

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

3 B.P. Marchant^{1,2*}, E.J. Villanneau³, D. Arrouays³, N.P.A. Saby³ & B.G. Rawlins¹

4 ¹ British Geological Survey, Environmental Science Centre, Keyworth, Nottingham, NG12 5GG, UK.

5 ²Rothamsted Research, West Common, Harpenden, AL5 2JQ, UK.

6 ³INRA, US 1106 InfoSol, F-4075 Orléans, France.

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

8 **Abstract**

9 The potential for soils to act as a sink or a source of atmospheric carbon has long been recognised
10 and great efforts are made to monitor soil organic carbon stocks. Inorganic carbon is also exchanged
11 between soils and the atmosphere and is important for soil function but inorganic carbon stocks are
12 not measured in many national- and continental-scale soil monitoring networks. Topsoil (0–30 cm)
13 inorganic carbon concentrations were measured at more than 2 000 sites on a regular 16-km grid as
14 part of the French National Soil Monitoring Network (RMQS: Réseau de Mesures de la Qualité des
15 Sols). We used design-based statistical methods to calculate unbiased estimates of the mean soil
16 inorganic carbon concentration and total stocks across France and we used model-based methods to
17 determine the uncertainty of these estimates and to predict the spatial distribution of these
18 quantities. The observations of inorganic carbon were highly skewed and did not conform to
19 standard statistical models. We therefore applied a non-parametric transform to normalise the data
20 and performed validation of the model that resulted. The estimates and predictions of inorganic
21 carbon are baselines against which the results of future phases of the network can be compared. We
22 found that the total topsoil inorganic carbon stocks in France amount to 1070 ± 61 Tg (Tg = terragram
23 = 10^{12} g) which is approximately one third of the mass of the corresponding organic carbon stocks.

24 Their spatial distribution was strongly linked to the underlying geology. We note that in countries
25 where inorganic carbon is not measured within soil monitoring networks, its concentrations and
26 stocks might be estimated from other sources of soil legacy information such as databases of soil
27 tests requested by farmers. We tested the reliability of this approach by considering the French Soil
28 Test Database which contains the results of 280 000 soil analyses requested by farmers between
29 1990 and 2004. We found that these data led to a biased estimate of soil inorganic carbon
30 concentrations which was presumably the result of the locations of the soil samples being selected
31 according to the concerns of farmers rather than a statistical design.

32 **Introduction**

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

41 Batjes (1996) estimated that one third of the global soil carbon stocks contained in the upper metre
42 of soil are in an inorganic form. These stocks can also act as either a sink or a source of atmospheric
43 carbon. The reaction of calcium and magnesium with atmospheric carbon dioxide and water in the
44 upper soil horizons can lead to the formation of secondary carbonates and the sequestration of
45 atmospheric carbon (Sahrawat, 2003). Manning (2008) and Washbourne *et al.* (2012) have
46 suggested that further stocks of inorganic carbon could be sequestered through the incorporation
47 into the soil of waste calcium silicate minerals from quarrying, demolition and steel manufacture.
48 Suarez (2006) noted that land management practices such as irrigation and use of acidifying

49 fertilizers can cause losses of soil inorganic carbon although inorganic carbon can be added to the
50 soil upon liming. There is a further need to monitor inorganic carbon where it supports soil function.
51 For example, in many soil types the dissolution of soil carbonate is the dominant buffering
52 mechanism which limits the soil acidification which can otherwise occur when nitrogen fertilisers are
53 added to the soil.

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

64 Previously reported broad-scale estimates of soil inorganic carbon stocks and concentrations have
65 been based on existing soil databases and information systems that were not constructed according
66 to a statistical design or with quantitative soil monitoring in mind. Expert judgement has been used
67 in the establishment of these databases to divide the study region into different soil classes and to
68 locate representative soil pits within these classes. For example, Wu *et al.* (2009) used 2 553 soil
69 profiles from the second national Chinese soil survey to determine inorganic carbon stocks in the A,
70 B and C horizons for each subgroup of the Chinese Soil Taxonomy. They upscaled these results to a
71 national stock of 55.3 ± 10.7 Pg or 6.3 ± 1.2 kg/m². Similarly, Guo *et al.* (2006) used the State Soil
72 Geographic Database to estimate the stock of inorganic carbon in the top two metres of soil in the
73 conterminous USA to be between 22.6 and 93.7 Pg. Lettens *et al.* (2004) divided Belgium into 492

74 landscape units based upon land cover and soil association and estimated mean inorganic carbon
75 stocks for each landscape unit using 13 000 geo-referenced soil profile descriptions. They found that
76 the inorganic stock in the top 30 cm was 4 Tg or 0.54 kg/m³. The uncertainty of the estimate was not
77 reported. The maps and estimates which result from these databases are informative but it is
78 challenging to account for the expert judgements when the uncertainty of the outputs is quantified.

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

94 In contrast, purpose-built SMNs use a statistical design to ensure unbiased estimates of both soil
95 status and soil change. The French national SMN (RMQS – Réseau de Mesures de la Qualité des Sols)
96 included over 2 000 observations of the concentration of soil carbonates and stocks of soil inorganic
97 carbon in the top 30 cm of soil across France (Fig. 1). We used geostatistical techniques to map the
98 inorganic carbon stocks across France and to determine the total mass of inorganic carbon stored in

99 the top 30 cm of soil. We used knowledge of the parent material across France (Fig. 2) to improve
100 our predictions. Such geostatistical techniques require that a valid model of the spatial variation of
101 inorganic carbon is fitted to the spatial data. We performed a design-based analysis (Brus & De
102 Gruijter, 1997) to validate our prediction of the total stocks. Design-based analyses require much
103 less stringent assumptions about the statistical distribution of the data. We also explored whether
104 the soil inorganic carbon concentrations could have been estimated from sources of soils data that
105 might be available in other countries. In particular, we considered the results of 268 000 soil
106 analyses that have been requested by farmers in France between 1990 and 2004.

107 **Statistical Theory**

108 *Design of soil monitoring networks*

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

120 Mean values and maps of properties can be determined from legacy datasets that were not
121 collected for the purpose of monitoring. For example, Orton *et al.* (2012) have estimated changes in
122 soil carbon concentrations in French soils using measurements requested by farmers to assess the

123 fertility of their soils. Generally, the locations of these measurements are not selected according to a
124 statistical design. Design-based analyses can only be applied if the locations are selected
125 probabilistically. Model-based methods can be applied if the sample selection is not biased
126 according to the expected value of the measurement. However, biases often occur in such surveys.
127 For example, a farmer might bias their measurements towards areas where they are concerned
128 about soil fertility. Such preferential sampling must be accounted for when using the measurements
129 to predict soil properties.

130 *Model-based analyses*

131 A model-based analysis of a survey assumes that the observed data are a realization of a random
132 function. A statistical model that describes the random function, including any spatial correlation,
133 must be fitted to the data prior to prediction and the production of maps. The statistical model is a
134 simplification of reality and its use requires that assumptions are made about the random function
135 from which the observed data are realized. For example, the random function is often assumed to
136 have a multivariate Gaussian distribution and to be second order stationary (Webster & Oliver,
137 2007).

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

148 residual maximum likelihood (REML; see Lark *et al.*, 2006). The Matérn model is a flexible variogram
149 function that is often advocated in studies of soil properties (*e.g.* Marchant & Lark, 2007).

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

$$AIC = -2L + 2q, \quad (1)$$

152 for each model where L is the ML log-likelihood and q is the number of parameters in the model.

153 The model with the smallest AIC best manages the trade-off between quality of fit and model
154 complexity (Webster & Oliver, 2007).

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

162 Standard bootstrapping or resampling methods (Efron & Tibshirani, 1993) cannot be used to validate
163 a spatial LMM because these methods do not account for the spatial correlation amongst the data.

164 An LMM can be validated by leave-one-out cross validation. Here each observation, $z^*(\mathbf{x}_i)$, is
165 removed in turn from the dataset and the remaining observations are used to predict the
166 (transformed) property at the site of the removed observation. The standardized prediction errors

$$\theta_i = \frac{\{z^*(\mathbf{x}_i) - m_i\}^2}{\sigma_i^2}, \quad (2)$$

167 are calculated. If the LMM is accurate, the expected square error should be equal to the kriging
168 variance, and the mean of θ should be 1 and the median equal to 0.45.

169 The mean of a property across the entire prediction grid can be estimated by using the multivariate
 170 LU simulation method (Deutsch & Journel, 1998) to produce a large number (*e.g.* 1000) of
 171 realizations of the transformed property, conditional on the transformed observed data, at all of the
 172 nodes of the prediction grid. Each simulated value is then back-transformed and the mean of each
 173 realization is calculated. The predicted mean is equal to the mean of the realization means. The
 174 prediction variance is equal to the variance of the prediction means.

175 *Design-based analyses*

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

$$\hat{z} = \frac{1}{n} \sum_{i=1}^n z(\mathbf{x}_i). \quad (3)$$

185 This estimate is said to be p -unbiased because

$$E_p[\hat{z}] = \bar{z}, \quad (4)$$

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

$$r = E_p[\hat{z} - \bar{z}]^2 = \frac{\sigma^2}{n} \approx \frac{\hat{s}^2}{n}, \quad (5)$$

188 where

$$\hat{s}^2 = \frac{1}{n-1} \sum_{i=1}^n \{z(\mathbf{x}_i) - \hat{z}\}^2, \quad (6)$$

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

194 Design-based analyses have been extended to consider systematic random designs. Such a design
195 follows a fixed pattern but the location of the origin is selected at random. When such a design is
196 used the mean of z across the study region can still be estimated from Eqn. 3 because each site has
197 an equal probability of being included in the design and the estimator is p -unbiased. However, if
198 $z(\mathbf{x})$ is spatially correlated then Eqn. 5 will be a biased estimate of the variance of $z(\mathbf{x})$ because the
199 pattern of sampling locations is not random. In the case of a regular grid with good coverage of the
200 study region, Eqn. 6 will over-estimate the variance of $z(\mathbf{x})$ since the smallest lag between pairs of
201 observations is equal to the grid-spacing. Therefore, Eqn. 5 will over-estimate the expected squared
202 estimation error. Domburg *et al.* (1994) proposed a method by which the expected squared error of
203 \hat{z} can be estimated for a systematic design using simulated realizations of an estimated model of
204 the random function $Z(\mathbf{x})$.

205 **Methods**

206 *The French National Soil Monitoring Network (RMQS)*

207 The RMQS database (Arrouays *et al.*, 2002) primarily consists of over 2 000 observations of soil
208 properties on a 16-km regular grid across the 550 000 km² French metropolitan territory (Fig. 2a).

209 Twenty six additional observations were positioned within the grid. The locations of these

210 observations targeted specific combinations of soil type and land use. If it was not possible to
211 sample at the selected site (*e.g.* in urban areas, in rivers or on roads) an alternative cultivated or
212 undisturbed location was selected as nearby as possible within a 1-km radius. When this was not
213 possible the cell was omitted from the survey. This baseline survey of the RMQS was completed in
214 2009.

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

220 *Preparation and analysis of soil samples*

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

231 The mass of fine earth (*i.e.* particles which pass through a 2-mm sieve) for each observation site was
232 determined for three samples of known volume that were extracted from a soil pit adjacent to the
233 site. The methods varied according to the particle size distribution of the samples. The samples were

234 dried prior to analysis. The cylinder method (procedure reference NF X 31-501 in AFNOR, 1999) was
235 used for soils with little to no gravel. The cylinder was 90 mm high with a diameter of 84 mm, *i.e.*,
236 500 cm³. When the cylinder method was not possible, *i.e.* for gravelly or stony soils, the water
237 method was applied. The water method is adapted from the sand method (procedure reference NF
238 X 31-503 in AFNOR, 1999). In this method, circular holes of between 1 000 and 3 000 cm³ are dug
239 into the soil. The exact volume of soil extracted is determined by lining the hole with a plastic bag
240 and measuring the volume of water required to fill the hole.

241 *Soil parent material*

242 Information on soil parent material was derived from the 1:1 000 000 scale soil database of Europe
243 (King *et al.*, 1995). The classes from this database were amalgamated into five broader classes (Table
244 1 and Fig. 2b) for this study. These classes were created according to their natural geochemical
245 concentration of calcium and magnesium carbonate according to the attribute tables of the
246 database. The parent material classes were marls (covering 3.1% of the area of France), chalk (5.3%),
247 other calcareous rocks (21.2%), dolomitic rocks (1.1%) and non-calcareous rocks (69.3%).

248 *The French Soil Test Database*

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

258 could be recorded was the name of the local authority or canton which contained the sampled field.
259 France is divided into 4 032 cantons which are intended to have approximately equal populations of
260 around 20 000. Hence, urban cantons tend to be small but rural ones can have areas of over 100
261 km².

262 *Model-based analyses*

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

278 *Design-based analyses*

279 We re-estimated the mean inorganic carbon concentration and the total inorganic carbon stocks
280 using design-based methods. This analysis used only the observations that were located on the
281 nodes of the 16-km grid. We assumed that the origin of the RMQS grid was selected at random and

282 that Eqn. 3 yielded a p unbiased estimate of these quantities. We determined the estimation
283 variance using the constant mean LMMs fitted to the observed data and the approach suggested by
284 Domburg *et al.* (1994).

285 *Analyses of the Soil Test Database*

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

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

301 **Results**

302 *Mapping soil inorganic carbon*

303 There were substantial differences between the inorganic carbon concentrations observed in each
304 parent material class (Table 1). The AICs for the estimated LMMs of both concentrations and stocks
305 with fixed effects that varied according to parent material were less than the AICs for the constant

306 fixed effect models. Therefore, we included parent material classes as fixed effects in our model-
307 based analyses. All of the estimated LMMs had mean θ equal to 1.00 and median θ close to 0.45
308 (Table 2) indicating that the residual variation had been adequately modelled.

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

324 *National estimates of inorganic carbon*

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

330 **Discussion**

331 There is a growing consensus that soil inorganic carbon might play a major role in the global carbon
332 cycle (Sanderman, 2012) and that it must be accounted for within carbon budgeting exercises (Guo
333 *et al.*, 2006). However, many SMNs omit soil inorganic carbon (Arrouays *et al.*, 2008). We have
334 demonstrated that when soil inorganic carbon is measured within an SMN it is possible to produce
335 unbiased design-based estimates of the total stock and to estimate a statistical model of the spatial
336 variation of these stocks. Upon validation, the model was found to be consistent with the design-
337 based mean stocks and the variability of stocks was accurately represented. Therefore, it was
338 possible to use the model to quantify the uncertainty of the design-based total stocks and to map
339 the variation in these stocks. These uncertainties are proportionally smaller than those reported for
340 the estimate of soil organic carbon stocks in the USA that was calculated from the State Soil
341 Geographic Database (Guo *et al.*, 2006). Other surveys of soil inorganic carbon based on linear
342 regressions (Rawlins *et al.*, 2011) or legacy data (Lettens *et al.*, 2004) do not quantify uncertainties.
343 We quote the design-based estimate of inorganic carbon stocks because it is known to be unbiased.
344 The estimated total inorganic carbon stock in the top 30 cm of French soils of 1070 Tg or 6.54 kg/m³
345 is approximately one third of the stock of organic carbon estimated on the same depth in other
346 studies of French soils. For example, the estimated organic carbon stocks according to Arrouays *et*
347 *al.* (2001), Martin *et al.* (2011) and Meersmans *et al.* (2012) were 3 100 Tg, 3 260 Tg and 3 700 Tg
348 respectively. The stocks per unit volume are more than ten times larger than those observed in
349 neighbouring Belgium (Lettens *et al.*, 2004). This reflects an absence of large calcareous deposits in
350 Belgium. The influence of parent material on the maps of inorganic carbon stock and concentrations
351 in France was clearly evident. The largest values were recorded in the calcareous and shallow chalky
352 soils of the Champagne region, on the Jurassic rocks of the Charentes region (south-west France)
353 and on various calcareous rocks of the Mediterranean region of southern France. The concentrations
354 of around 60 g/kg predicted in the Champagne region are comparable with the values measured in

355 the south of England by Rawlins *et al.* (2011). The inorganic carbon concentration observed on
356 dolomitic rocks was surprisingly small (Table 1). This might be because weathering of these topsoils
357 had led to the removal of a large proportion of the geogenic magnesium carbonate or the carbonate
358 concentration of the soils might have been underestimated since magnesium carbonate is harder to
359 extract with acid than calcium carbonate.

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

371 **Conclusions**

372 The data collected in the RMQS was used to estimate that there are inorganic carbon stocks of 1070
373 ± 61 Tg in the top 30 cm of French soil. This is approximately one third of the organic carbon stocks to
374 the same depth. The mean concentration was estimated to be 6.46 ± 0.38 g/kg. This inorganic carbon
375 is likely to decrease through acidification that will result from atmospheric and voluntary application
376 of nitrogen and sulphur to the soil. Future phases of the RMQS will monitor changes in these
377 baseline stocks of inorganic carbon.

378 **Acknowledgements**

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

388 **References**

- 389 AFNOR, 1999. *Qualité des sols*, 2. AFNOR, Paris.
- 390 Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. In:
391 *Second International Symposium on Information Theory* (eds B.N. Petrov and F. Csáki), pp. 267–281.
392 Akadémiai Kiadó, Budapest.
- 393 Arrouays, D., Deslais, W. & Bateau, V. 2001. The carbon content of topsoil and its geographical
394 distribution in France. *Soil Use and Management*, **17**, 7–11.
- 395 Arrouays, D., Jolivet, C., Boulonne, L., Bodineau, G., Saby, N. & Grolleau, E. 2002. A new initiative in
396 France: a multi-institutional soil quality monitoring network. *Comptes Rendus de l'Academie*
397 *d'Agriculture de France*, **88**, 93–105.
- 398 Arrouays, D., Richer de Forges, A.C., Morvan, X., Saby, N.P.A., Jones, A.R., Le Bas, C. (eds). 2008.
399 Environmental Assessment of Soil for Monitoring: Volume IIb Survey of National Networks.
400 EUR23490 EN/2B, Office for the Official Publications of the European Communities, Luxembourg,
401 254pp.

402 Arrouays, D., Marchant, B.P., Saby, N.P.A., Meersmans, J., Orton, T.G., Martin, M.P., Bellamy, P.H.,
403 Lark, R.M. & Kibblewhite, M. 2012. Generic Issues on Broad-Scale Soil Monitoring Schemes: A
404 Review. *Pedosphere*, **22**, 456–469.

405 Batjes, N.H. 1996. Total carbon and nitrogen in the soils of the world. *European Journal of Soil*
406 *Science*, **47**, 151–163.

407 Black, H., Bellamy, P., Creamer, R., Elston, D., Emmett, B., Frogbrook, Z., Hudson, G., Jordan, J., Lark,
408 M., Lilly, A., Marchant, B., Plum, S., Potts, J., Reynolds, B., Thompson, R. & Booth, P. 2008. Design
409 and operation of a UK soil monitoring network. Science Report SC060073, Environment Agency,
410 Bristol, UK.

411 Brus, D.J. & De Gruijter, J.J. 1997. Random sampling or geostatistical modelling? Choosing between
412 design-based and model-based sampling strategies for soil (with Discussion). *Geoderma*, **80**, 1–44.

413 Deutsch, C.V. & Journel, A.G. 1998. *GSLIB: geostatistical software library and user's guide*. Second
414 edition, pp. 369. Oxford University Press, Oxford, UK.

415 Domburg, P., de Gruijter, J.J. & Brus, D.J. 1994. A structured approach to designing soil survey
416 schemes with prediction of sampling error from variograms. *Geoderma*, **62**, 151–164.

417 Efron, B. & Tibshirani, R. 1993. *An Introduction to the Bootstrap*. Chapman & Hall/CRC, Boca Raton,
418 FL.

419 Goovaerts, P. 1997. *Geostatistics for Natural Resources Evaluation*. Applied Geostatistics Series,
420 Oxford University Press, New York.

421 Guo, Y., Amundson, R., Gong, P. & Yu, Q. 2006. Quantity and spatial variability of soil carbon in the
422 conterminous United States. *Soil Science Society of America Journal*, **70**, 590–600.

423 Hannam, J.A., Hollis, J.M., Jones, R.J.A., Bellamy, P.H., Hayes, S.E., Holden, A., Van Liedekerke, M.H. &
424 Montanarella, L.V.M.U.R., 2009. SPADE-2: The soil profile analytical database for Europe Version 2.0

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

428 King, D., Burrill, A., Daroussin, J., Le Bas, C., Tavernier, R. & Van Ranst, E. 1995. The EU soil
429 geographic database. In: European Land Information Systems for Agro-environmental Monitoring,
430 EUR 16232 EN (eds D.King, R.J.A.Jones & A.J.Thomasson), pp. 43–60. Office for Official Publications
431 of the European Communities, Luxembourg.

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

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

436 Lemerrier, B., Gaudin, L., Walter, C., Arousseau, P., Arrouays, D., Schwartz, C., Saby, N. P. A., Follain,
437 S. & Abrassart, J. 2008. Soil phosphorus monitoring at the regional level by means of a soil test
438 database. *Soil Use and Management*, **24**, 131–138.

439 Lettens, S., Van Orshoven, J., van Wesemael, B. & Muys, B. 2004. Soil organic and inorganic
440 landscape units derived using data from 1950 to 1970. *Soil Use and Management*, **20**, 40–47.

441 Manning, D.A.C. 2008. Biological enhancement of soil carbonate precipitation: passive removal of
442 atmospheric CO₂. *Mineralogical Magazine*, **72**, 639–649.

443 Marchant, B.P. & Lark, R.M. 2007. The Matérn variogram model: Implications for uncertainty
444 propagation and sampling in geostatistical surveys. *Geoderma*, **140**, 337–345.

445 Martin, M.P., Wattenbach, M., Smith, P., Meersmans, J., Jolivet, C.C., Boulonne, L. & Arrouays, D.
446 2011. Soil organic carbon stocks distribution in France. *Biogeosciences*, **8**, 1053–1065.

447 Meersmans, J., Martin, M.P., Lacarce, E., De Baets, S., Jolivet, C., Boulonne, L., Lehmann, S., Saby,
448 N.P.A., Bispo, A. & Arrouays, D. 2012. A high resolution map of French soil organic carbon. *Agronomy*
449 *for Sustainable Development*, **32**, 841–851.

450 Na, M.I., Shaoqiang, W., Jiyuan, L.I.U., Guirui, Y.U., Wenjuan, Z. & Esteban, J. 2008. Soil inorganic
451 carbon storage pattern in China. *Global Change Biology*, **14**, 2380–2387.

452 Orton, T.G., Saby, N.P.A., Arrouays, D., Walter, C., Lemerrier, B., Schwartz, C. & Lark, R.M. 2012.
453 Spatial prediction of soil organic carbon from data on large and variable spatial supports. II. Mapping
454 temporal change. *Environmetrics*, **23**, 148–161.

455 Rawlins, B.G., Henrys, P., Breward, N., Robinson, D.A., Keith, A.M. & Garcia-Bajo, M. 2011. The
456 importance of inorganic carbon in soil carbon databases and stock estimates: a case study from
457 England. *Soil Use and Management*, **27**, 312–320.

458 Sahrawat, K.L. 2003. Importance of inorganic carbon in sequestering carbon in soils of the dry
459 regions. *Current Science*, **84**, 864–865.

460 Salminen, R., Batista, M.J., Bidovec, M., Demetriades, A., De Vivo, B., De Vos, W., Duris, M., Gilucis,
461 A., Gregorauskiene, V., Halamic, J., Heitzmann, P., Lima, A., Jordan, G., Klaver, G., Klein, P., Lis, J.,
462 Locutura, J., Marsina, K., Mazreku, A., O'Connor, P.J., Olsson, S.Å., Ottesen, R.-T., Petersell, V., Plant,
463 J.A., Reeder, S., Salpeteur, I., Sandström, H., Siewers, U., Steenfelt, A. & Tarvainen, T., Geochemical
464 Atlas of Europe, P. B. I., Methodology and & Maps. Geological Survey of Finland, E., Finland. 2005.
465 Geochemical Atlas of Europe, part 1: background information, methodology and maps. Geological
466 Survey of Finland, Espoo, Finland.

467 Sanderman, J. 2012. Can management induced changes in the carbonate system drive soil carbon
468 sequestration? A review with particular focus on Australia. *Agriculture, Ecosystems and Environment*,
469 **155**, 70–77.

470 Sanderman, J., Baldock, J., Hawke, J., Macdonald, L., Massis-Puccini, A., & Szarvas, S. 2011. National
471 Soil Carbon Research Programme: Field and Laboratories Methodologies, CSIRO Land and Water,
472 Urrbrae, Australia.

473 Suarez, D.L. 2006. Inorganic carbon: land use impacts. In: Encyclopedia of Soil Science (ed R. Lal), pp.
474 895–897. Taylor and Francis, UK.

475 Washbourne, C.-L., Renforth, P. & Manning, D.A.C. 2012. Investigating carbonate formation in urban
476 soils as a method for capture and storage of atmospheric carbon. *Science of the Total Environment*,
477 **431**, 166–175.

478 Webster, R. & Oliver, M. 2007. *Geostatistics for Environmental Scientists*. 2nd Edition, John Wiley &
479 Sons Ltd, Chichester UK.

480 Wu, H., Guo, Z., Gao, Q. & Peng, C. 2009. Distribution of soil inorganic carbon storage and its
481 changes due to agricultural land use activity in China. *Agriculture, Ecosystems and Environment*, **129**,
482 413–421.

483

484 **Figure Captions:**

485 Figure 1 France and surrounding countries.

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

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

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

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

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

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

502 **Table captions**

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

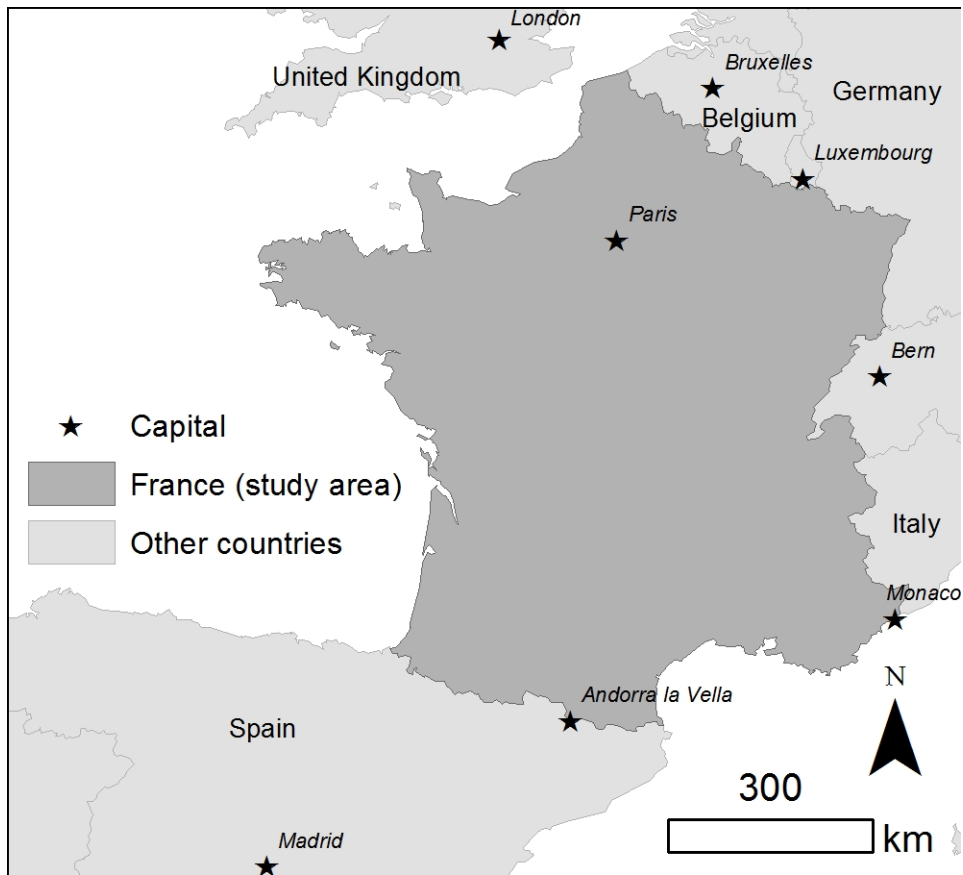
506 Table 2 Cross-validation statistics and AIC for estimated LMMs of inorganic carbon (IC) concentration
507 and stock.

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

510

511 **Figures:**

512 **Figure 1**



513

514

515

516

517

518

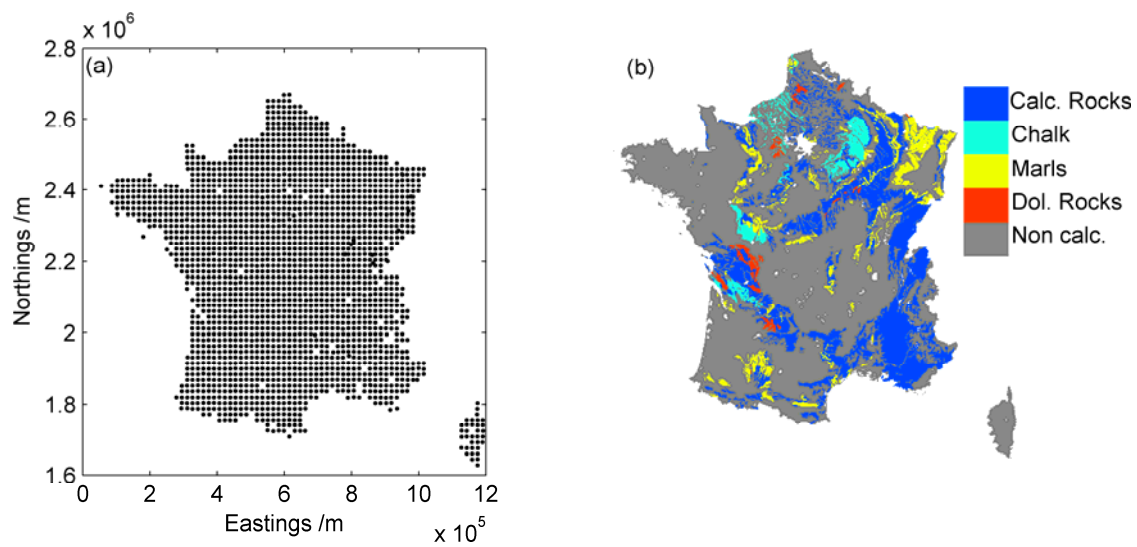
519

520

521

522 Figure 2

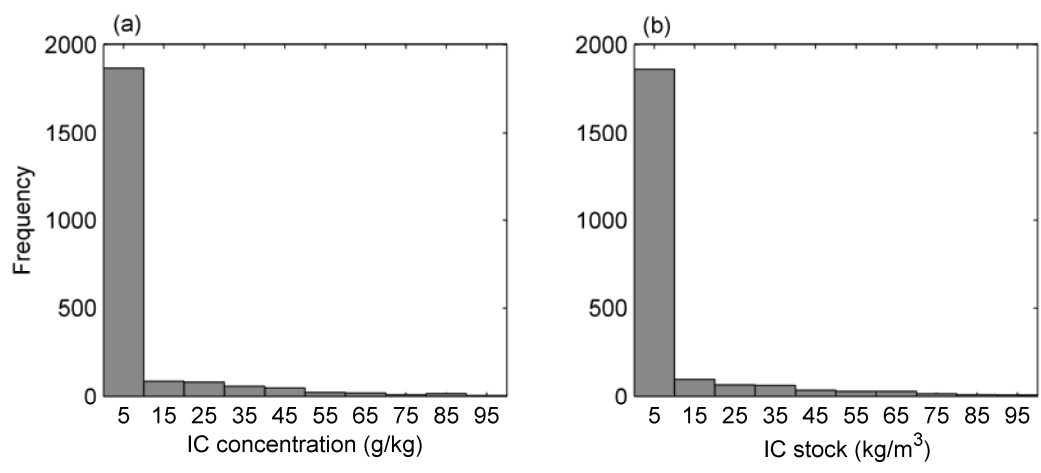
523



524

525

526 Figure 3



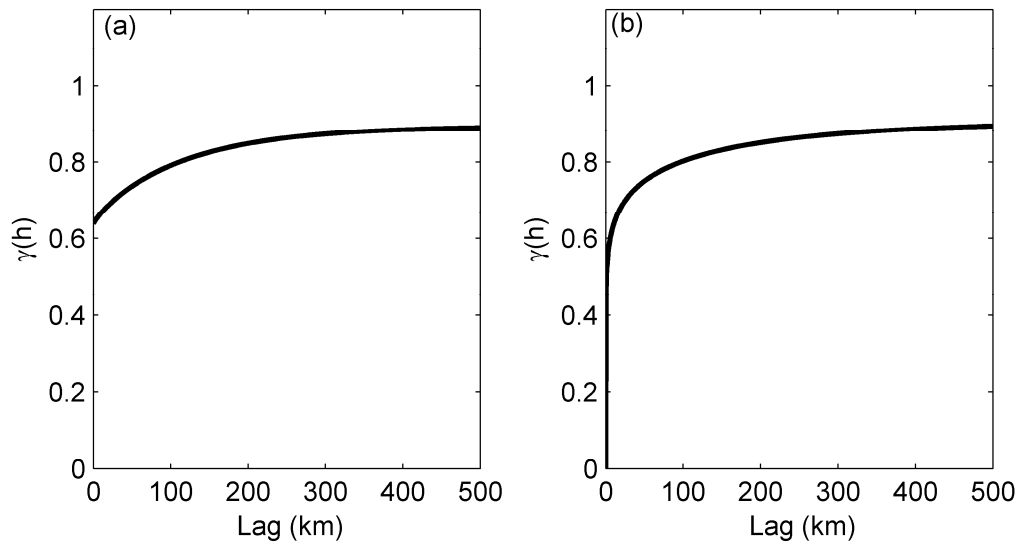
527

528

529

530

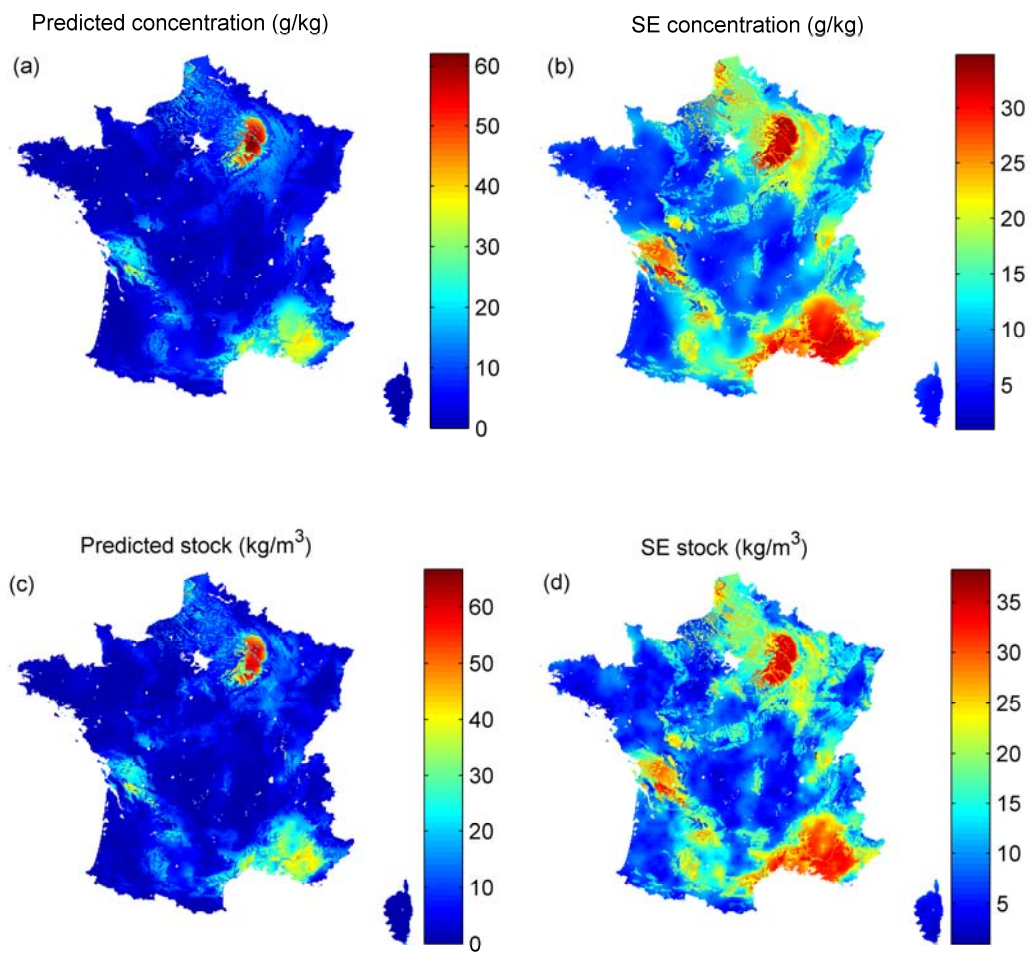
531 Figure 4



532

533

534

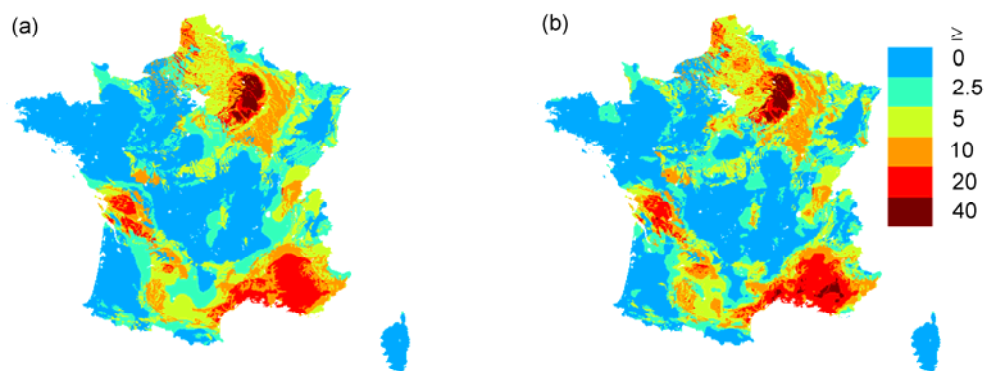


536

537

538

539 Figure 6



540

541

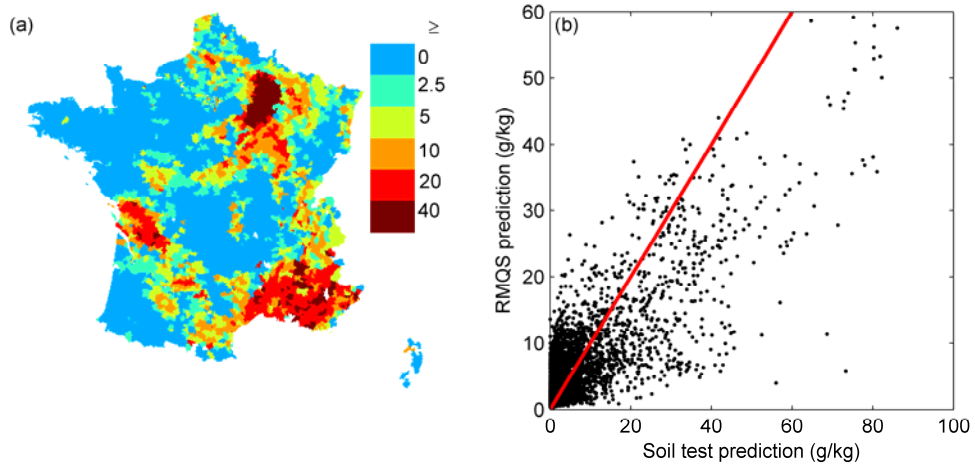
542

543

544

545

546 Figure 7



547

548

549

550

551

552

553

554

555

556

557

558

559

560 **Tables**

561 Table 1

PM class	Concentration (g/kg)				Stock (kg/m ³)			
	<i>n</i>	mean	median	sd	<i>n</i>	mean	median	sd
All	2168	6.46	0.06	16.00	2168	6.54	0.06	16.30
Calcareous rocks	471	12.53	1.17	19.31	471	12.07	1.20	19.07
Chalk	64	39.19	28.13	38.65	64	38.05	25.45	37.82
Marls	109	8.61	0.21	15.91	109	9.64	0.27	17.61
Dolomitic rocks	24	0.49	0.05	1.41	24	0.51	0.06	1.38
Non-calcareous	1499	3.04	0.05	10.05	1499	3.34	0.05	11.16

562

563

564 Table 2

Property	Fixed effects	Mean θ	Median θ	AIC
IC concentration	Constant	1.00	0.41	1550.4
IC concentration	P. material	1.00	0.43	1532.3
IC stock	Constant	1.00	0.40	1642.3
IC stock	P. material	1.00	0.40	1528.5

565

566

567 Table 3

568

Estimator	Concentration (g/kg)	Stock (Tg)
Design-based RMQS	6.46±0.38	1070 ± 61
Model-based RMQS	6.79±0.28	1136± 45
Soil Test Database	8.45	na

569

570

571

572