

Are data collected to support farm management suitable for monitoring soil indicators at the national scale?

B. G. RAWLINS^a , B. MARCHANT^a, S. STEVENSON^b & W. WILMER^c

^aBritish Geological Survey, Keyworth, Nottingham, NG12 5GG, UK, ^bNRM Laboratories, Coopers Bridge, Braziers Lane, Bracknell, Berkshire, RG42 6NS, UK, and ^cRothamsted Research, Harpenden, Hertfordshire, AL5 2JQ, UK

Summary

Monitoring of topsoil properties (referred to as indicators) at the national scale has been limited in general to government-funded representative surveys. We consider a cost-effective complementary source of soil information for monitoring agricultural soil across England and Wales (E&W): soil measurements paid for by farmers that we refer to as farmers' data (FD). A potential problem in using FD for soil monitoring is any unattributable sources of bias, such as the sample design. Farmers may choose to focus their measurements (purposely) where they perceive a particular problem. Such a source of bias is avoided in the random sampling adopted by statistically designed surveys, such as the Countryside Survey (CS2007) and LUCAS (Land Use/Cover Area frame statistical Survey). We used measurements from 143 000 FD soil samples from a single laboratory to estimate national mean values and confidence intervals of five topsoil indicators (pH, available P (Olsen), K, Mg and organic matter (OM)) across three combinations of nation (England or Wales) and land use (arable and horticulture (A&H) or improved grassland (IG)). We computed mean estimates for FD over two time periods (2004–9 and 2010–2105) and assessed the significance of any change. We compared these estimates with those from representative national surveys to establish whether there was evidence for bias and whether it could be explained. Mean estimates of topsoil pH for the FD and the LUCAS survey (same analytical method) were consistent for both A&H and IG. Although FD estimates of mean Olsen P (OP) concentrations were similar to previous surveys, we show it is likely that the larger mean OP concentrations observed in the LUCAS survey compared with FD for arable topsoil in England are partly due to an attributable source of analytical bias. For such quantifiable sources of bias, it might be possible to adjust estimated mean values from FD. However, FD might also include sources of unattributable bias, such as the effect of purposive sampling. It is important that contemporaneous data from surveys with statistically unbiased designs are available so that we can assess whether unattributable sources exert a significant effect over estimates of mean values computed from FD.

Highlights

- Assessment of farmers' data (FD) to provide a potentially cost-effective way to monitor topsoil indicators.
- Few studies have compared national-scale estimates of topsoil indicators with survey data from statistically unbiased designs and FD.
- Bias between mean estimates from national surveys and FD could be accounted for.
- The denser sampling of FD enables mapping of national survey data with greater accuracy.

Introduction

Regulatory authorities often monitor selected topsoil properties (which they refer to as indicators) because it is an effective way to detect changes in the capacity of soil to fulfill a range of functions and to underpin policy development (Defra, 2009). Soil monitoring,

where sampling and analysis had been undertaken on at least two occasions, has advanced considerably recently (Arrouays *et al.*, 2012). To date, the majority of soil monitoring has been funded by national governments; for example, the Countryside Survey (CS2007; Emmett *et al.*, 2008) and Representative Soil Sampling Scheme (RSSS; Oliver *et al.*, 2006) in the UK, and in France the Réseau de Mesures de la Qualité des Sols (Arrouays *et al.*, 2001). Larger-scale, Europe-wide monitoring of forest soil has been carried out by the Biosoil project (de Vos & Cools, 2011). These

Correspondence: B. G. Rawlins. E-mail: bgr@bgs.ac.uk

Received 12 September 2016; revised version accepted 30 January 2017

monitoring networks were formulated to be design-unbiased in estimating mean values of selected soil indicators on the date of each sampling phase.

In the UK, 13 topsoil properties or indicators (TIs; 12 chemical measurements plus bulk density) were identified as priorities for monitoring in Merrington *et al.* (2006). In another report, Black *et al.* (2008) investigated the costs, advantages and disadvantages of various options for establishing a broad 'Tier 1' monitoring network for selected TIs based on existing data. A stated objective of this proposed network was to compute unbiased estimates of mean soil indicator values across the dominant land-use classes at the national scale so that these properties can be monitored over time. Where sampling sites are selected by probability sampling, design-based statistical inference can be adopted (de Gruijter *et al.*, 2006). There are currently no plans to continue soil monitoring across both England and Wales (E&W) based on either established or proposed networks (Environmental Audit Committee, 2016); therefore it is worthwhile to consider alternative, cost-effective approaches.

Across E&W around 82% of the land is devoted to agricultural production (Centre for Ecology & Hydrology, 2011). Farmers and agronomists regularly take samples from fields and pay for chemical analyses of various topsoil properties, many of which are also in the list of indicators. In its manual for plant nutrient management, Defra (2010) recommended that farmers measure available plant nutrient concentrations such as potassium (K) and phosphorus (P) every 3–5 years. If these farmers' data (FD) could be collated at the national scale, it may be possible to use them to monitor certain TIs. Farmers' data are taken purposively according to each farmer's needs and not according to a statistical sample design. Therefore, farmers' data are subject to sources of bias, some of which we may know about and others not. Design-based inference cannot be applied to the farmers' data because the sampling sites are not selected by probabilistic sampling. Instead, model-based statistical analyses can be used, which account for any spatial non-uniformity in the intensity of sampling using an explicit model of the spatial correlation of the TI (de Gruijter *et al.*, 2006). When sufficient observations are available the model-based approach can also be used to map the variation of TIs across a region and to identify regions where changes in land management might be required. However, standard model-based approaches do not account for preferential sampling where measurements are targeted deliberately in areas where the TI is thought to be either larger or smaller than average (Diggle *et al.*, 2010).

Model-based, national-scale estimates of quantities such as mean values computed from FD might be biased for several reasons. First, farmers or agronomists may take samples from particular fields because they perceive a problem that might be related to a TI value (Simpson, 1983). For example, if crop yields are small in a given field it might indicate available nutrient concentrations are below their optimum concentrations. Second, not all TI analyses may be on all samples collected by farmers. For example, with data from soil samples collected by French farmers, measurements of inorganic carbon (IC) concentrations were undertaken on farmers' soil samples only when the concentrations were likely to be

substantial to reduce the cost of chemical analysis. When the FD were analysed by a model-based approach to estimate mean values they were significantly larger than those estimated with data from the national soil monitoring network (Marchant *et al.*, 2015a). It is not possible to account for such unattributable sources of bias in global estimates computed from FD. By contrast, if the bias is attributable to an analytical method and the magnitude of its effect can be computed with existing data, it might be possible to modify estimates of FD mean values to account for it.

The commercial company NRM (Natural Resources Management) Laboratories (Bracknell, Berkshire, UK) typically analyses around 350 000 soil samples for a range of TIs each year for farmers and agronomists, the majority of which are from E&W. Land use and some farm addresses are often provided with each sample. In this paper we compare estimates of mean TI values (and confidence intervals) for FD from NRM Laboratories with estimates of the same TIs from national surveys. The five TIs are: pH, available P, K and Mg concentrations and organic matter (OM) content. We make these comparisons for: (i) sampling dates that are as close to one another as possible (given the available data) and (ii) the two dominant agricultural land-use classes (arable and horticulture (A&H) and improved grassland (IG)) across England, and for IG across Wales. We assess whether the estimated mean values from the FD are significantly different from other surveys, and identify potential reasons for this. We also assess the evidence for statistically significant changes in the TI values over time with the FD. We present maps to show how FD could help to highlight regions where topsoil pH might have become suboptimal. We discuss the implications of our findings for soil monitoring.

Materials and methods

Farmers' data

Agronomists and farmers in E&W who supply soil material to NRM Laboratories for chemical analysis typically adopt government guidance on soil sampling (Defra, 2010). To prepare soil samples for analysis, NRM first air-dry the material at 30°C, then remove any large lumps of rock and organic material. Next, if the material is dominated by quartz (a sandy soil) they sieve it through a 2-mm mesh. If the soil is more fine-grained with hard aggregates, they use a flailing hammer mill to disaggregate the coarse lumps and this material is then thrown at a sieve mesh, such that the angle used is equivalent to a sieve of 2 mm. The components that pass through the mesh are used for analysis. Quality control of these five TIs is assured by regular analyses of four standard soil samples (two in each batch) and participation in the the Wageningen Evaluating Programs for Analytical Laboratories (WEPAL; www.wepal.nl) international laboratory comparison scheme.

The FD comprised chemical measurements, associated land-use information, address details and sampling dates from a database for the years 2004–15 inclusive. We removed any analyses for which the address details were likely to be insufficient to retrieve a postcode. For those farm addresses without postcodes, we used an online geocoding database (www.doogal.co.uk) to extract

Table 1 Features of the soil survey datasets used in this study

Survey	CS 2007	LUCAS	NSI resampling ^a	RSSS ^b	Farmers' data
Period of sampling	2007	2009	1994–2003	2001	2004–15
Statistical sampling	Randomized	Multi-stage stratified	5 km regular	Randomized	Purposive
Inference method	Design based	Design based	Model based	Design based	Model based
Sampling depths / cm	0–15	0–15 or 20	0–15	0–15	0–15 (arable) 0–7.5 (grassland)
Sampling support	Soil core	Soil slice	25 bulked cores 20 m square	25 bulked cores over field	Farm, field scale composite sample
Indicators ^c	pH, P, OM	pH, P, K, TOC (OM)	Mg	pH, P, K, Mg	pH, P, K, Mg, OM
Reference	Emmett <i>et al.</i> (2008)	Tóth <i>et al.</i> (2013)	Kirk <i>et al.</i> (2009)	Oliver <i>et al.</i> (2006)	This study

^aNational Soil Inventory.^bRepresentative Soil Sampling Scheme.^cNutrients Olsen P, K and Mg are available concentrations; OM, organic matter; TOC, total organic carbon.

geographical coordinates based on the farm name, its town and its county. For those farms where postcodes had been recorded in the database, we used the UK Ordnance Survey database to extract the coordinates at the centre of each postcode polygon for each farm. We note that these geographical coordinates include an unknown error because the location of the farm in many cases will not be at the centre of the fields from which soil samples were taken. Nevertheless, we consider that the error in most cases would be small compared with the scale of our geostatistical models fitted with spatial ranges of between 80 and 200 km. We used the land use prior to soil sampling to allocate each sample to either (i) A&H or (ii) IG or (iii) to remove it from the database if the land use was neither A&H nor IG.

Once complete, we had a database of 143 201 soil samples with national grid coordinates of the farm, not the precise sampling locations from which each sample was collected. So that we could make comparisons between (i) the FD and the other national-scale survey datasets (Table 1) and (ii) FD TIs over time, we split the FD into two equal time periods (years 2004–9 and 2010–15). We computed mean values for each of the two agricultural groups over the first period and considered that it was justifiable to compare these to both the CS2007 (sampled in 2007) and the LUCAS survey data (sampled in 2009). The most recent survey for which we could use data on extractable Mg was the resampling of the National Soil Inventory (Kirk *et al.*, 2009), with sampling between 1994 and 2003.

In the FD, each farm had many samples in the database for the same agricultural group and for the same time period. For our model-based analysis we had only one set of coordinates for each farm from which many samples were typically collected, and so we computed a single TI value for each farm. We computed both mean and median values for each farm for different nation and land-cover combinations, and compared their frequency distributions and summary statistics. Both the frequency distributions and summary statistics for the mean and median values were similar for each indicator, and because the median is less sensitive to extreme values (outliers) we used it in our subsequent analyses of the FD. We then used a spatial overlay to classify each farm location to be within

either England or Wales with polygons of each nation's borders. After processing the data we had 20 separate datasets for England (5 indicators \times 2 agricultural groups \times 2 time periods) and 10 for Wales (5 indicators \times 1 grassland group \times 2 time periods). A small area of Wales only is under A&H production; therefore we excluded this from our analysis.

National-scale survey data

We used data from four other national-scale surveys to compare topsoil indicator values with the FD; selected features of the surveys are listed in Table 1.

Comparison of analytical methods between farmers' data and other surveys

Here we describe salient features of the analytical methods and any significant differences between the FD and the other three national surveys. The analytical methods for the RSSS were the same as for the FD; Oliver *et al.* (2006) provided estimated mean values of nutrient TIs across representative farms of E&W for 2001.

Soil pH. In all surveys soil pH is determined in water at a soil:solution ratio of 1:2.5. In the FD and LUCAS methods, air-dried soil is used to determine pH, whereas for CS2007 the soil is field-moist when pH is measured.

Available (Olsen) P. For the FD, NRM Laboratories use the MAFF (1986) method for Olsen P (OP) measurements, which is slightly different from the two other surveys (with contemporaneous data) in that: (i) it includes addition of polyacrylamide (not activated charcoal; AC) as a decolourant in the extraction stage, (ii) there might be small differences in the type of filter paper used and (iii) the first 10 ml of filtrate is discarded by some laboratories. The LUCAS method (ISO, 1994) includes the addition of 1 g AC. The CS2007 method does not include addition of a decolourant (Emmett *et al.*, 2010). In humus-rich soil, OM is dissolved during the extraction procedure, which can be addressed by adding a

decolourant to the solution (either polyacrylamide or AC) during shaking or after filtration. To investigate whether this methodological difference can lead to systematic bias between laboratories we analysed 24 soil samples with a wide range of OP concentrations by (i) the method described in the Appendix (AC as a decolourant) and (ii) the RB427 protocol adopted by NRM Laboratories (MAFF, 1986), which uses polyacrylamide. We made a scatterplot from both sets of analyses and fitted a regression model by ordinary least squares to the data to assess the strength of their linear relation and the magnitude of any systematic bias from the 1:1 line.

Available K. The LUCAS survey method uses ammonium acetate (pH 7) for extraction of K, whereas both the FD and the RSSS (Oliver *et al.*, 2006) method use ammonium nitrate.

Organic matter. Both CS2007 and FD use loss-on-ignition (LOI) to determine the OM content in topsoil; the former method has a combustion temperature of 375°C for 16 hours, whereas the latter temperature is 430°C for 12 hours. The LUCAS survey measured organic carbon (%) (ISO, 1995) and we converted this to OM by multiplying the value by two (Pribyl, 2010).

Statistical analyses: survey comparisons

The different sampling approaches used in the various survey datasets mean that we must use different statistical approaches to estimate their global means (and confidence intervals) for TIs at the national scale. The purposive and grid-based sampling of the FD and resampling of the National Soil Inventory, respectively, require a model-based (geostatistical) analysis. The random sampling adopted in CS and LUCAS surveys can be analysed by a design-based approach.

Model-based analysis. We used ordinary block kriging (de Gruijter *et al.*, 2006) to estimate mean values (\tilde{Z}_j) for each land-cover class (j) and confidence intervals for each topsoil indicator for the three land-cover and nation combinations. In each case we created single polygons over which to make block kriged predictions. We used ArcGIS (ESRI) to digitize a polygon around the 1-km squares of either the A&H or IG-dominant broad habitat types of the Land Cover Map 2007 (Centre for Ecology & Hydrology, 2011) for E&W. Within each of these three main polygons, we delineated internal polygons where other broad habitats dominated. This created holes to exclude these regions from within each main polygon. We then used the `spsample` function from the `sp` package in R to create a regular grid of 5000 prediction locations across each main polygon for block kriging.

Where a variate had a skewness coefficient in the range $[-1, 1]$ we considered it to be sufficient to compute variograms without data transformation (Webster & Oliver, 2007). If the frequency distribution was strongly skewed, we applied a Box–Cox transformation with a lambda value that minimized the magnitude of the skewness coefficient. During exploratory analyses, we observed no evidence

for strong anisotropy in the FD and so we estimated isotropic variograms of each variate. For those variates with more than 1000 samples we used Matheron's method of moments to estimate semi-variances (using the `variogram` function in the R package `gstat`; Pebesma, 2004) at increasing lag intervals and fitted either an exponential or spherical variogram model by least squares approximation. For variates with fewer than 1000 samples we estimated the variogram by residual maximum likelihood (REML) with the `likfit` function from the `geoR` package (Diggle & Ribeiro, 2007). For all variates we used a modified version of the `krige0` function (see Supporting Information) in the `gstat` package to predict by kriging and used the covariance matrix to estimate the uncertainty. The parameters of all the variogram models fitted to each of the variates (either with or without transformation of the original data) are reported in the Supporting Information for this paper.

Predictions of transformed variates will eventually require back-transformation to their original units. If the inverse of the transformation is applied to a kriged prediction, the result will be a biased estimate of the mean in the original units because the pdf of the untransformed variate is not symmetric (Webster & Oliver, 2007). There is no simple formula to correct for this bias when back-transforming the mean across a block (Cressie, 2006). Therefore, we avoid this problem by simulating 1000 conditional realizations of the transformed variate on a dense grid that covers the block evenly. We used a modified version of the `krige0` function in the R `gstat` package to calculate kriged predictions of the transformed variate at each node of the grid and to calculate the covariance matrix of the prediction errors. The modifications ensured that the formula to calculate the covariance matrix was consistent with those used by Marchant *et al.* (2009). We simulated 1000 realizations of the prediction errors by the LU approach (Webster & Oliver, 2007) and then added the predicted values to produce 1000 simulated realizations of the TI values at each point on the dense grid that are consistent with the modelled spatial correlation.

We applied the inverse transform to each element of each realization and calculated the mean of each realization. We treated the resulting 1000 mean values as an independent sample of the block mean. Their mean was our estimate of the block mean. The 95% confidence interval stretches between the 25th and 975th largest values of the sample of block means. The use of simulations rather than block kriging ensures that predictions of the mean will be unbiased even when the variable has been transformed. This approach disregards the small component of uncertainty that will arise because of the difference between the empirical variograms of the observed variable and of the realized simulations.

Design-based analysis. Within the strata of the CS and LUCAS surveys, sampling locations were selected independently and with uniform inclusion probability densities. For CS2007, 45 classes were established based on a combination of land use, landscape settings and topographic features. For the LUCAS survey, soil sampling locations were selected from five main land-use classes (Tóth *et al.*, 2013). We estimated the spatial mean of the land-cover

class, \hat{z}_j , for the target variable z by:

$$\hat{z}_j = \frac{1}{n} \sum_{i=1}^n z_i, \quad (1)$$

where n is the sample size, z_i is the value at location i and j is the land-cover class. To compute a 95% confidence interval around the mean we first computed the sampling variance of the estimated mean, $\hat{V}(\hat{z}_j)$:

$$\hat{V}(\hat{z}_j) = \frac{1}{n(n-1)} \sum_{i=1}^n (z_i - \hat{z}_j)^2. \quad (2)$$

We then computed the 95% confidence interval for \hat{z}_j as:

$$\hat{z}_j \pm 1.96 \cdot \sqrt{\hat{V}(\hat{z}_j)}. \quad (3)$$

For topsoil indicators with small numbers of samples, computing the confidence intervals in this way assumes that the sample mean is approximately normally distributed. Where the frequency distribution deviated significantly from normality, we transformed the data before calculating the confidence limits. We then back-transformed these limits to the original scale.

Change in topsoil indicators: farmers' data

We selected data for those farms that had analyses for TIs in both the first (2004–9) and second (2010–15) of two time periods. We computed the absolute change (values from the second period minus the first) in indicator values and also the rate of change (per year) based on the number of years between sampling, and we plotted histograms of these data for pH and available K, OP and Mg so that we could observe the range of any change and whether or not it was centred around zero. To assess the average change across the block we paired the 1000 simulated values of the block mean randomly for each time period and calculated the difference between them. Our estimate of the change is the mean of these differences. The 95% confidence intervals can be estimated by sorting the differences and extracting the 25th and 975th largest values. The probability of an increase (or decrease) in the variate can be estimated from the proportion of differences that are greater (or less than) zero. This approach disregards any cross-correlation between the observations for the different time periods. This correlation could have been considered by estimating a linear model of coregionalization and then performing cokriging (Marchant & Lark, 2007). The cross-correlation is likely to cause the confidence intervals to narrow.

We also used ordinary kriging to create maps of topsoil pH for the FD for the two time periods (2004–9 and 2010–15) with the `krige0` function. We made predictions on a 1-km grid across a single polygon that combined the two land-cover combinations (A&H and IG) for both E&W, but in which there were gaps relating to other land-use types.

Coregionalization: farmers' data and survey data

It may be possible to use the more densely sampled FD to create maps of the national-scale survey data with greater accuracy if there is strong joint spatial correlation between the two survey datasets. The strength of their coregionalization can be explored by computing the cross variogram between two variates (Webster & Oliver, 2007). We explored the spatial cross-correlation for topsoil pH between the FD and the LUCAS survey data. We could not do this for the CS2007 survey data because the coordinates must remain confidential to CS staff. We extracted the FD for 2007–11 because this date range was within 2 years of the LUCAS survey samples (2009). Neither the LUCAS nor the FD topsoil pH were strongly skewed so we used the original (untransformed) data to compute semivariances at a lag interval of 10 km to a maximum lag distance of 200 km for the LUCAS survey samples and FD, and also their cross variogram with the `variogram` function in the R package `gstat`. The sampling sites for the FD and LUCAS surveys do not coincide and so we computed their pseudo cross variogram (Webster & Oliver, 2007). We plotted the results to determine the strength of any spatial coregionalization and fitted a linear model of coregionalization with the `fit.lmc` function of the `gstat` package. If there is strong coregionalization in the data we could then use the denser sampling of the FD to reduce the uncertainties in mapped estimates of the LUCAS data by cokriging (Marchant & Lark, 2007).

Results

Survey comparisons

Figure 1 shows the farm locations of the FD for both A&H and GR for the two time periods and also where data from farms are common to both time periods for four TIs (pH, P, K and Mg). The locations of these farms are distributed widely across both E&W. The total land area for the three polygons that delineate the dominant land-cover types (A&H or IG) across either England or Wales on which the block kriged estimates were made are shown in Table 2.

Topsoil pH. Table 3 shows the estimates of mean topsoil pH for the three combinations of nation and land use for three surveys. Estimates for the LUCAS survey and FD are similar in each case; these surveys adopted the same analytical method for pH with air-dried soil in water. The CS2007 mean estimates are larger than the other surveys in each case by around 0.5 pH units for arable soil across England, and 0.25–0.4 or 0.14–0.2 units for grassland across England and Wales, respectively. Part of this difference might arise from the difference in analytical method for the CS2007 topsoil samples, which are measured in a field-moist condition. The 95% confidence intervals for the FD are smaller than for the other surveys because of the substantially greater number of samples on which measurements were made.

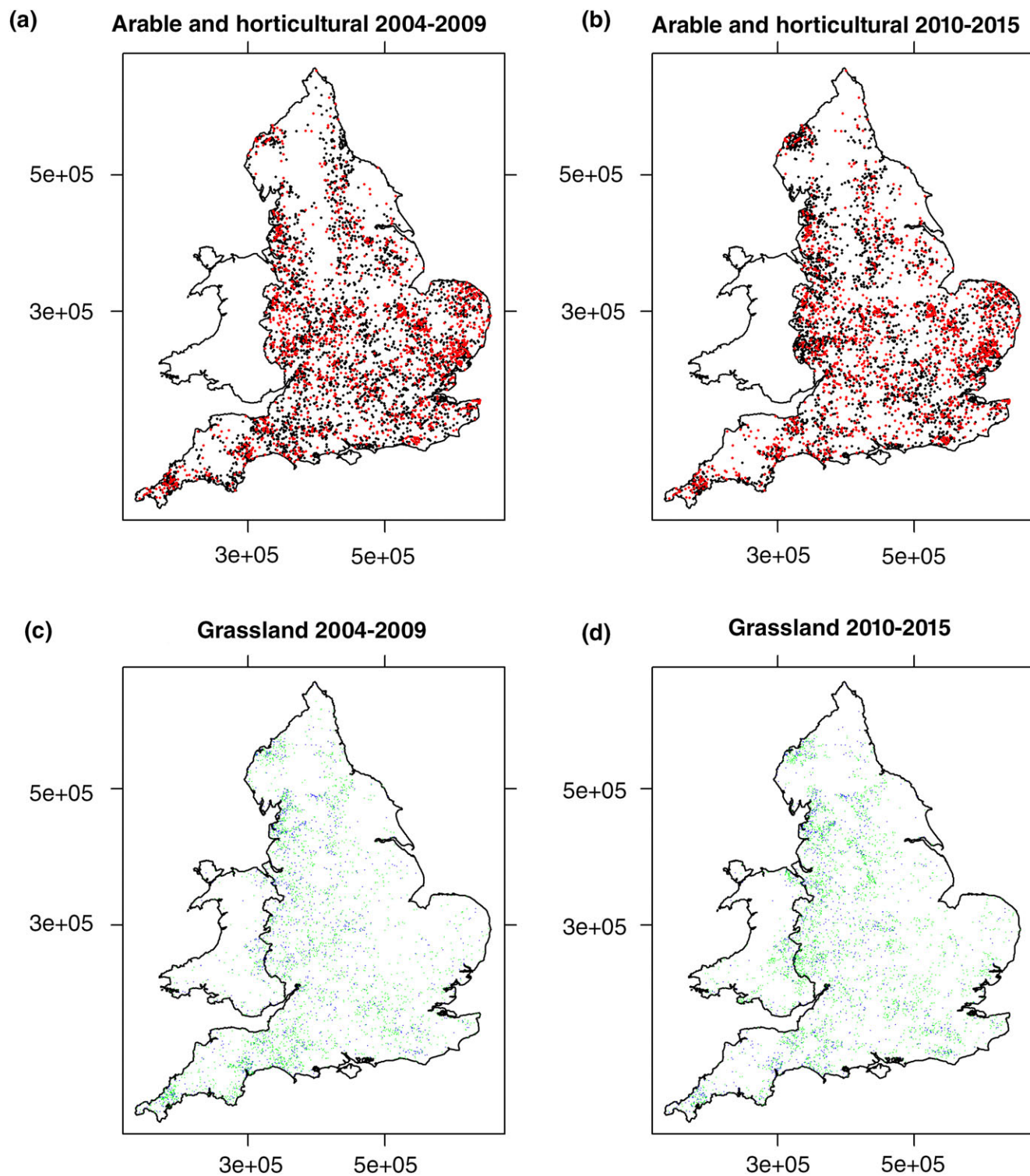


Figure 1 Farm locations of the farmers' data used in this study for four topsoil indicators (pH, Olsen P, available K and available Mg) for two time intervals (2004–9 (a and c) and 2010–15 (b and d)) and two agricultural groups. The red (arable and horticulture) and blue (grassland) dots are farms with data for both time intervals. The black (arable and horticulture) and green (grassland) dots show sites with data from one time interval, for the dates shown. Coordinates are metres on the British National Grid.

Table 2 The land area (km²) for the three combinations of land cover and nation over which block kriged estimates were computed for the topsoil indicators based on polygon delineation of dominant 1-km land cover of the Centre for Ecology and Hydrology Landcover Map (Centre for Ecology & Hydrology, 2011)

Land cover and nation	Land area / km ²
Arable and horticulture, England	88 207
Improved grassland, England	48 954
Improved grassland, Wales	14 008

Table 3 Mean topsoil pH values, 95% confidence intervals (CIs) and *n* samples for three national-scale datasets across England and Wales for arable (and horticulture) or improved grassland

Quantity	Arable England	Grass England	Grass Wales
Mean pH CS2007	7.56	6.51	5.93
% CI CS2007	7.48–7.64	6.41–6.61	5.83–6.03
<i>n</i> CS2007	440	301	147
Mean pH 2004–9	7.00	6.22	5.82
% CI 2004–9	6.97–7.03	6.19–6.25	5.75–5.89
<i>n</i> farms 2004–9	4433	3498	248
Mean pH LUCAS	7.03	6.11	5.73
% CI LUCAS	6.94–7.12	5.97–6.24	5.52–5.96
<i>n</i> LUCAS	260	212	42
Mean pH 2010–15	6.96	6.25	5.83
% CI 2010–15	6.93–6.99	6.22–6.28	5.76–5.89
<i>n</i> farms 2010–15	4061	3722	395

CS2007, Countryside Survey; LUCAS, EU-wide LUCAS survey; 2004–9 and 2010–15 are farmers' data for year ranges shown.

Olsen P. Figure 3 shows the differences between OP concentrations for two analytical methods with the addition of AC or polyacrylamide as decolourants. There is a clear systematic bias between the methods. The method in which AC is used has larger OP concentrations in all but one of the 24 samples than for those analysed with polyacrylamide. The regression model fitted to the data has an adjusted *R*² value of 0.98, a slope of 1.32 and an intercept of –0.13, which shows there is a consistent bias between the two methods (Figure 2). On average, the method with AC has OP concentrations that are 32% larger than the method with polyacrylamide.

Table 4 shows the estimates of mean topsoil OP concentrations for the three combinations of nation and land use for four national scale surveys. The estimated mean values for A&H soil of England for both the CS2007 and LUCAS surveys are substantially larger (between 15.6 and 19.6 mg kg^{–1}) than those of the FD. For the LUCAS survey, part of this difference probably results from differences in the analytical methods, one of which is the different decolourant used in (i) FD (polyacrylamide) and (ii) the LUCAS surveys (AC). The larger mean OP concentrations observed in the CS (no decolourant added) than in FD (polyacrylamide added) might in part result from this difference in analytical method. However, we cannot state that the decolourant is a cause of the

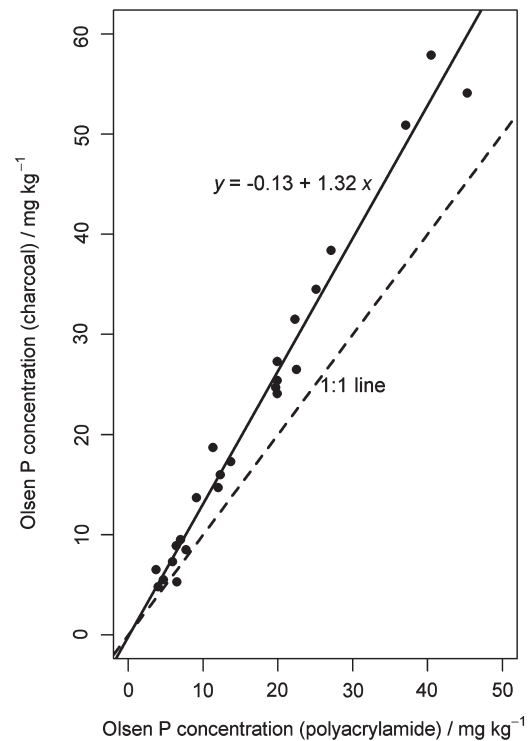


Figure 2 Scatterplot of Olsen P concentrations (mg kg^{–1}) for 24 topsoil samples using two analytical methods with small differences in their protocols. The values plotted on the *x*-axis are for the RB427 method (polyacrylamide added as a decolourant) and those values plotted on the *y*-axis are for the method that included the addition of 0.05–0.1 g of activated charcoal as a decolourant.

Table 4 Mean topsoil available (Olsen) P concentrations (mg kg^{–1}), 95% confidence intervals (CIs) and *n* samples for soil datasets across England and Wales for arable (and horticulture) or improved grassland

Quantity	Arable England	Grass England	Grass Wales
Mean CS2007	43.5	26.8	21.8
95% CI CS2007	36.8–50.2	22.5–31.1	16.5–26.6
<i>n</i> CS2007	197	123	47
Mean 2004–9 ^a	27.9	24.8	22.1
95% CI 2004–9	24.1–25.5	24.1–25.5	0.5–23.9
<i>n</i> farms 2004–9	4433	3498	248
Mean LUCAS	47.1	39.9	36.7
95% CI LUCAS	42.2–51.4	34.9–44.4	28.9–43.8
<i>n</i> LUCAS	260	212	42
Mean 2010–15 ^a	27.5	23.6	22.4
95% CI 2010–15	26.7–28.3	22.9–24.4	21.1–23.9
<i>n</i> farms 2010–15	4061	3722	395
Mean RSSS 2001 ^a	–	25.9	–
<i>n</i> RSSS 2001	–	676	–

^aUsed the same analytical method from RB427 (MAFF, 1986). RSSS, representative soil sampling scheme 2001, is the mean across all three combinations of nation and land use (Oliver *et al.*, 2006); 2004–9 and 2010–15 are farmers' data ^a for year ranges.

Table 5 Mean topsoil available K concentrations (mg kg^{-1}), confidence intervals (CIs) and n samples for three soil datasets across England and Wales for arable (and horticulture) or improved grassland

Quantity	Arable England	Grass England	Grass Wales
Mean K 2004–9 ^a	181	158	138
% CI 2004–9	177–185	153–162	130–147
n farms 2004–9	4433	3498	248
Mean LUCAS	262	239	258
% CI LUCAS	236–283	203–269	162–335
n LUCAS	260	212	42
Mean K 2010–15 ^a	177	157	149
% CI 2010–15	173–182	153–162	139–160
n farms 2010–15	4061	3722	395
Mean RSSS 2001 ^a	–	178	–
n sites	–	676	–

^aSame extractant (ammonium nitrate).

LUCAS, EU-wide LUCAS survey; RSSS, representative soil sampling scheme 2001 across all three combinations of nation and land use (Oliver *et al.*, 2006); 2004–9 and 2010–15 are farmers' data ^a for year ranges.

Table 6 Mean topsoil available Mg concentrations (mg kg^{-1}), confidence intervals (CIs) and n samples for three soil surveys across England and Wales for arable (and horticulture) or improved grassland

Quantity	Arable England	Grass England	Grass Wales
Mean Mg 2004–9	120	144	139
95% CI 2004–9	116–125	140–148	130–149
n farms 2004–9	4433	3498	248
Mean Mg NSI	160	190	140
95% CI NSI	147–173	178–202	129–152
n NSI	709	750	187
Mean Mg 2010–15	123	152	150
95% CI 2010–15	119–128	147–157	140–160
n farms 2010–15	4061	3722	395
Mean RSSS 2001	–	123	–
n sites	–	676	–

NSI, National Soil Inventory resampling; RSSS, representative soil sampling scheme 2001 across all three combinations of nation and land use (Oliver *et al.*, 2006); 2004–9 and 2005–10 are farmers' data for year ranges. All surveys use the same analytical method.

bias because we would need to make measurements on the same soil samples and also control for other differences in the methods used. The large differences between mean OP concentrations in arable topsoil across England from LUCAS (47.1 mg kg^{-1}), CS (43.5 mg kg^{-1}) and FD ($27.5\text{--}27.9 \text{ mg kg}^{-1}$) might be caused partly by sampling bias. We cannot estimate accurately the magnitude of the analytical bias because the quantities of decolourants added to the extractions were different for each of the methods used, including the method we adopted in the Appendix. The mean OP concentration for all three combinations of land use and nation from the RSSS survey in 2001 (25.9 mg kg^{-1}) is similar to the mean estimates from the FD (range $22.1\text{--}27.9 \text{ mg kg}^{-1}$) over both time periods. For IG across England the FD and CS2007 mean values

Table 7 Mean topsoil organic matter concentrations (%), confidence intervals (CIs) and n samples for three soil datasets across England and Wales for arable (and horticulture) or improved grassland

Quantity	Arable England	Grass England	Grass Wales
^a Mean OM CS2007	5.10	9.72	10.3
95% CI CS2007	4.59–5.50	8.77–10.5	9.1–11.3
n CS2007	440	301	146
^b Mean OM LUCAS 2009	5.64	13.1	12.2
95% CI LUCAS	5.2–6.0	10.2–13.4	7.8–15.1
n LUCAS	260	212	42
^a Mean OM 2010–15	4.66	8.65	–
95% CI 2010–15	4.34–4.99	7.87–9.49	–
n farms 2010–15	244	209	18

^aLoss-on-ignition (LOI) temperature was 375°C for CS2007 (for 16 hours) and 430°C (12 hours) for farmers' data.

^bComputed as $2\times$ organic carbon concentration determined by combustion at $> 900^\circ\text{C}$ and detection of carbon (ISO 10694:1995).

CS2007, Countryside Survey; LUCAS, EU-wide LUCAS survey; 2004–9 and 2010–15 are farmers' data for year range; OM, organic matter.

Table 8 Mean absolute change and its statistical significance for four topsoil indicators using farmers' data for the three combinations of land cover and nation (arable and horticulture (A&H) and improved grassland (IG)) for either England or Wales between two time periods (2004–9 and 2010–2015)

Indicator	Mean change	P -value of change
A&H England		
pH	–0.04	0.03
Olsen P / mg kg^{-1}	–0.4	0.21
Available K / mg kg^{-1}	–4.0	0.11
Available Mg / mg kg^{-1}	+3.0	0.18
IG England		
pH	+0.03	0.32
Olsen P / mg kg^{-1}	–1.2	0.50
Available K / mg kg^{-1}	–1.0	0.42
Available Mg / mg kg^{-1}	+8.0	0.01
IG Wales		
pH	+0.01	0.48
Olsen P / mg kg^{-1}	+0.3	0.4
Available K / mg kg^{-1}	+11.0	0.07
Available Mg / mg kg^{-1}	+11.0	0.05

are similar, although the CS2007 mean values are larger (by 3.2 and 2 mg kg^{-1}) than the FD mean values for the two time periods, which might be affected by analytical bias. The mean values for CS2007 and FD for IG in Wales are similar (maximum difference 0.6 mg kg^{-1}).

Available K. Table 5 shows the estimates of mean topsoil available K concentrations for the three combinations of nation and land use for three national-scale surveys. In each of these three combinations, the mean values from the LUCAS survey were substantially larger than those for the FD (difference $81\text{--}120 \text{ mg kg}^{-1}$); we

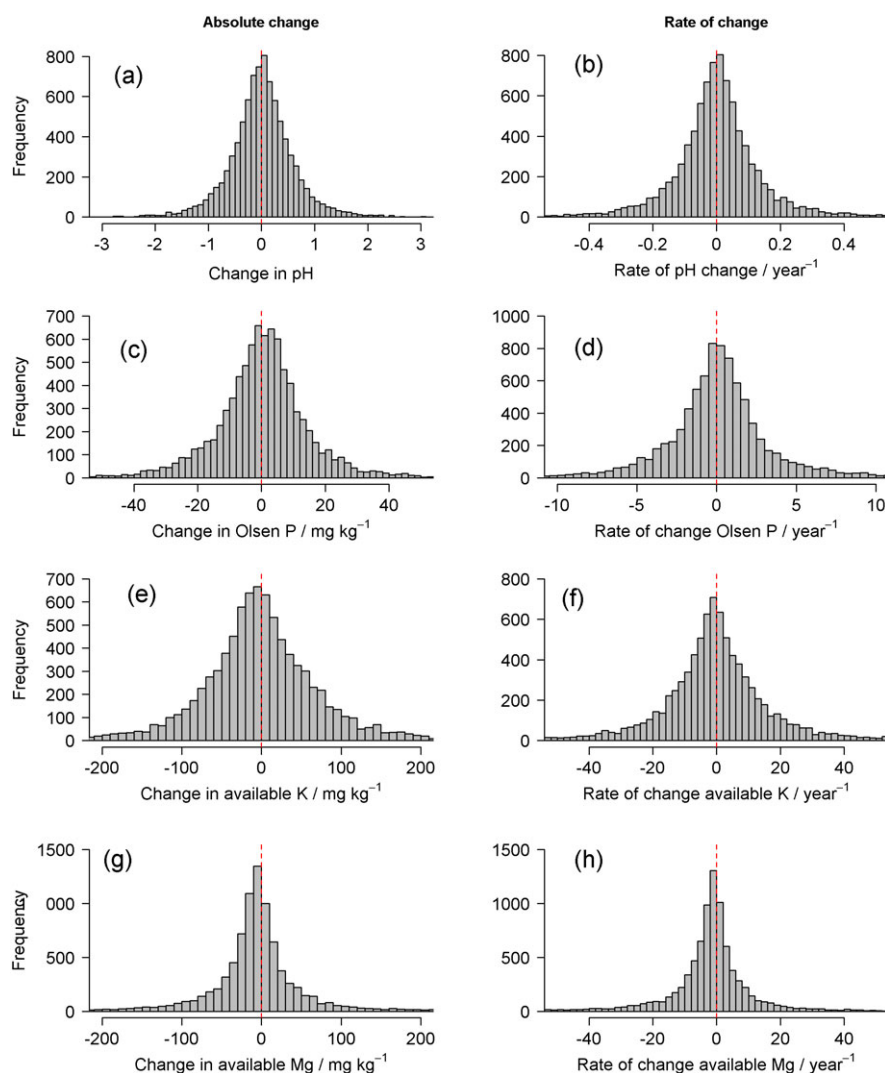


Figure 3 Histograms of absolute change and rates of change in four topsoil indicators for farmers' data for arable and horticultural land for England in which farms have data for both of two time periods (2004–9 and 2010–15): (a,b) pH, (c,d) Olsen P (mg kg^{-1}), (e,f) available K (mg kg^{-1}), (g,h) available Mg (mg kg^{-1}). The vertical red line is at zero on the x-axis.

attribute this to the different extractants. However, the average concentration of available K in the soil sampled as part of the RSSS in 2001 was 173.1 mg kg^{-1} (Oliver *et al.*, 2006), which is similar to the estimated mean values reported for the three combinations of nation and land use from the FD (range $138\text{--}181 \text{ mg kg}^{-1}$), both of which used the same extractant (ammonium nitrate).

Available Mg. Table 6 shows the estimates of mean topsoil available Mg concentrations for the three combinations of nation and land use for FD and three national-scale surveys. The estimated mean available Mg concentrations for the NSI resampling (2003) are substantially larger (difference range $40\text{--}46 \text{ mg kg}^{-1}$) for England (A&H and IG) than the FD. We cannot account for this difference because the same analytical methods were used in both laboratories. Based on quality controlled unpublished data (NRM Laboratories, 6 June 2016) from the WEPAL laboratory comparison scheme we know that values reported from the NRM Laboratories are not systematically biased from the overall mean of several other laboratories. For IG in Wales, the NSI resampling and

FD for the two periods have similar mean concentrations of 140, 139 and 150 mg kg^{-1} , respectively. The estimated mean Mg concentration (123 mg kg^{-1}) from the RSSS (2001) for all three combinations of nation and land use is similar to the mean Mg values for A&H soil across England for the FD for the two time periods ($120\text{--}123 \text{ mg kg}^{-1}$).

Organic matter. Table 7 shows the estimates of mean topsoil OM contents (%) for three combinations of nation and land use for FD and three national-scale surveys. For both land-use types across England the concentrations decrease in the order $\text{LUCAS} > \text{CS2007} > \text{FD 2010–15}$. We cannot assess whether the larger mean values in the LUCAS survey than in CS2007 and FD result from the difference in analytical methods (carbon detection following combustion in the LUCAS survey and LOI for CS2007 and the FD). For the CS2007 and FD (2010–15) for arable land and grassland across England, the 95% confidence intervals of the mean values overlap, indicating that the differences between them might not be statistically significant.

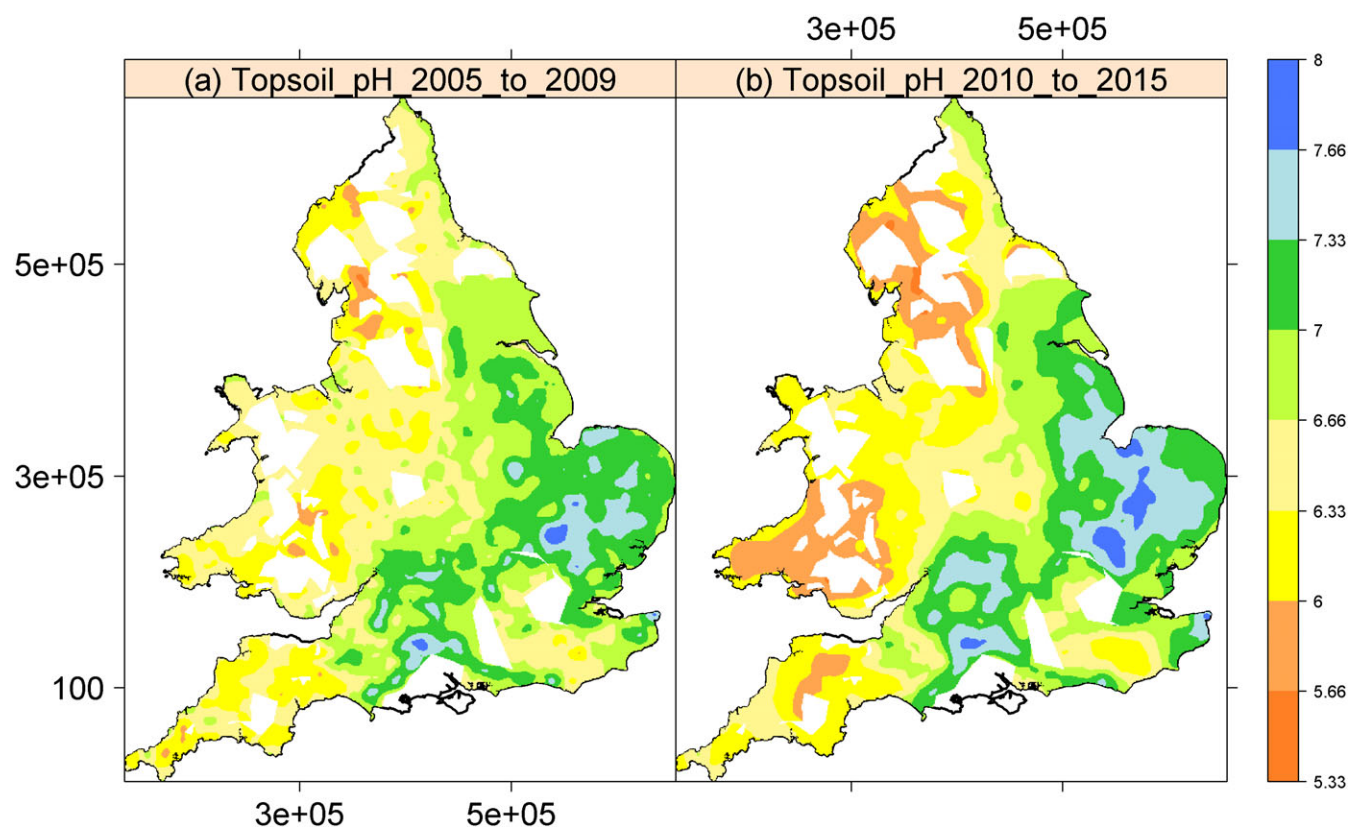


Figure 4 Maps of topsoil pH for arable and horticulture and improved grassland land-cover types of the farmers' data on a 1-km grid across England and Wales for two time periods (2004–9 (a) and 2010–15 (b)). Areas of land cover other than A&H or IG are shown in white. Coordinates are metres on the British National Grid.

Change in farmers' data

Figure 3 shows histograms of change and also rates of change in four TIs for A&H land use of England between the two selected time periods for FD farms with samples from both periods (see also Figure 1). In each case the absolute change and rates of change show an approximately normal distribution centred around zero. This suggests that overall changes in the TIs are likely to be small at the national scale.

The model-based estimates of change in TI between the two time periods (Table 8) show that for these four indicators only the reduction (0.04 pH units) in topsoil pH (P -value 0.03) for A&H soil across England and the increase in available Mg (P -value 0.01) for IG across England (increase of 8 mg kg^{-1}) were statistically significant. There are some clear changes in topsoil pH for different regions of E&W for the two periods (2004–9 and 2010–15) of FD (Figure 4). In particular, there is a significant increase in the area of topsoil with pH values < 6 in parts of northern England, Wales and parts of Devon (southwest England).

Coregionalization of topsoil pH

Variograms of the LUCAS survey (2009) and FD (2007–11) for topsoil pH (Figure 5a,b, respectively) show autocorrelation (the

semivariates increase with increasing lag distance) to separating distances of around 100 and 200 km, respectively. The semivariates of the FD topsoil pH ($n = 11\,229$) increase more smoothly with distance than those of the LUCAS data ($n = 567$), partly because there are many more sample locations (farms) in the former. The cross variogram (Figure 5c) shows that the two datasets also show strong cross-correlation to a distance of around 120 km, including a shorter range structure of around 20 km lag distance. The parameters of the linear model of coregionalization fitted to the variograms and cross variogram are given in Table 9.

Discussion

In general, we found reasonable agreement between estimated mean values of TIs across national and land-cover combinations for FD and representative national-scale surveys. Given that these TIs exert a strong effect on soil fertility, land managers have good reason to keep them within narrow limits. We provided evidence above to show that in one case attributable bias from the difference in analytical methods partly accounted for the discrepancies in mean estimates between FD and national-scale surveys. We consider that FD offer a potential means for monitoring the soil indicators examined at the national scale that complement,

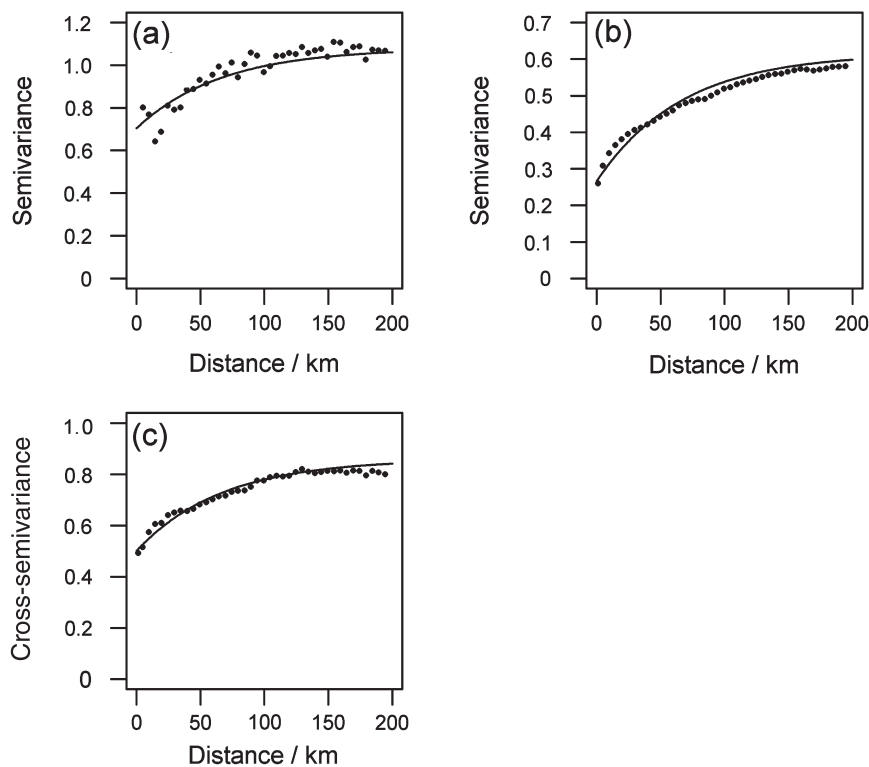


Figure 5 Semivariance and cross-semivariance estimates of topsoil pH across England and Wales for both arable and horticulture and improved grassland land-cover types for: (a) LUCAS survey (2009), (b) farmers' data (FD) (samples collected 2007–11) and (c) their cross variogram.

Table 9 Parameters of the linear model of co-regionalization fitted to the farmers' data and LUCAS survey data (see Figure 5)

	Model	Sill semivariance	Model range / m
LUCAS	Nug	0.70	0
LUCAS	Exp	0.38	66 667
Farmers' data	Nug	0.27	0
Farmers' data	Exp	0.35	66 667
Cross variogram	Nug	0.50	0
Cross variogram	Exp	0.36	66 667

Nug, nugget; Exp, exponential.

but do not replace, monitoring with statistically designed surveys. Unbiased, statistically designed surveys are necessary to assess the magnitude of any unattributable bias introduced into FD.

In a study that assessed changes in topsoil pH between 1978 and 2003, Kirk *et al.* (2009) reported annual increases in soil pH of up to 0.04 units for parts of E&W, which they attributed partly to reductions in sulphur deposition. Assuming the FD for topsoil pH are robust, which is supported by their strong similarity to the LUCAS survey, the FD show a small overall reduction in topsoil pH for A&H land-cover types between 2004–9 and 2010–15. The larger number and spatial distribution of measurements in FD mean that we can identify changes in topsoil properties at the regional scale, which cannot be detected by the sparser sampling of representative surveys. For example, our maps based on FD highlight reductions in topsoil pH to below the recommended

threshold (pH = 6) for optimum yield (Defra, 2010), which might reflect reductions in the quantities of lime applied in these areas.

We identified an attributable source of bias from differences in analytical methods used in the FD and other surveys for the measurement of OP concentrations, which we suggested might result from the use of different decolourants. In the method for the FD, OP concentrations were typically smaller than those for the LUCAS survey. This has important implications because the OP concentrations (and their associated index values) from FD are used to estimate the quantities of P fertilizer applied at the field scale to ensure optimal yields (Defra, 2010). Further research is required to determine which features of the OP analytical methods, including whether a decolourant is used and its form, lead to the bias we observed. There is also a need to determine which of these features reflects more accurately the capacity of the soil to supply P to the soil solution because it largely determines its uptake by crop roots, which then affects crop yield.

Of our selected TIs, we found that OM content had far fewer measurements in our FD than the other four indicators, although there were substantially more in the second period (2010–15) than the first. To improve our mean estimates of topsoil OM concentration by including more measurements, we could investigate whether laboratories other than NRM laboratories would provide their soil measurements for subsequent data analysis. However, mean estimates by the LUCAS and CS2007 surveys suggested that analytical differences (LOI compared with determination of organic carbon by combustion) in estimating OM content might have led to biased estimates, and this would need to be accounted for when combining

these data. This could be problematic if the main cause of observed differences between these two methods is the loss of structural water from clay minerals (Kucirik *et al.*, 2016), the amount of which is typically unknown in measurements on FD samples.

It is fortunate that the soil sampling depth (0–15 cm) for most of the national-scale surveys and the FD were similar. There is field-based evidence from cultivated land in parts of England that changes from conventional to minimum or no tillage can lead to the formation of two distinct horizons in the upper 25 cm of the solum (Mike Slater, personal communication); the horizons separate at around 10-cm depth. Therefore, it might be beneficial for farmers to sample and measure the properties of these horizons separately. If it became common for agronomists or farmers to take separate samples at two depths, this information would need to be recorded and accounted for in subsequent statistical analyses of FD. We were able to use only a subset of FD soil measurements from all those available because in many cases either land use or farm address details were missing from the electronic records for many samples. If agronomists and farmers recorded land, crop details and national grid coordinates for every field from which they took soil samples we would have a larger sample database for statistical analyses and smaller geographical error associated with our model-based, spatial analysis.

We chose to split the 12 years of FD into two datasets and develop separate spatial models for the TIs. In future, we consider that developing combined space-time models (De Cesare *et al.*, 2001) for each indicator would have the advantage that new data could be added annually, and these models could be used to estimate mean values and confidence intervals. The strong joint spatial correlation (co-regionalization) we showed for topsoil pH suggests that we could increase the resolution of maps based on statistically unbiased surveys with the more densely sampled FD.

Conclusions

We were able to retrieve a sufficiently large number of geographical coordinates for farms within the FD database to make model-based mean estimates of topsoil indicator values for combinations of nation (E&W) and land cover for two time periods (2004–9 and 2010–2105). However, there were insufficient FD for soil OM content for the first time period to compute mean estimates. Where common analytical approaches were used (e.g. topsoil pH) there was generally good agreement between the estimated mean values for FD and national surveys for different land-cover and nation combinations. The mean topsoil pH values for the FD and LUCAS surveys were similar and also showed strong coregionalization. For other indicators (OP and available K), we identified an attributable source of bias from differences in analytical protocols that are a likely source of the discrepancies between estimated mean values from the FD and national surveys. Such differences could be accounted for in computing national-scale estimates of the TIs from FD. Data from statistically designed surveys are required to estimate any unattributable sources of bias in estimates of TI values from FD.

Our results, and those of other empirical studies, do not remove the need for design-unbiased surveys in soil monitoring.

In general, there were small changes in the four TI values examined (pH, OP, available K and Mg) for FD between the two time periods. Because of the large number of samples ($n > 4000$) in each time period, we detected a small (0.04 units) but statistically significant reduction in topsoil pH for arable soil of England with our model-based approach. By creating maps of greater accuracy with the samples of the FD, we were able to highlight a reduction in topsoil pH between the two time periods for specific regions of E&W. The strong joint spatial correlation in topsoil pH between FD and the LUCAS data shows that one of the potential benefits of the denser sampling of the former is to create more accurate maps of the latter.

Supporting Information

The following supporting information is available in the online version of this article:

Table S1. Model parameters for mean topsoil pH values (no transformation of original data); see Table 3 in main paper.

Table S2. Model parameters for mean topsoil Olsen P values (Box–Cox transformation of original data); see Table 4 in main paper.

Table S3. Model parameters for mean topsoil available K values (Box–Cox transformation of original data); see Table 5 in main paper.

Table S4. Model parameters for mean topsoil available Mg values (Box–Cox transformation of original data); see Table 6 in main paper.

Table S5. Model parameters for mean organic matter content; see Table 7 in main paper.

Acknowledgements

This research was funded through a NERC innovation internship grant (reference number NE/N012380/1). We thank (i) Claire Wood and David Robinson of the Centre for Ecology and Hydrology for providing the Countryside Survey (CS) 2007 data and for discussing the analytical methods used in CS, (ii) Murray Lark for comments on an original draft of the manuscript and (iii) Cranfield University for providing the NSI re-sampling data under licence L0212/00947. This paper is published with the permission of the Executive Director of the British Geological Survey (Natural Environment Research Council). The LUCAS topsoil dataset used in this work was made available by the European Commission through the European Soil Data Centre managed by the Joint Research Centre (JRC), <http://esdac.jrc.ec.europa.eu/>.

Appendix

Description of analytical method

Analytical method for soil Olsen P (OP) soil measurements with activated charcoal (AC) as a decolourant. The analytical protocol

described here was used for the analysis of 24 soil samples with a range of OP concentrations. This protocol includes the addition of AC as a decolourant and the results were compared with the RB427 method (MAFF, 1986) in which polyacrylamide was used as a decolourant. Apparatus:

1. Temperature controlled orbital shaker
2. Polythene bottles 100 ml
3. Polythene filter funnels
4. 250-ml glass conical flasks
5. Fluted Whatman no 42 filter papers (150 mm)
6. Bacofoil

Reagents:

1. 0.5 M sodium bicarbonate solution: dissolve 42 g sodium bicarbonate (NaHCO_3) and 0.72 g sodium hydroxide pellets in 900 ml of 18 M Ω deionized water. Adjust the pH to 8.5 using a saturated solution of sodium hydroxide or concentrated sulphuric acid and make up to 1 l with water. Mix thoroughly. Dissolve 210 g NaHCO_3 /3.6 g NaOH in 5 l) – store in a room at 20°C.
2. Activated charcoal – product name ‘Acros organics’ Activated Charcoal, NORIT SA2 (Fisher Scientific, Waltham, MA, USA).

Procedure (to be carried out in a 20°C laboratory):

1. Weigh 5 g of air-dried soil, ground to pass a 2-mm sieve, into a weighing boat and transfer to a 250-ml conical flask.
2. Add approximately 0.05–0.1 g AC from the tip of a spatula and cover each flask with Bacofoil.
3. Peel back the Bacofoil and add 100 ml 0.5 M sodium bicarbonate, reseal, swirl the flask and immediately place on the shaker. These steps must be carried out speedily to avoid delays that might vary the contact time.
4. Shake the flask in a temperature-controlled cabinet orbital shaker at 20°C for exactly 30 minutes at exactly 2 Hz.
5. Swirl the flask (to mix the extract and soil) and filter immediately through a Whatman no 42 filter paper discarding the first 10 ml of filtrate. Remove the filter paper and whatever solution remains in the filter paper, exactly 45 minutes after discarding the first 10 ml. Between runs it is essential to keep the contact time constant.
6. After filtration the sample should be analysed immediately.
7. If the extracts are strongly coloured, shake the filtered extract with 1 g of AC and filter through a Whatman no 42 filter paper.
8. Bottles should be capped, placed in a tray and taken to the anion laboratory with an analysis information card ready for analysis by Skalar continuous flow.
9. Blanks and in-house standard materials are included in each batch shaken.
10. Duplicate extractions of at least one in every 10 samples should also be included.
11. Sample extracts are analysed on the Skalar continuous flow system.

References

- Arrouays, D., Thorette, J., Daroussin, J. & King, D. 2001. Analyse de représentativité de différentes configurations d'un réseau de sites de surveillance des sols. *Etude Gestion Sols*, **8**, 7–17.
- Arrouays, D., Marchant, B.P., Saby, N.P.A., Meersmans, J., Orton, T.G., Martin, M.P. *et al.* 2012. Generic issues on broad-scale soil monitoring schemes: a review. *Pedosphere*, **22**, 456–469.
- Black, H., Bellamy, P., Creamer, R., Elston, D., Emmett, B., Frogbrook, Z. *et al.* 2008. *Design and Operation of a UK Soil Monitoring Network: Science Report – SC060073*. Environment Agency, Bristol. [WWW document]. URL https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/291678/scho0908bomx-e-e.pdf [accessed on 16 May 2016].
- Centre for Ecology & Hydrology 2011. Land Cover Map 2007 dataset documentation Version 1.0. Centre for Ecology & Hydrology, Wallingford. [WWW document]. URL <http://www.ceh.ac.uk/sites/default/files/LCM2007%20dataset%20documentation.pdf> [accessed on 16 May 2016].
- Cressie, N. 2006. Block kriging for lognormal spatial processes. *Mathematical Geology*, **38**, 413–452.
- De Cesare, L., Myers, D.E. & Posa, D. 2001. Estimating and modelling space-time correlation structures. *Statistics & Probability Letters*, **51**, 9–14.
- Defra 2009. *Safeguarding our Soils – A Strategy for England*. Defra, London. [WWW document]. URL https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/69261/pb13297-soil-strategy-090910.pdf [accessed on 16 May 2016].
- Defra 2010. *Fertiliser Manual (RB209)*, 8th edn. The Stationery Office, London. [WWW document]. URL <http://www.ahdb.org.uk/documents/rb209-fertiliser-manual-110412.pdf> [accessed on 6 May 2016].
- Diggle, P.J. & Ribeiro, P.J. Jr. 2007. *Model Based Geostatistics*. Springer, New York.
- Diggle, P.J., Menezes, R. & Ting-li, S. 2010. Geostatistical inference under preferential sampling. *Journal of the Royal Statistical Society, Series C*, **59**, 191–232.
- Emmett, B.A., Frogbrook, Z.L., Chamberlain, P.M., Griffiths, R., Pickup, R. Poskitt, J. *et al.* 2008. *Countryside Survey*. Technical Report No 03/07, Centre for Ecology & Hydrology, Bristol, UK. [WWW document]. URL http://www.countrysidesurvey.org.uk/sites/default/files/pdfs/reports2007/CS_UK_2007_TR3.pdf [accessed on 16 May 2016].
- Emmett, B.A., Reynolds, B., Chamberlain, P.M., Rowe, E., Spurgeon, D., Brittain, S.A. *et al.* 2010. *Countryside Survey: Soils Report from 2007*. Technical Report No 9/07, CEH Project Number: C03259. NERC/Centre for Ecology & Hydrology, Bristol, UK. [WWW document]. URL http://www.countrysidesurvey.org.uk/sites/default/files/pdfs/reports2007/CS_UK_2007_TR9-revised.pdf [accessed on 16 May 2016].
- Environmental Audit Committee 2016. *Soil Health*. House of Commons, London. [WWW document]. URL <http://www.publications.parliament.uk/pa/cm201617/cmselect/cmenvaud/180/180.pdf> [accessed on 10 October 2016].
- de Groot, J., Brus, D.J., Bierkens, M.F.P. & Knotters, M. 2006. *Sampling for Natural Resource Monitoring*. Springer, Berlin.
- ISO 1994. ISO 11263: 1994. *Soil Quality: Determination of Phosphorus – Spectrometric Determination of Phosphorus Soluble in Sodium Hydrogen Carbonate Solution*. International Organization for Standardization, Geneva.

- ISO 1995. ISO 10694:1995. *Soil Quality: Determination of Organic Carbon and Total Carbon after Dry Combustion (Elementary Analysis)*. International Organization for Standardization, Geneva.
- Kirk, G.J.D., Bellamy, P.H. & Lark, R.M. 2009. Changes in soil pH across England and Wales in response to decreased acid deposition. *Global Change Biology*, **16**, 3111–3119.
- Kucerik, J., Demyan, M.S. & Siewert, C. 2016. Practical applications of thermogravimetry in soil science. Part 4: relationship between clay, organic carbon and organic matter contents. *Journal of Thermal Analysis & Calorimetry*, **123**, 2441–2450.
- MAFF 1986. *The Analysis of Agricultural Materials (RB427)*, 3rd edn. Her Majesty's Stationery Office, London.
- Marchant, B.P. & Lark, R.M. 2007. Estimation of linear models of coregonalization by residual maximum likelihood. *European Journal of Soil Science*, **58**, 1506–1513.
- Marchant, B.P., Newman, S., Corstanje, R., Reddy, K.R., Osborne, T.Z. & Lark, R.M. 2009. Spatial monitoring of a non-stationary soil property: phosphorus in a Florida water conservation area. *European Journal of Soil Science*, **60**, 757–769.
- Marchant, B.P., Villanneau, E.J., Arrouays, D., Saby, N.P.A. & Rawlins, B.G. 2015a. Quantifying and mapping topsoil inorganic carbon concentrations and stocks: approaches tested in France. *Soil Use & Management*, **31**, 29–38.
- Marchant, B.P., Crawford, D.M. & Robinson, N.J. 2015b. What can legacy datasets tell us about soil quality trends? Soil acidity in Victoria. In: *Earth and Environmental Science. IOP Conference Series*, Volume **25**, p. 012015. Institute of Physics, Bristol, UK.
- Merrington, G., Fishwick, S., Barraclough, D., Morris, J., Preedy, N. Boucard, T. et al. 2006. *The Development and Use of Soil Quality Indicators for Assessing the Role of Soil in Environmental Interactions*. Science Report No SC030265, Environment Agency, Bristol. [WWW document]. URL https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/290729/scho0306bkiq-e-e.pdf [accessed on 16 May 2016].
- Oliver, M.A., Archer, J.R., Baxter, S.J., Todd, A.D. & Skinner, R.J. 2006. The representative soil sampling scheme of England and Wales: a statistical analysis of topsoil nutrient status and pH between 1971 and 2001. *Soil Use & Management*, **22**, 372–382.
- Pebesma, E.J. 2004. Multivariable geostatistics in S: the gstat package. *Computers & Geosciences*, **30**, 683–691.
- Pribyl, D.W. 2010. A critical review of the conventional SOC to SOM conversion factor. *Geoderma*, **156**, 75–83.
- Simpson, K. 1983. *Soil: Longman Handbooks in Agriculture*. Longman, Harlow.
- Tóth, G., Jones, A. & Montanarella, L. 2013. *LUCAS Topsoil Survey: Methodology, Data and Results*. Joint Research Centre, Ispra. [WWW document]. URL http://esdac.jrc.ec.europa.eu/ESDB_Archive/eusoils_docs/other/EUR26102EN.pdf [accessed on 16 May 2015].
- de Vos, B. & Cools, N. 2011. *Second European Forest Soil Condition Report Volume I: Results of the BioSoil Soil Survey*. Report No INBO.R.2011.35, Research Institute for Nature and Forest, Brussels. [WWW document]. URL <http://www.flanders.be/en/publications/detail/second-european-forest-soil-condition-report-volume-1-results-of-the-biosoil-soil-survey> [accessed on 6 May 2016].
- Webster, R. & Oliver, M.A. 2007. *Geostatistics for Environmental Scientists*, 2nd edn. John Wiley & Sons, Chichester.