Sampling and analytical plus subsampling variance components for five soil indicators observed at regional scale

B. G. Rawlins\textsuperscript{a,}\textsuperscript{*}, A. J. Scheib \textsuperscript{a}, R. M. Lark \textsuperscript{b} & T. R. Lister \textsuperscript{a}.

\textsuperscript{a}British Geological Survey, Keyworth, Nottingham  NG12 5GG, UK
\textsuperscript{b}Rothamsted Research, Harpenden, Hertfordshire  AL5 2JQ, UK

* Corresponding author: B. G. Rawlins

British Geological Survey
Keyworth
Nottingham NG12 5GG
UK
phone: +44 (0) 115 9363140
Fax: +44 (0) 115 936 3100
e-mail: bgr@bgs.ac.uk

Running head: Variance components for five soil indicators
Summary

When comparing soil baseline measurements with resampled values there are four main sources of error. These are: i) location (errors in relocating the sample site), ii) sampling errors (representing the site with a sample of material) iii) subsampling error (selecting material for analysis) and iv) analytical error (error in laboratory measurements). In general we cannot separate the subsampling and analytical sources of error (since we always analyse a different subsample of a specimen), so in this paper we combine these two sources into subsample plus analytical error. More information is required on the relative magnitudes of location and sampling errors for the design of effective resampling strategies to monitor changes in soil indicators. Recently completed soil surveys of the UK with widely differing soils included a duplicate site and subsampling protocol to quantify ii), and the sum of iii) and iv) above. Sampling variances are estimated from measurements on duplicate samples — two samples collected on a support of side length 20 m separated by a short distance (21 m). Analytical and subsampling variances are estimated from analyses of two subsamples from each duplicate site.

After accounting for variation caused by region, parent material class and land use, we undertook a nested analysis of data from 196 duplicate sites across three regions to estimate the relative magnitude of medium-scale (between site), sampling and subsampling plus analytical variance components for five topsoil indicators: total metal concentrations of copper (Cu), nickel (Ni) and zinc (Zn), soil pH and soil organic carbon (SOC) content. The variance components for each indicator diminish by about an order of magnitude from medium-scale, to sampling, to analytical plus subsampling. Each of the three fixed effects (parent material, land use and region) were statistically significant for each of the five indicators. The most effective way to minimise the overall uncertainty of our observations at sample sites is to reduce the sampling variance.
Introduction

Soil monitoring activity is increasing because regulatory authorities recognise the need to detect whether significant changes are occurring in properties of soil — and their ability to fulfil soil functions — due to a variety of anthropogenic drivers. The basic aim of soil monitoring is to quantify changes in selected soil properties or indicators over time, and to make inferences regarding these changes. When we resample the soil, we want to compare the baseline and resampled data in order to estimate change.

When we compare a baseline measurement of the soil with the resampled value there are various sources of error. A sample site is the small local area for which we obtain a single value of the soil property of interest. This might be a single core but more typically it is a larger area, e.g. a square of sides 20 m, and we sample this by collecting several cores which are aggregated and then subsampled to obtain material for analysis. To resample we return to the site and then repeat this procedure. There are various sources of error in the estimate of change at a sample site that is given by the difference between the baseline and resample values. First, there is location error. If we have not returned exactly to the original sample site then spatial variation between the two sites actually sampled will contribute uncertainty to our estimate. Second, there is what we may call sampling error. If we repeatedly resampled a site (at the same time), by the same procedure, then our different sample averages (obtained with no further sources of error due, for example, to analysis) would vary due to sampling error. The magnitude of the sampling error depends on the sampling procedure; it decreases as we increase the number of cores that are aggregated and from which we take a subsample for analysis.

The variance of these notional repeat sample averages is the sample error variance. In addition to this the material that we sample is thoroughly mixed, but subsampling for analysis introduces uncertainty too (different subsamples would have different values). We call this the subsample error, which introduces subsample error variance. We know that multiple analyses of uniform material will not return exact results and the
variance of such multiple determinations is the analytical variance. Although separable in theory, the subsample and analytical variance cannot be separated in practice. Analytical variance must be estimated by repeat analyses of uniform material, but soil is never entirely uniform and so the resampling and analytical variances will always be confounded.

Two published studies from the UK have attempted to provide estimates of the magnitude of sampling variances at survey sites. First, the relative proportions of between-site, local (duplicate) and analytical variances for total concentrations of 23 elements in topsoil from 21 sites with common soil parent materials across part of northern England were reported by Rawlins et al. (2002). They showed that the local, sampling variance differed considerably for each element. Second, a study published by Defra, (the Department for Environment Food and Rural Affairs, in England and Wales) where sites from the National Soil Inventory were revisited and samples collected at short intervals of 10 and 50 m (Defra, 2003). Based on analyses at ten sites, the authors concluded that there were significant differences between samples taken at these intervals from the original target site for some of the parameters measured, including SOC, pseudo-total Pb and Zn determined by aqua regia extraction.

Both these studies, particularly the latter, have sample sizes much too small to make confident general statements about the importance of different sources of error in determining soil properties. However, their results, while only indicative, are interesting, and provided a motivation to study the problem more thoroughly.

The sampling protocol adopted by the UK Geochemical Baseline Survey of the Environment (G-BASE) project and the recently completed Tellus soil survey of Northern Ireland includes a procedure for quantifying both sampling (short-scale) and analytical variances at 1 in every 100 locations in its systematic, unaligned sampling design.

Data are available on topsoil properties (including five soil indicators) at 196 sites across three large regions of the UK, comprising a diverse range of soil, mean annual
rainfall, parent material and land use types. In such grid-based surveys, it is inevitable that location and sampling errors will be conflated.

After accounting for variation caused by region, parent material class and land use, we have undertaken a nested analysis to determine the relative magnitude of subsampling plus analytical and sampling variance components for five, tier-one topsoil indicators identified in the UK (Environment Agency, 2006). These are total metal concentrations of copper (Cu), nickel (Ni) and zinc (Zn), soil pH and soil organic carbon (SOC) content. We present the results of our findings and comment on their implications for the design and implementation of soil monitoring.

Methods

Study region and surveys

The three regions (Figure 1) comprise a diverse range of soil parent materials and soil types; further details of the soil types in Northern Ireland are provided by Cruickshank (1997), whilst the soils in the study regions of England are described in three regional bulletins (Soil Survey of England and Wales, 1983a-c). The major soil types in each of the three regions are summarized in Table 1. The larger proportion of peat soils in Northern Ireland compared to the two regions in England reflects the wetter environment of the former, where average mean annual rainfall for the vast majority of the region is greater than 1 m, with a minimum of around 0.75 m.

For the soil surveys, a systematic, unaligned strategy was adopted in which one sample was collected from a random location in every other 1 km square of the British or Irish National Grids, subject to the avoidance of roads, tracks, buildings, railways, electricity pylons, and disturbed ground. One in every 100 of these sites was randomly selected and designated a duplicate sampling site at which the following sampling protocol was adopted. The dominant land use was recorded at each duplicate sampling location.

At each sampling site, including those selected for duplicate sampling, five incremental soil samples were collected using a Dutch auger at the corners and centre
of a square with a side of length 20 m and combined to form a composite sample of
approximately 0.5 kg. At each of these five points, any surface litter was removed and
the soil sampled to a depth of 15 cm into the exposed soil. In the case of organic-rich
soil, 5 cm of surface litter was removed and the soil samples were collected from a
depth range of 5 – 20 cm. This composite sample is referred to as duplicate A (DUP
A). At each of the duplicate sites, another composite sample was collected from one of
four squares with the same support. These squares are illustrated in Figure 2. Notice
that the centres of each of them are 21 m from the centre of the square for the original
sample. The second sample square (1 of a possible 4) was selected randomly, and the
same sampling procedure adopted; this composite sample is referred to as duplicate B
(DUP B).

All samples of soil were air-dried in a dedicated temperature controlled oven at
30°C for 2–3 days, disaggregated and sieved to <2 mm. The samples were coned and
quartered and each split into two sub-samples. The two samples from the DUP A
site were labelled DUP A and SUB A, whilst the samples from the DUP B site were
labelled DUP B and SUB B. From each of the four sub-samples, a 50-g sub-sample
was ground in an agate planetary ball mill. The total concentrations of Ni, Cu and
Zn were determined in each sample by wavelength and energy dispersive XRFS (X-
Ray Fluorescence Spectrometry). The samples were prepared for analysis by grinding
12 g of sample material and 3 g of Elvacite 2013 (n-butyl methacrylate copolymer,
Dupont & Co.) in an agate planetary ball mill for half an hour. The mixture was then
pressed using a 25 tonne load into pellets with a thickness of 40 mm. The coefficients
of variation for a soil sample standard were: Ni (0.6 %) Cu ( 1%) and Zn (0.8 %).

Soil pH and SOC were only determined on a subset of 114 sites (a total of 456 sam-
pies) from two of the regions; East Anglia and Northern Ireland. Soil organic carbon
was estimated in each sample using loss-on-ignition analysis by heating a sub-sample to
450°C for eight hours and multiplying the mass difference by 0.58 (Broadbent, 1953).
The coefficient of variation for this method for 174 replicate analyses of a sample stan-
standard was 3.6 %. The pH of each sample was measured using a glass electrode and
Orion 720A meter. To 10 g of the <2 mm sample, 25 ml of 0.01 M calcium chloride
solution was added, the samples were magnetically stirred for one minute and then left
to settle for 15 minutes. Prior to recording the pH value, the suspension was stirred.

Digital layers of the 1:50,000 maps of bedrock geology and superficial deposits
of England, part of DigiMap GB (British Geological Survey, 2006) were combined in
a GIS to form a series of parent material polygons. The parent material code for any
position on the landscape can be found by assigning the code of any superficial deposit
present, or where absent, the code of the bedrock geology. We devised a simple, seven-
fold parent material classification for all soil parent material types. In a GIS we used
a spatial join procedure to associate each soil sampling observation with the relevant
code of the parent material polygons (see Figure 3).

Data analysis

Our data are collected according to a balanced nested sample design. Within every
randomly selected sample location there is a pair of Duplicates (DUP A and DUP B),
and from each duplicate two subsamples are taken and analysed. Differences between
the subsamples (within a duplicate) are due to the sources of uncertainty that we call
analytical and subsampling error. Differences between duplicates within a site are due
to repeating the sampling procedure with a small shift in the position of the sampled
square. In this study we assume that this represents sampling error (although it will
be slightly inflated by the 21-m shift). We require an analysis to estimate the variances
that these different sources of uncertainty will contribute to our values for each site.

There are two kinds of sources of variation in our data. Random effects arise
because they may be regarded as random processes due to our sampling. Coning
and quartering is regarded as a way of randomly subsampling material, the random
selection of one of four possible second squares to sample at each site is a random source
of variation (sampling error variation), and the random selection of sites for this more
detailed investigation introduces what we shall call medium-scale spatial variation —
it is variation within the land uses, parent materials and regions. The second source of variation arises from fixed effects which we chose to sample, including land uses, parent materials and regions. They do not enter into our sampling in a random way, but because we chose to sample them.

Let us consider one of our observations. It is subsample \( k \) of duplicate \( j \) at site \( i \) randomly selected from sites in landuse \( c \), parent material \( b \) within region \( a \). Because of our random sampling we regard that observation as a random variable \( Z_{a,b,c,i,j,k} \). We can write an expression for this variable,

\[
Z_{a,b,c,i,j,k} = \mu_{a,b,c} + Q_i + R_{i,j} + S_{i,j,k},
\]

where \( \mu \) is the overall mean of the variable for landuse \( c \) within parent material \( b \) within region \( a \), \( Q_i \) is the difference between the mean for the \( i \)th randomly located site and the mean \( \mu_{a,b,c} \), \( R_{i,j} \) is the difference between the mean for duplicate \( j \) within site \( i \) and the mean for site \( i \), and \( S_{i,j,k} \) is the difference between subsample \( k \) within duplicate \( j \) and the mean for duplicate \( j \). The last three terms in the equation (shown as upper-case letters) are random variables, all of mean zero, and each with a variance. We assume that these variances are uniform for all observations, and denote them by \( \sigma^2_Q \), \( \sigma^2_R \) and \( \sigma^2_S \). They are respectively the medium-scale spatial variance, the sampling variance and the subsampling+analytical variance that we want to estimate.

We estimated the fixed effects and the variance components for each random effect using residual maximum likelihood. If we denote the \( m \) variance components by \( \sigma^2_l, l = 1,2,\ldots,m \) then we may write a covariance matrix for the observations \( \mathbf{V} \) given by

\[
\mathbf{V} = \sum_{l=1}^{m} \sigma^2_l \mathbf{U}_l \mathbf{U}_l^T,
\]

where \( \mathbf{U}_l \) is a design matrix for the \( l \)th fixed effect. If we have \( n \) observations, and at level \( l \) of the nested structure there are \( n_k \) units then \( \mathbf{U}_l \) is a \( n \times n_k \) matrix, and if the \( g \)th observation is in the \( j \)th unit at level \( l \) then the element in the \( j \)th column of row \( g \) in \( \mathbf{U}_l \) is 1 and all the rest in that row are zero.
Let $X$ denote the design matrix for the fixed effects. This can be constructed in various ways, and since our samples are all nested in a balanced way within the fixed effects categories (i.e. all values from a particular site correspond to the same land use, parent material and region), the precise fixed effects model that we construct from these has no effect on our estimates of the variances. For consistency with Equation (??), if there are $p$ unique combinations the regions, parent materials and land uses, then $X$ is an $n \times p$ matrix. If the $g$th observation corresponds to the $j$th out of these $p$ combinations then the element in the $j$th column of row $g$ in $X$ is 1 and all the rest in that row are zero.

We can now write the log residual likelihood function corresponding to our model. This is a function of our data, conditional on the specified model, but with the fixed effects filtered out so that the variance component estimates do not depend on them. This reduces bias in the variance estimates (see Webster et al., 2006). The residual log likelihood is

$$\ell_R = -\frac{1}{2} \left( \ln V + \ln |X^T V^{-1} X| + z^T P z \right),$$

where $z$ is the vector of data values and $P$ is

$$P = V^{-1} - V^{-1} X \left( X^T V^{-1} X \right)^{-1} X^T V^{-1}.$$

We estimate the variance components numerically by finding the values which, when used to compute $V$ by means of Equation (??), maximize Equation (??).

We examined summary statistics of the data; the three metals and organic carbon data had skewness coefficients $> 1$ and so we transformed them by taking natural logarithms to stabilize their variances, as is recommended in standard texts such as Webster & Oliver (2007). In practice we might subsequently need to back-transform estimates to the original units. However, for our present purpose, to investigate the relative magnitudes of different sources of uncertainty, the results on appropriately-transformed data are most reliable. The pH data were not transformed. We then found the REML estimates of the variance components using the VCOMPONENTS and
REML directives in GenStat (Payne et al., 2008). This also returns standard errors for the variance components (root mean-square errors of the estimates). It also computes Wald tests on the fixed effects, which allows us to test the null hypotheses that, for example, the mean values do not differ between regions, landuses or parent materials.

**Results and their interpretation**

Summary statistics for the five topsoil indicators are shown in Table 2 by region and for all regions combined. The median metal concentrations (Cu, Ni, and Zn) in the topsoil of Northern Ireland are larger than those of the two English regions, which reflects the occurrence in Northern Ireland of soils derived from basaltic bedrock and superficial deposits over this lithology. The soils of Northern Ireland also have larger median SOC concentrations and lower median pH values than those of East Anglia due to a combination of factors including parent material, climate and land use.

The importance of the the three fixed effects (region, parent material and land use) on the five soil indicators is highlighted in Table 3. With the exception of land use and topsoil Zn concentrations (P-value=0.056), the test statistics for our null hypotheses are all significant (P-values < 0.05 ) for each fixed effect and soil indicator. The regional effect partly reflects the variations in the two other fixed effects (parent material and land use), although the differing climates and elevations for the three regions are also likely to exert an influence, particularly for soil organic carbon and pH.

The variance components for the three random effects are shown in Table 4. In each case the standard errors are an order of magnitude smaller than their respective variance components. The variance components for each indicator diminish by about an order of magnitude from medium-scale, to sampling, to analytical plus subsampling variance. In the case of Cu, the magnitude of the variance components are somewhat different to this general rule. The sampling variance of Cu is only three times smaller than the medium-scale variance, whilst the analytical plus subsampling variance is about 30-times smaller than the sampling variance. However, in common with the other
indicators, the difference between the medium-scale and analytical plus subsampling variance for Cu is around two orders of magnitude. Further work would be required to explain the larger sampling variances for Cu in relation to those of Ni and Zn.

Discussion

Results from the duplicate sampling protocol described in this study provide the first comprehensive estimates of local, sample error variance for five topsoil indicators at the regional scale. It should be borne in mind that the sample error variances are somewhat inflated because of the shift in location of 21 metres in one of four possible directions (see Figure 2). This could be thought of as an approximation to the location error. Further work is required to determine the relative magnitudes of locational error and sampling error which are needed for effective resampling strategies of soil inventory sites (Lark, this volume). This would require more detailed sampling on the support of the cores that are aggregated to form the site sample to characterize spatial variability over the scales of the location error.

Our analyses confirm that the sampling error variance is substantially more important than the analytical and subsampling error (Ramsey, 1998), if we resample the soil to detect change with the field procedures used in the G-BASE and Tellus surveys. If, for some reason, we needed to reduce the overall uncertainty of our observations at sample sites, (e.g. to reduce the confidence intervals for estimates of change to acceptable levels) then the most effective way to do this is not by improving our analyses or subsampling protocols, but by doing all that we can to reduce the sampling variance. This could be done by increasing the number of cores that are collected and bulked for each site.

Conclusion

Our findings show that sampling error variances for five of the eight minimum soil indicators identified for England and Wales (Environment Agency, 2006), across contrasting regions of the United Kingdom, are about an order of magnitude larger than the combined analytical and subsampling variance, and an order of magnitude smaller
than the variance within the regions, parent materials and land uses. This shows that
the largest reductions in estimation variance for means within these groups would be
obtained by increasing the number of sample sites, and the next largest reductions by
increasing the number of cores that are aggregated at each sample site.

Further work is needed on the effects of location error, and to investigate similar
sources of uncertainty in other soil indicators.

Acknowledgements

The Tellus surveys were financed in Northern Ireland by the Department of Enterprise, Trade and Investment and the 'Building Sustainable Prosperity' scheme of the Rural Development Programme. This paper is published with the permission of the Executive Director of the British Geological Survey (Natural Environment Research Council) and the Director of the Geological Survey of Northern Ireland. R.M. Lark's contributions were part of Rothamsted Research’s programme in Mathematical and Computational Biology which is supported by the Biotechnology and Biological Sciences Research Council of the United Kingdom. The authors wish to acknowledge the contributions of all staff from the Geological Survey of Northern Ireland and the British Geological Survey who were involved in the organisation, collection, analysis or processing of the G-BASE and Tellus soil surveys. We also thank Tim Farewell at Cranfield University for providing summaries of the major soil types for parts of England.

References


List of Figures and Captions

Figure 1 Soil survey regions: a) Northern Ireland, b) Humber-Trent and c) East Anglia.

Figure 2 Sampling protocol at each duplicate survey site. The support of the original survey site is shown as a solid black line, with five auger holes (filled discs) at the corner and centre of a square with side length of 20 metres. These five samples are combined to form a composite sample (duplicate A; DUP A). The sample support of four possible locations for the duplicate sample, separated by one metre from the support of DUP A, are shown as dashed lines. One of these four supports was randomly selected at each duplicate site, and five auger holes combined to form a composite sample (duplicate B; DUP B).

Figure 3 Location of the 196 duplicate sample sites across the three study regions shown in Figure 1 classified by: a) land use, and b) parent material. The frequency of each class is shown in parenthesis. Scale 1:2,000,000; grid coordinates are Latitude-Longitude.
Table 1 Major soil types across the three regions (%).

<table>
<thead>
<tr>
<th></th>
<th>East Anglia</th>
<th>Humber-Trent</th>
<th>Northern Ireland</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gleys Soils</td>
<td>29.5</td>
<td>52.7</td>
<td>54</td>
</tr>
<tr>
<td>Brown Soils</td>
<td>36.9</td>
<td>35.8</td>
<td>5.5</td>
</tr>
<tr>
<td>Lithomorphic Soils</td>
<td>6.2</td>
<td>4.4</td>
<td>0</td>
</tr>
<tr>
<td>Podzolic Soils</td>
<td>0.2</td>
<td>0</td>
<td>1.4</td>
</tr>
<tr>
<td>Peat Soils</td>
<td>2.6</td>
<td>3.1</td>
<td>14</td>
</tr>
<tr>
<td>Pelosols</td>
<td>12.2</td>
<td>2.9</td>
<td>17.8</td>
</tr>
</tbody>
</table>
Table 2 Summary statistics of five topsoil indicators including transforms to natural logarithms for all regions combined and each of the regions. Metal concentrations are mg kg\(^{-1}\) and SOC is %. 

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>St dev.</th>
<th>Skewness</th>
<th>Mean</th>
<th>St dev.</th>
<th>Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Three regions (n=776)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cu</td>
<td>27.3</td>
<td>19.4</td>
<td>25.8</td>
<td>2.7</td>
<td>3.0</td>
<td>0.8</td>
<td>0.2</td>
</tr>
<tr>
<td>Ni</td>
<td>33.4</td>
<td>22.0</td>
<td>39.3</td>
<td>2.9</td>
<td>3.1</td>
<td>0.9</td>
<td>0.2</td>
</tr>
<tr>
<td>Zn</td>
<td>89.5</td>
<td>66.2</td>
<td>180.6</td>
<td>11.5</td>
<td>4.2</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>SOC(^a)</td>
<td>1.8</td>
<td>5.5</td>
<td>0.9</td>
<td>2.6</td>
<td>1.8</td>
<td>0.9</td>
<td>0.4</td>
</tr>
<tr>
<td>pH(^a)</td>
<td>5.41</td>
<td>5.20</td>
<td>1.32</td>
<td>−0.31</td>
<td>1.77</td>
<td>0.2</td>
<td>−0.2</td>
</tr>
<tr>
<td><strong>Humber-Trent (n=260)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cu</td>
<td>19.7</td>
<td>17</td>
<td>9.9</td>
<td>0.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ni</td>
<td>23.6</td>
<td>20</td>
<td>17.5</td>
<td>3.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zn</td>
<td>91.9</td>
<td>77</td>
<td>97.5</td>
<td>5.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOC(^a)</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pH(^a)</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>East Anglia (n=216)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cu</td>
<td>15.9</td>
<td>14.5</td>
<td>8.9</td>
<td>2.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ni</td>
<td>18.8</td>
<td>17.7</td>
<td>10.5</td>
<td>0.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zn</td>
<td>56.0</td>
<td>56.6</td>
<td>25.9</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOC(^a)</td>
<td>2.9</td>
<td>2.4</td>
<td>2.0</td>
<td>2.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pH(^a)</td>
<td>6.8</td>
<td>7.1</td>
<td>0.9</td>
<td>−1.76</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Northern Ireland (n=300)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cu</td>
<td>42.2</td>
<td>34.3</td>
<td>34.9</td>
<td>1.52</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ni</td>
<td>52.4</td>
<td>31.7</td>
<td>55.4</td>
<td>1.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zn</td>
<td>111.6</td>
<td>75.5</td>
<td>273.0</td>
<td>8.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOC(^a)</td>
<td>13.4</td>
<td>7.7</td>
<td>13.6</td>
<td>2.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pH(^a)</td>
<td>4.7</td>
<td>4.8</td>
<td>0.87</td>
<td>−2.4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) sample size of 316 for two regions; nd – not determined
Table 3 Wald test results from sequential addition of three fixed effects to the model: 
a) regions, b) parent material and c) land use.

<table>
<thead>
<tr>
<th></th>
<th>Wald Statistic</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) region</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln Cu</td>
<td>55.8</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>ln Ni</td>
<td>27.0</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>ln Zn</td>
<td>14.2</td>
<td>0.001</td>
</tr>
<tr>
<td>ln SOC</td>
<td>256.3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>pH</td>
<td>353.4</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>b) parent material</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln Cu</td>
<td>26.2</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>ln Ni</td>
<td>20.5</td>
<td>0.003</td>
</tr>
<tr>
<td>ln Zn</td>
<td>13.5</td>
<td>0.04</td>
</tr>
<tr>
<td>ln SOC</td>
<td>100.7</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>pH</td>
<td>33.3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>c) land use</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln Cu</td>
<td>54.1</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>ln Ni</td>
<td>53.5</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>ln Zn</td>
<td>17.1</td>
<td>0.056</td>
</tr>
<tr>
<td>ln SOC</td>
<td>63.9</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>pH</td>
<td>95.9</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
Table 4 Variance components and standard errors (in parenthesis) for the three random effects for each of the five soil indicators.

<table>
<thead>
<tr>
<th></th>
<th>Medium-scale, spatial</th>
<th>Sampling</th>
<th>Analytical + Subsampling</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln Cu</td>
<td>0.317 (0.04)</td>
<td>0.099 (0.010)</td>
<td>0.003 ($22.2 \times 10^{-5}$)</td>
</tr>
<tr>
<td>ln Ni</td>
<td>0.531 (0.057)</td>
<td>0.017 (0.002)</td>
<td>0.003 ($20.3 \times 10^{-5}$)</td>
</tr>
<tr>
<td>ln Zn</td>
<td>0.348 (0.039)</td>
<td>0.029 (0.003)</td>
<td>0.001 ($80.0 \times 10^{-6}$)</td>
</tr>
<tr>
<td>ln SOC</td>
<td>0.181 (0.027)</td>
<td>0.024 (0.003)</td>
<td>0.002 ($21.5 \times 10^{-5}$)</td>
</tr>
<tr>
<td>pH</td>
<td>0.326 (0.048)</td>
<td>0.048 (0.007)</td>
<td>0.004 ($34.0 \times 10^{-5}$)</td>
</tr>
</tbody>
</table>
Figure 2
Figure 3