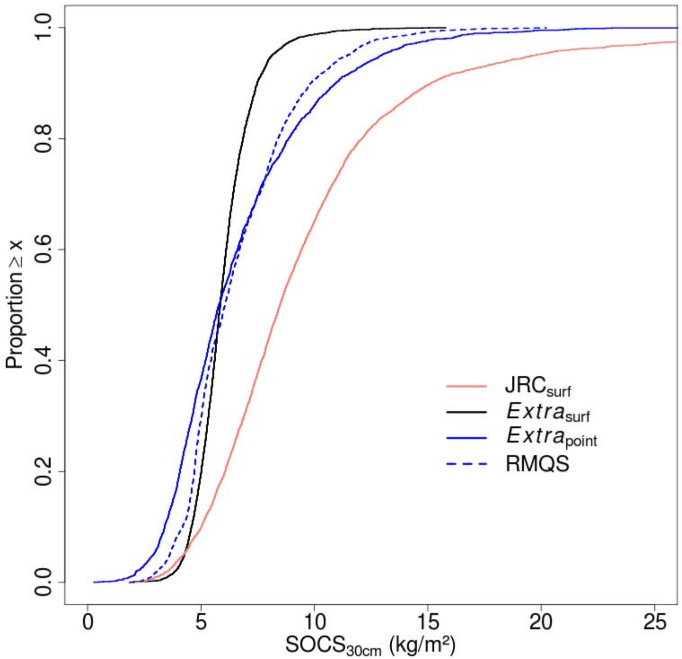

Fig. 5. Empirical cumulative distribution functions (ecdf) for the two spatial estimates presented in this paper (using the *Extra* model and the JRC estimate) as well as for the observed (curve RMQS) and predicted (curve $Extra_{point}$) SOCS at RMQS sites. Computing ecdf on spatial estimates is done as follows: first the statistical population is made of each spatial unit where the prediction model has been applied (the *Extra* model for instance). Second, a weight is computed for each unit as the ratio between its area and the sum of spatial units area (here, the area of France). Third, the ecdf is estimated on models predictions within the spatial units (kg/m²) using weights previously calculated. Ecdfs of site observed or predicted values are calculated using equal weights between individuals.

BGD
7, 8409–8443, 2010

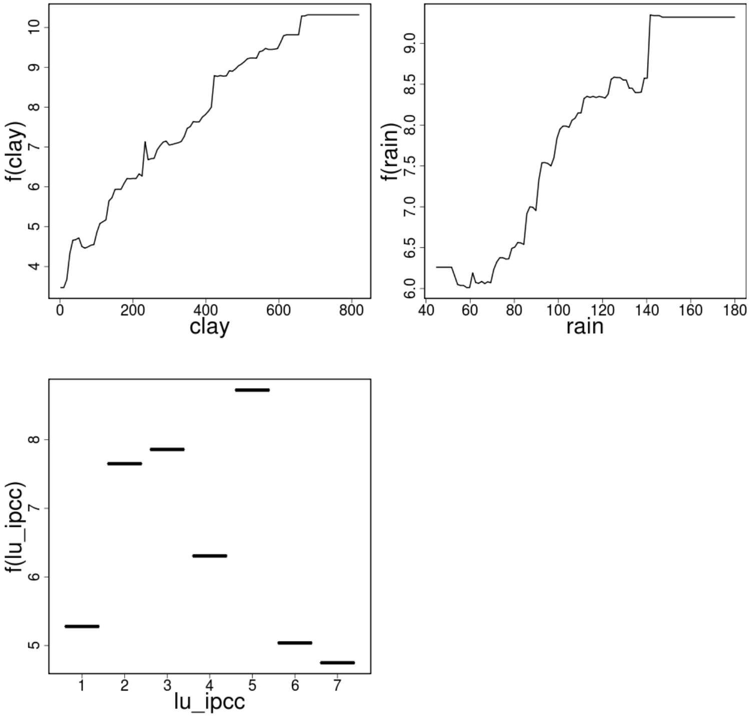

Fig. 6. Effect of the three most important variables in the *Extra* model (i.e. *clay*, *rain* and *lu_ipcc*). The lower left diagram gives the modelled relationship between SOCS and land use (coded using the adapted ipcc classification: croplands (1), permanent grasslands (2), woodlands (3) orchards, shrubby perennial crops (4), wetlands (5), others (6), vineyards (7)).