Quantifying and mapping topsoil inorganic carbon concentrations and stocks: approaches tested in France.

B.P. Marchant\textsuperscript{1,2*,} E.J. Villanneau\textsuperscript{3}, D. Arrouays\textsuperscript{3}, N.P.A. Saby\textsuperscript{3} & B.G. Rawlins\textsuperscript{1}

\textsuperscript{1} British Geological Survey, Environmental Science Centre, Keyworth, Nottingham, NG12 5GG, UK.

\textsuperscript{2} Rothamsted Research, West Common, Harpenden, AL5 2JQ, UK.

\textsuperscript{3} INRA, US 1106 InfoSol, F-4075 Orléans, France.

\*Corresponding author: Email benmarch@bgs.ac.uk; Tel. +44 1159 363100

Abstract

The potential for soils to act as a sink or a source of atmospheric carbon has long been recognised and great efforts are made to monitor soil organic carbon stocks. Inorganic carbon is also exchanged between soils and the atmosphere and is important for soil function but inorganic carbon stocks are not measured in many national- and continental-scale soil monitoring networks. Topsoil (0–30 cm) inorganic carbon concentrations were measured at more than 2 000 sites on a regular 16-km grid as part of the French National Soil Monitoring Network (RMQS: Réseau de Mesures de la Qualité des Sols). We used design-based statistical methods to calculate unbiased estimates of the mean soil inorganic carbon concentration and total stocks across France and we used model-based methods to determine the uncertainty of these estimates and to predict the spatial distribution of these quantities. The observations of inorganic carbon were highly skewed and did not conform to standard statistical models. We therefore applied a non-parametric transform to normalise the data and performed validation of the model that resulted. The estimates and predictions of inorganic carbon are baselines against which the results of future phases of the network can be compared. We found that the total topsoil inorganic carbon stocks in France amount to 1070 ±61 Tg (Tg = terragram = 10\textsuperscript{12} g) which is approximately one third of the mass of the corresponding organic carbon stocks.
Their spatial distribution was strongly linked to the underlying geology. We note that in countries where inorganic carbon is not measured within soil monitoring networks, its concentrations and stocks might be estimated from other sources of soil legacy information such as databases of soil tests requested by farmers. We tested the reliability of this approach by considering the French Soil Test Database which contains the results of 280,000 soil analyses requested by farmers between 1990 and 2004. We found that these data led to a biased estimate of soil inorganic carbon concentrations which was presumably the result of the locations of the soil samples being selected according to the concerns of farmers rather than a statistical design.

Introduction

The global stocks of soil carbon exceed the combined stocks of carbon contained in the atmosphere and above-ground biomass (Lal, 2004). The soil carbon reservoir is dynamic and the potential for it to act as a sink or a source of atmospheric carbon dioxide has been long recognised. Lal (2004) estimated that the global potential of soil organic carbon sequestration was 3.3 Pg C/year (Pg = petagram = $10^{15}$ g). Soil carbon stocks must be monitored to determine whether this sequestration potential is being realized or alternatively whether soil carbon is being lost to the atmosphere. This is one of the reasons why soil organic carbon has become the primary focus of many broad-scale soil monitoring networks (SMNs; Black et al., 2008).

Batjes (1996) estimated that one third of the global soil carbon stocks contained in the upper metre of soil are in an inorganic form. These stocks can also act as either a sink or a source of atmospheric carbon. The reaction of calcium and magnesium with atmospheric carbon dioxide and water in the upper soil horizons can lead to the formation of secondary carbonates and the sequestration of atmospheric carbon (Sahrawat, 2003). Manning (2008) and Washbourne et al. (2012) have suggested that further stocks of inorganic carbon could be sequestered through the incorporation into the soil of waste calcium silicate minerals from quarrying, demolition and steel manufacture. Suarez (2006) noted that land management practices such as irrigation and use of acidifying
fertilizers can cause losses of soil inorganic carbon although inorganic carbon can be added to the soil upon liming. There is a further need to monitor inorganic carbon where it supports soil function. For example, in many soil types the dissolution of soil carbonate is the dominant buffering mechanism which limits the soil acidification which can otherwise occur when nitrogen fertilisers are added to the soil.

Inorganic carbon is excluded from many broad-scale SMNs such as the FOREGS geochemical survey (Salminen et al., 2005) and the Soil Profile Analytical Database of Europe (Hannam et al., 2009). According to the ENVASSO European survey (Arrouays et al., 2008) inorganic carbon was monitored in only six of the 27 members of the European Community. Instead, SMNs tend to focus on soil properties that are related to soil fertility and are expected to evolve rapidly (e.g. pH, organic carbon, and soil nutrients) or on soil properties that are of environmental concern (e.g. heavy metals and organic pollutants). However, in a review of field evidence and the results of mathematical models, Sanderman (2012) found that inorganic carbon stocks in Australian agricultural soils could change by as much as 1 t/ha/year. Therefore, measurements of inorganic carbon were included in the protocols for the Australian National Soil Carbon Research Project (Sanderman et al., 2011).

Previously reported broad-scale estimates of soil inorganic carbon stocks and concentrations have been based on existing soil databases and information systems that were not constructed according to a statistical design or with quantitative soil monitoring in mind. Expert judgement has been used in the establishment of these databases to divide the study region into different soil classes and to locate representative soil pits within these classes. For example, Wu et al. (2009) used 2 553 soil profiles from the second national Chinese soil survey to determine inorganic carbon stocks in the A, B and C horizons for each subgroup of the Chinese Soil Taxonomy. They upscaling these results to a national stock of 55.3±10.7 Pg or 6.3±1.2 kg/m². Similarly, Guo et al. (2006) used the State Soil Geographic Database to estimate the stock of inorganic carbon in the top two metres of soil in the contiguous USA to be between 22.6 and 93.7 Pg. Lettens et al. (2004) divided Belgium into 492
landscape units based upon land cover and soil association and estimated mean inorganic carbon stocks for each landscape unit using 13 000 geo-referenced soil profile descriptions. They found that the inorganic stock in the top 30 cm was 4 Tg or 0.54 kg/m³. The uncertainty of the estimate was not reported. The maps and estimates which result from these databases are informative but it is challenging to account for the expert judgements when the uncertainty of the outputs is quantified.

Other studies have inferred inorganic carbon estimates by fitting simple relationships with other soil properties and environmental covariates. In England, Rawlins et al. (2011) used a regression model of calcium and aluminium concentrations to estimate inorganic carbon stocks in the top 30 cm of soil. They found that the total inorganic stocks contained in soils over two groups of primary, carbonate-bearing parent materials across parts of southern and eastern England amounted to 186 Tg (Tg = terragram = 10¹² g) or 15% of the total (inorganic plus organic) topsoil carbon stocks in England. The regression models were calibrated upon 100 measurements of inorganic carbon concentrations in the top 15 cm of soil. These measurements ranged from 7 to 70 g/kg (mean 44 g/kg) on a Cretaceous chalk deposit and 1 to 60 g/kg (mean of 36 g/kg) on a Jurassic limestone deposit. Given the relative complexity of their statistical procedure, Rawlins et al. (2011) did not quantify the uncertainty of their estimates. In China, Na et al. (2008) estimated relationships between inorganic carbon stocks, climatic parameters and land use information. The total stock of inorganic carbon in China to a depth of 2m was estimated to be 53.3±6.3 Pg. Such an approach can make very efficient use of the available data but it does rely on the assumption that the same simple relationships hold across the study region.

In contrast, purpose-built SMNs use a statistical design to ensure unbiased estimates of both soil status and soil change. The French national SMN (RMQS – Réseau de Mesures de la Qualité des Sols) included over 2 000 observations of the concentration of soil carbonates and stocks of soil inorganic carbon in the top 30 cm of soil across France (Fig. 1). We used geostatistical techniques to map the inorganic carbon stocks across France and to determine the total mass of inorganic carbon stored in
the top 30 cm of soil. We used knowledge of the parent material across France (Fig. 2) to improve our predictions. Such geostatistical techniques require that a valid model of the spatial variation of inorganic carbon is fitted to the spatial data. We performed a design-based analysis (Brus & De Gruijter, 1997) to validate our prediction of the total stocks. Design-based analyses require much less stringent assumptions about the statistical distribution of the data. We also explored whether the soil inorganic carbon concentrations could have been estimated from sources of soils data that might be available in other countries. In particular, we considered the results of 268,000 soil analyses that have been requested by farmers in France between 1990 and 2004.

Statistical Theory

Design of soil monitoring networks

Arrouays et al. (2012) reported that purpose-built SMNs tend to either employ a systematic design such as a grid or to use a probabilistic design. In probabilistic designs, sampling locations are selected independently so there is no need to account for correlations between observations. Mean values of soil properties across a region and corresponding estimation variances can be estimated by classical design-based statistical methods that require few assumptions about the variation of the property. When a systematic design is applied, design-based methods can be used to estimate the mean of the property but no unbiased design-based estimator of the variance is available. Instead a model-based analysis is required to account for the correlation between observations when the estimation variance is determined. Model-based analyses can also be used to predict maps of properties. The predictions and prediction variances that result from these analyses are only valid if the model appropriately represents the spatial correlation of the property.

Mean values and maps of properties can be determined from legacy datasets that were not collected for the purpose of monitoring. For example, Orton et al. (2012) have estimated changes in soil carbon concentrations in French soils using measurements requested by farmers to assess the
fertility of their soils. Generally, the locations of these measurements are not selected according to a statistical design. Design-based analyses can only be applied if the locations are selected probabilistically. Model-based methods can be applied if the sample selection is not biased according to the expected value of the measurement. However, biases often occur in such surveys. For example, a farmer might bias their measurements towards areas where they are concerned about soil fertility. Such preferential sampling must be accounted for when using the measurements to predict soil properties.

**Model-based analyses**

A model-based analysis of a survey assumes that the observed data are a realization of a random function. A statistical model that describes the random function, including any spatial correlation, must be fitted to the data prior to prediction and the production of maps. The statistical model is a simplification of reality and its use requires that assumptions are made about the random function from which the observed data are realized. For example, the random function is often assumed to have a multivariate Gaussian distribution and to be second order stationary (Webster & Oliver, 2007).

In some cases the assumptions are implausible and must be relaxed. For example, the histogram of inorganic carbon stocks observed in RMQS (Fig. 3) is highly skewed. Therefore, the histogram or marginal distribution is inconsistent with a multivariate Gaussian distribution and a transformation must be applied to the data prior to analysis. Once the observations have been transformed to an approximate multivariate Gaussian distribution they can be represented by a linear mixed model (LMM; Lark *et al*., 2006). This model assumes that the transformed data $z^*$ are realizations of a random function $Z^*$. The model splits the variation into two components or effects. Fixed effects are linear relationships between the variable of interest and relevant covariates. Random effects are spatially correlated residual terms. The spatial correlation is described by a parametric variogram function, the parameters of which are fitted to the data by either maximum likelihood (ML) or
residual maximum likelihood (REML; see Lark et al., 2006). The Matérn model is a flexible variogram function that is often advocated in studies of soil properties (e.g. Marchant & Lark, 2007).

The appropriateness of fitted models with different fixed effects can be compared by calculating the Akaike information criterion (AIC; Akaike, 1973):

$$AIC = -2L + 2q,$$

for each model where $L$ is the ML log-likelihood and $q$ is the number of parameters in the model. The model with the smallest AIC best manages the trade-off between quality of fit and model complexity (Webster & Oliver, 2007).

There is a small bias in the variance parameters of an LMM which are estimated by ML and therefore REML is often advocated (e.g. Lark et al., 2006) since this estimator adjusts the expression for the likelihood to minimize this bias. However, the AIC cannot be calculated from these adjusted likelihoods. Once the LMM has been estimated it is possible to krige (Webster & Oliver, 2007) predictions of the mean ($m_j$) and the kriging variance ($\sigma_j^2$) of $Z^*$ at each node, $j$, of a grid where predictions are required. The predictions of the transformed property must then be back-transformed to the units of the measured observations.

Standard bootstrapping or resampling methods (Efron & Tibshirani, 1993) cannot be used to validate a spatial LMM because these methods do not account for the spatial correlation amongst the data. An LMM can be validated by leave-one-out cross validation. Here each observation, $z^*(x_i)$, is removed in turn from the dataset and the remaining observations are used to predict the (transformed) property at the site of the removed observation. The standardized prediction errors

$$\theta_i = \frac{(z^*(x_i) - m_i)^2}{\sigma_i^2},$$

are calculated. If the LMM is accurate, the expected square error should be equal to the kriging variance, and the mean of $\theta$ should be 1 and the median equal to 0.45.
The mean of a property across the entire prediction grid can be estimated by using the multivariate LU simulation method (Deutsch & Journel, 1998) to produce a large number (e.g. 1000) of realizations of the transformed property, conditional on the transformed observed data, at all of the nodes of the prediction grid. Each simulated value is then back-transformed and the mean of each realization is calculated. The predicted mean is equal to the mean of the realization means. The prediction variance is equal to the variance of the prediction means.

**Design-based analyses**

The model-based methods described above regard the observations of a property as realizations of a random function whereas the sampling locations are treated as fixed. The model-based uncertainty analysis considers the variation that will arise from sampling different realizations of the random function. In contrast, design-based statistical methods regard the values of the property of interest across the study region as fixed and the sampling locations are treated as random. Design-based uncertainty analysis considers the variation that will arise from repeatedly estimating the mean value of the property across the study region from different randomly selected sets of sampling locations. If the \( n \) sampling locations \( x \) are selected purely at random such that there is an equal probability of each site being included then the mean of \( z \) across the study region is estimated by

\[
\hat{z} = \frac{1}{n} \sum_{i=1}^{n} z(x_i). \tag{3}
\]

This estimate is said to be \( p \)-unbiased because

\[
E_p[\hat{z}] = \bar{z}, \tag{4}
\]

where \( E_p \) is the statistical expectation over repeated sampling under design \( p \) and \( \bar{z} \) is the actual mean of \( z \). The expected squared estimation error is

\[
r = E_p[\hat{z} - \bar{z}]^2 = \frac{\sigma^2}{n} \approx \frac{s^2}{n}, \tag{5}
\]
where

\[ s^2 = \frac{1}{n-1} \sum_{i=1}^{n} \{ z(x_i) - \hat{z} \}^2, \]

is the estimated variance of \( z \). A major advantage of design-based inference over model-based is that the mean and the expected error can be calculated by such formulae without assuming a particular model of variation. The only required assumption is that \( n \) is sufficiently large such that the Central Limit Theorem can be invoked and the distribution of the repeated estimates of \( \hat{z} \) can be treated as Gaussian (Brus & De Gruijter, 1997).

Design-based analyses have been extended to consider systematic random designs. Such a design follows a fixed pattern but the location of the origin is selected at random. When such a design is used the mean of \( z \) across the study region can still be estimated from Eqn. 3 because each site has an equal probability of being included in the design and the estimator is \( p \)-unbiased. However, if \( z(x) \) is spatially correlated then Eqn. 5 will be a biased estimate of the variance of \( z(x) \) because the pattern of sampling locations is not random. In the case of a regular grid with good coverage of the study region, Eqn. 6 will over-estimate the variance of \( z(x) \) since the smallest lag between pairs of observations is equal to the grid-spacing. Therefore, Eqn. 5 will over-estimate the expected squared estimation error. Domburg et al. (1994) proposed a method by which the expected squared error of \( \hat{z} \) can be estimated for a systematic design using simulated realizations of an estimated model of the random function \( Z(x) \).

**Methods**

*The French National Soil Monitoring Network (RMQS)*

The RMQS database (Arrouays et al., 2002) primarily consists of over 2 000 observations of soil properties on a 16-km regular grid across the 550 000 km\(^2\) French metropolitan territory (Fig. 2a). Twenty six additional observations were positioned within the grid. The locations of these
observations targeted specific combinations of soil type and land use. If it was not possible to 
sample at the selected site (e.g. in urban areas, in rivers or on roads) an alternative cultivated or 
undisturbed location was selected as nearby as possible within a 1-km radius. When this was not 
possible the cell was omitted from the survey. This baseline survey of the RMQS was completed in 
2009. 

At each site, 25 individual core samples were taken of the topsoil (0–30 cm) layer, using an unaligned 
sampling design within a 20 × 20-m area. The 0–30 cm layer was chosen to ensure consistency with 
existing surveys and because in France 30 cm is the maximum depth at which topsoil is affected by 
ploughing (Arrouays et al., 2001). Core samples were bulked to obtain a composite sample for each 
site. 

Preparation and analysis of soil samples 

The inorganic carbon concentration was determined for a 0.5–10 g subsample of each composite 
sample. The size of the subsample was selected according to the expected calcium carbonate 
concentration. The subsamples were dried at 30 °C, sieved to 2 mm and subsequently ground to 250 
µm. The samples were acidified in a closed flask with a solution of 1/5 dilute hydrochloric acid (37% 
v/v). The volume of carbon dioxide released was measured using a Bernard calcimeter (a graduate 
tube filled with 200 mL of water) and was compared to the volume of carbon dioxide produced by 
pure calcium carbonate under the same temperature and pressure conditions. The kinetics of 
decomposition by acid varies according to chemical species; decomposition is rapid for calcite and 
slow for dolomite. The INRA laboratory of Arras has a COFRAC accreditation for calcimetry analysis 
and follows the French standard procedure. The detection limit of this procedure was 1 g/kg. 

The mass of fine earth (i.e. particles which pass through a 2-mm sieve) for each observation site was 
determined for three samples of known volume that were extracted from a soil pit adjacent to the 
site. The methods varied according to the particle size distribution of the samples. The samples were
dried prior to analysis. The cylinder method (procedure reference NF X 31-501 in AFNOR, 1999) was used for soils with little to no gravel. The cylinder was 90 mm high with a diameter of 84 mm, i.e., 500 cm³. When the cylinder method was not possible, i.e. for gravelly or stony soils, the water method was applied. The water method is adapted from the sand method (procedure reference NF X 31-503 in AFNOR, 1999). In this method, circular holes of between 1 000 and 3 000 cm³ are dug into the soil. The exact volume of soil extracted is determined by lining the hole with a plastic bag and measuring the volume of water required to fill the hole.

Soil parent material

Information on soil parent material was derived from the 1:1 000 000 scale soil database of Europe (King et al., 1995). The classes from this database were amalgamated into five broader classes (Table 1 and Fig. 2b) for this study. These classes were created according to their natural geochemical concentration of calcium and magnesium carbonate according to the attribute tables of the database. The parent material classes were marls (covering 3.1% of the area of France), chalk (5.3%), other calcareous rocks (21.2%), dolomitic rocks (1.1%) and non-calcareous rocks (69.3%).

The French Soil Test Database

The results of the soil analyses compiled in the French Soil Test database are a consequence of requests from farmers and landowners for help to improve the management of their crops and pastures (Lemercier et al., 2008). The motivation for each analysis request was not recorded. The analysed samples were taken from topsoil horizons of cultivated fields or fields under pasture. In each sampled field 10–15 samples of the ploughed layer (or the 0–30 cm topsoil layer in the case of pastures) were collected with a hand auger. These samples were bulked and the composite sample was sent to a laboratory certified by the French Ministry of Agriculture. The results of standardized analytical procedures were stored in the national database. The sample coordinates could not be stored in the database because of confidentiality concerns. The most accurate geo-referencing that
could be recorded was the name of the local authority or canton which contained the sampled field.

France is divided into 4 032 cantons which are intended to have approximately equal populations of around 20 000. Hence, urban cantons tend to be small but rural ones can have areas of over 100 km².

Model-based analyses

We used model-based analyses to predict (i) maps of the inorganic carbon concentrations and stocks across France, (ii) the mean inorganic carbon concentration and (iii) the total soil inorganic carbon stocks within France. We also determined the uncertainty of these predictions. The histograms of inorganic carbon stock and concentration observations (Fig. 3) indicated that the distributions of these properties were non-Gaussian. Therefore, we normalized these observations by the rank order transform (Goovaerts, 1997) before estimating an LMM of the transformed properties. The observations that were located away from the nodes of the 16-km grid were included in this analysis since they provided useful information about the variation over small lag distances. We compared two different fixed effects models. The first assumed a constant mean and the second assumed that the mean was constant within each of the five geological classes. We estimated both of these models by ML and calculated the AIC to determine which was the more appropriate. Then the more appropriate model for each property was re-estimated by REML before we predicted the transformed properties on a 2.5-km grid across France by kriging. These predictions were back-transformed and finally the mean inorganic carbon concentration and total stocks at the national-scale were estimated using the multivariate LU simulation approach described in the Theory section.

Design-based analyses

We re-estimated the mean inorganic carbon concentration and the total inorganic carbon stocks using design-based methods. This analysis used only the observations that were located on the nodes of the 16-km grid. We assumed that the origin of the RMQS grid was selected at random and
that Eqn. 3 yielded a \( p \) unbiased estimate of these quantities. We determined the estimation variance using the constant mean LMMs fitted to the observed data and the approach suggested by Domburg et al. (1994).

**Analyses of the Soil Test Database**

The samples contained in the Soil Test Database were selected according to the concerns of farmers rather than a statistical design. We assumed that farmers only requested analyses of inorganic carbon at locations where they thought the soil was alkaline and they expected reasonably large concentrations. Therefore, the samples for which inorganic carbon has been measured was unrepresentative of the distribution of soils across France. We adjusted for this preferential sampling by considering the samples for which analyses of pH but not inorganic carbon had been requested. We assumed that inorganic carbon analyses had not been requested because the concentrations were thought to be small. Therefore, we treated such observations as zero inorganic carbon concentrations provided that the pH was less than 6.5.

The mean value of the inorganic carbon concentrations was calculated for each canton and these values were used to map inorganic carbon concentration across France. The mean inorganic carbon for the whole of France was calculated from the mean of the canton means, weighted according to the area of the canton. Since these data were not gathered according to a statistical design we do not quote a standard error on the estimates. For comparison we also used the estimated LMM of inorganic carbon concentrations to estimate mean values for each canton.

**Results**

**Mapping soil inorganic carbon**

There were substantial differences between the inorganic carbon concentrations observed in each parent material class (Table 1). The AICs for the estimated LMMs of both concentrations and stocks with fixed effects that varied according to parent material were less than the AICs for the constant
fixed effect models. Therefore, we included parent material classes as fixed effects in our model-based analyses. All of the estimated LMMs had mean $\theta$ equal to 1.00 and median $\theta$ close to 0.45 (Table 2) indicating that the residual variation had been adequately modelled.

Figure 4 shows the variograms of the random effects of the variable fixed effect LMM for both the transformed inorganic carbon stocks and concentrations. These plots show how the expected squared difference between observations increases with the distance between the locations of the samples. The variogram sills were close to 0.87 in both cases. This is less than 1.0, the variance of the transformed property, and indicates that a proportion of the variation had been explained by the fixed effects. For both variograms, around 75 % of the total variance occurred over distances of less than 1 km but some spatial correlation was apparent up to distances over 200 km. The corresponding maps of inorganic carbon stocks and concentrations and the prediction of uncertainties are shown in Figure 5. The full pattern of variation is more evident in the maps with nonlinear scale in Figure 6. The prediction variances can be seen to increase with the magnitude of the prediction. The map of inorganic carbon concentrations (Fig. 7a) based upon the Soil Test Database showed a similar spatial pattern to those based upon the RMQS data. However, when the estimated mean concentrations for each canton calculated from the Soil Test Database and RMQS were compared a clear bias was evident (Fig 7b). The Soil Test Database estimate was greater than the RMQS estimate in 65 % of cantons.

*National estimates of inorganic carbon*

The estimates of average inorganic carbon concentrations and total stocks are shown in Table 3. The design-based estimates were 6.46 g/kg and 1070 Tg respectively. Both of the model-based estimates were larger but were approximately one standard deviation from the design-based ones. The estimate of mean concentration based on the Soil Test Database was substantially larger at 8.45 g/kg.
Discussion

There is a growing consensus that soil inorganic carbon might play a major role in the global carbon cycle (Sanderman, 2012) and that it must be accounted for within carbon budgeting exercises (Guo et al., 2006). However, many SMNs omit soil inorganic carbon (Arrouays et al., 2008). We have demonstrated that when soil inorganic carbon is measured within an SMN it is possible to produce unbiased design-based estimates of the total stock and to estimate a statistical model of the spatial variation of these stocks. Upon validation, the model was found to be consistent with the design-based mean stocks and the variability of stocks was accurately represented. Therefore, it was possible to use the model to quantify the uncertainty of the design-based total stocks and to map the variation in these stocks. These uncertainties are proportionally smaller than those reported for the estimate of soil organic carbon stocks in the USA that was calculated from the State Soil Geographic Database (Guo et al., 2006). Other surveys of soil inorganic carbon based on linear regressions (Rawlins et al., 2011) or legacy data (Lettens et al., 2004) do not quantify uncertainties. We quote the design-based estimate of inorganic carbon stocks because it is known to be unbiased. The estimated total inorganic carbon stock in the top 30 cm of French soils of 1070 Tg or 6.54 kg/m³ is approximately one third of the stock of organic carbon estimated on the same depth in other studies of French soils. For example, the estimated organic carbon stocks according to Arrouays et al. (2001), Martin et al. (2011) and Meersmans et al. (2012) were 3 100 Tg, 3 260 Tg and 3 700 Tg respectively. The stocks per unit volume are more than ten times larger than those observed in neighbouring Belgium (Lettens et al., 2004). This reflects an absence of large calcareous deposits in Belgium. The influence of parent material on the maps of inorganic carbon stock and concentrations in France was clearly evident. The largest values were recorded in the calcareous and shallow chalky soils of the Champagne region, on the Jurassic rocks of the Charentes region (south-west France) and on various calcareous rocks of the Mediterranean region of southern France. The concentrations of around 60 g/kg predicted in the Champagne region are comparable with the values measured in
the south of England by Rawlins et al. (2011). The inorganic carbon concentration observed on dolomitic rocks was surprisingly small (Table 1). This might be because weathering of these topsoils had led to the removal of a large proportion of the geogenic magnesium carbonate or the carbonate concentration of the soils might have been underestimated since magnesium carbonate is harder to extract with acid than calcium carbonate.

The map of inorganic carbon concentrations based upon observations from the Soil Test Database was similar in appearance to the RMQS map. However, the national-scale mean was substantially larger than the RMQS value and the estimated means for the majority of cantons exceeded the corresponding value predicted from the RMQS data. This is presumably because the soil tests were not sampled according to a statistical design and in particular because farmers would not have measured inorganic carbon where they expected concentrations to be small. We attempted to adjust the dataset to account for this effect but our adjustment was not sufficient. If the Soil Test Database were to be used to estimate inorganic carbon concentrations a better calibration of the model to adjust for these preferences would be required. These findings illustrate that although such legacy data can provide useful information about the variation of soil properties it is important to understand exactly how the observed samples were selected.

Conclusions

The data collected in the RMQS was used to estimate that there are inorganic carbon stocks of 1070 ±61 Tg in the top 30 cm of French soil. This is approximately one third of the organic carbon stocks to the same depth. The mean concentration was estimated to be 6.46±0.38 g/kg. This inorganic carbon is likely to decrease through acidification that will result from atmospheric and voluntary application of nitrogen and sulphur to the soil. Future phases of the RMQS will monitor changes in these baseline stocks of inorganic carbon.

Acknowledgements
RMQS soil sampling and physico-chemical analyses were supported by a French Scientific Group of
Interest on soils: the ‘GIS Sol’, involving the French Ministry for Ecology and Sustainable
Development, the French Ministry of Agriculture, the Environment and Energy Management Agency
(ADEME), The French National Forest Inventory (IFN), the French Institute for Research and
Development (IRD) and the National Institute for Agronomic Research (INRA). We thank all the soil
surveyors and technical assistants involved in sampling the sites. BPM was funded by the
Biotechnology and Biological Sciences Research Council and the British Geological Survey (Natural
Environment Research Council). BGR was funded by the British Geological Survey. This paper is
published with the permission of the Executive Director of the British Geological Survey.

References


Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. In:

Akadémiai Kiadó, Budapest.

distribution in France. Soil Use and Management, 17, 7–11.

d'Agriculture de France, 88, 93–105.


EUR23490 EN/2B, Office for the Official Publications of the European Communities, Luxembourg,
254pp.


Beta. Joint Research Centre, 27pp. Available at:
http://eusoils.jrc.ec.europa.eu/Esdb_Archive/eusoils_docs/esb_rr/SPADE-2_Beta_Report.pdf,
accessed 21 March 2011.

geographic database. In: European Land Information Systems for Agro-environmental Monitoring,
of the European Communities, Luxembourg.


Lark, R.M., Cullis, B.R. & Welham, S.J. 2006. On spatial prediction of soil properties in the presence of
a spatial trend: The empirical best linear unbiased predictor (E-BLUP) with REML. European Journal
of Soil Science, 57, 787–799.

S. & Abrassart, J. 2008. Soil phosphorus monitoring at the regional level by means of a soil test

landscape units derived using data from 1950 to 1970. Soil Use and Management, 20, 40—47.

Manning, D.A.C. 2008. Biological enhancement of soil carbonate precipitation: passive removal of


Figure Captions:

Figure 1 France and surrounding countries.

Figure 2 (a) Sampling locations for the RMQS and (b) parent material classes.

Figure 3 Histograms of (a) measured inorganic carbon (IC) concentration (g/kg) and (b) stock (kg/m³) at RMQS sites.

Figure 4 Variograms of (a) random effects for normalized RMQS 0-30 cm observations of inorganic carbon concentration and (b) stock. Fixed effect means vary according to parent material class.

Figure 5 (a) Map of predicted soil inorganic carbon concentration from 0-30 cm RMQS data (g/kg); (b) map of predicted standard error of inorganic carbon concentration (g/kg); (c) map of predicted soil inorganic carbon stock from 0-30 cm RMQS data (kg/m³); (d) map of standard error of inorganic carbon stock (kg/m³).

Figure 6 (a) Map of predicted 0-30 cm soil inorganic carbon concentration from RMQS data (g/kg) presented on a nonlinear scale; (b) map of predicted 0-30 cm soil inorganic carbon stock (kg/m³) from RMQS data presented on a nonlinear scale.

Figure 7 (a) Map of predicted 0-30 cm soil inorganic carbon concentration from Soil Test Database (g/kg) presented on a nonlinear scale; (b) expected concentration of inorganic carbon estimated from Soil Test Database (g/kg) plotted against corresponding estimate from RMQS data. The red line is $y = x$. 
Table captions

Table 1 Summary statistics (number of observations $n$, mean, median and standard deviation $sd$) for RMQS observations of inorganic carbon concentration and stock divided according to parent material (PM) class.

Table 2 Cross-validation statistics and AIC for estimated LMMs of inorganic carbon (IC) concentration and stock.

Table 3: Model- and design-based estimates of mean inorganic carbon concentration across France and total inorganic carbon stock.
Figures:

Figure 1
Figure 2
Figure 3

(a) Frequency vs. IC concentration (g/kg)

(b) Frequency vs. IC stock (kg/m³)
Figure 4

(a) \( \gamma(h) \) vs. Lag (km)

(b) \( \gamma(h) \) vs. Lag (km)
Figure 5

(a) Predicted concentration (g/kg)
(b) SE concentration (g/kg)
(c) Predicted stock (kg/m³)
(d) SE stock (kg/m³)
Figure 6
<table>
<thead>
<tr>
<th>PM class</th>
<th>Concentration (g/kg)</th>
<th>Stock (kg/m³)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>mean</td>
</tr>
<tr>
<td>All</td>
<td>2168</td>
<td>6.46</td>
</tr>
<tr>
<td>Calcareous rocks</td>
<td>471</td>
<td>12.53</td>
</tr>
<tr>
<td>Chalk</td>
<td>64</td>
<td>39.19</td>
</tr>
<tr>
<td>Marls</td>
<td>109</td>
<td>8.61</td>
</tr>
<tr>
<td>Dolomitic rocks</td>
<td>24</td>
<td>0.49</td>
</tr>
<tr>
<td>Non-calcareous</td>
<td>1499</td>
<td>3.04</td>
</tr>
</tbody>
</table>
Table 2

<table>
<thead>
<tr>
<th>Property</th>
<th>Fixed effects</th>
<th>Mean θ</th>
<th>Median θ</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>IC concentration</td>
<td>Constant</td>
<td>1.00</td>
<td>0.41</td>
<td>1550.4</td>
</tr>
<tr>
<td>IC concentration</td>
<td>P. material</td>
<td>1.00</td>
<td>0.43</td>
<td>1532.3</td>
</tr>
<tr>
<td>IC stock</td>
<td>Constant</td>
<td>1.00</td>
<td>0.40</td>
<td>1642.3</td>
</tr>
<tr>
<td>IC stock</td>
<td>P. material</td>
<td>1.00</td>
<td>0.40</td>
<td>1528.5</td>
</tr>
<tr>
<td>Estimator</td>
<td>Concentration (g/kg)</td>
<td>Stock (Tg)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>----------------------</td>
<td>--------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Design-based RMQS</td>
<td>6.46±0.38</td>
<td>1070 ± 61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model-based RMQS</td>
<td>6.79±0.28</td>
<td>1136± 45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soil Test Database</td>
<td>8.45</td>
<td>na</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>