

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

B. G. RAWLINS^{a,*}, A. J. SCHEIB^a, R. M. LARK^b & T. R. LISTER^a.

^a*British Geological Survey, Keyworth, Nottingham NG12 5GG, UK*

^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*

1 **Summary**

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

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

27 **Introduction**

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

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

50 The variance of these notional repeat sample averages is the sample error variance.
51 In addition to this the material that we sample is thoroughly mixed, but subsampling
52 for analysis introduces uncertainty too (different subsamples would have different val-
53 ues). We call this the subsample error, which introduces subsample error variance. We
54 know that multiple analyses of uniform material will not return exact results and the

55 variance of such multiple determinations is the analytical variance. Although separa-
56 ble in theory, the subsample and analytical variance cannot be separated in practice.
57 Analytical variance must be estimated by repeat analyses of uniform material, but soil
58 is never entirely uniform and so the resampling and analytical variances will always be
59 confounded.

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

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

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

81 Data are available on topsoil properties (including five soil indicators) at 196 sites
82 across three large regions of the UK, comprising a diverse range of soil, mean annual

83 rainfall, parent material and land use types. In such grid-based surveys, it is inevitable
84 that location and sampling errors will be conflated.

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

92 **Methods**

93 *Study region and surveys*

94 The three regions (Figure 1) comprise a diverse range of soil parent materials and soil
95 types; further details of the soil types in Northern Ireland are provided by Cruickshank
96 (1997), whilst the soils in the study regions of England are described in three regional
97 bulletins (Soil Survey of England and Wales, 1983a-c). The major soil types in each
98 of the three regions are summarized in Table 1. The larger proportion of peat soils in
99 Northern Ireland compared to the two regions in England reflects the wetter environ-
100 ment of the former, where average mean annual rainfall for the vast majority of the
101 region is greater than 1 m, with a minimum of around 0.75 m.

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

109 At each sampling site, including those selected for duplicate sampling, five in-
110 cremental soil samples were collected using a Dutch auger at the corners and centre

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

122 All samples of soil were air-dried in a dedicated temperature controlled oven at
123 30 °C for 2–3 days, disaggregated and sieved to <2 mm. The samples were coned and
124 quartered and each split into two sub-samples. The two samples from the DUP A
125 site were labelled DUP A and SUB A, whilst the samples from the DUP B site were
126 labelled DUP B and SUB B. From each of the four sub-samples, a 50-g sub-sample
127 was ground in an agate planetary ball mill. The total concentrations of Ni, Cu and
128 Zn were determined in each sample by wavelength and energy dispersive XRFs (X-
129 Ray Fluorescence Spectrometry). The samples were prepared for analysis by grinding
130 12 g of sample material and 3 g of Elvacite 2013 (n-butyl methacrylate copolymer,
131 Dupont & Co.) in an agate planetary ball mill for half an hour. The mixture was then
132 pressed using a 25 tonne load into pellets with a thickness of 40 mm. The coefficients
133 of variation for a soil sample standard were: Ni (0.6 %) Cu (1%) and Zn (0.8 %).

134 Soil pH and SOC were only determined on a subset of 114 sites (a total of 456 sam-
135 ples) from two of the regions; East Anglia and Northern Ireland. Soil organic carbon
136 was estimated in each sample using loss-on-ignition analysis by heating a sub-sample to
137 450 °C for eight hours and multiplying the mass difference by 0.58 (Broadbent, 1953).
138 The coefficient of variation for this method for 174 replicate analyses of a sample stan-

139 dard was 3.6 %. The pH of each sample was measured using a glass electrode and
140 Orion 720A meter. To 10 g of the <2 mm sample, 25 ml of 0.01 M calcium chloride
141 solution was added, the samples were magnetically stirred for one minute and then left
142 to settle for 15 minutes. Prior to recording the pH value, the suspension was stirred.

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

151 *Data analysis*

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

161 There are two kinds of sources of variation in our data. Random effects arise
162 because they may be regarded as random processes due to our sampling. Coning
163 and quartering is regarded as a way of randomly subsampling material, the random
164 selection of one of four possible second squares to sample at each site is a random source
165 of variation (sampling error variation), and the random selection of sites for this more
166 detailed investigation introduces what we shall call medium-scale spatial variation —

167 it is variation within the land uses, parent materials and regions. The second source
 168 of variation arises from fixed effects which we chose to sample, including land uses,
 169 parent materials and regions. They do not enter into our sampling in a random way,
 170 but because we chose to sample them.

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

$$Z_{a,b,c,i,j,k} = \mu_{a,b,c} + Q_i + R_{i,j} + S_{i,j,k}, \quad (1)$$

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

184 We estimated the fixed effects and the variance components for each random
 185 effect using residual maximum likelihood. If we denote the m variance components by
 186 σ_l^2 , $l = 1, 2, \dots, m$ then we may write a covariance matrix for the observations \mathbf{V} given
 187 by

$$\mathbf{V} = \sum_{l=1}^m \sigma_l^2 \mathbf{U}_l \mathbf{U}_l^T, \quad (2)$$

188 where \mathbf{U}_l is a design matrix for the l th fixed effect. If we have n observations, and at
 189 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
 190 g th observation is in the j th unit at level l then the element in the j th column of row
 191 g in \mathbf{U}_l is 1 and all the rest in that row are zero.

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

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

$$\ell_R = -\frac{1}{2} \left(\ln \mathbf{V} + \ln \left| \mathbf{X}^T \mathbf{V}^{-1} \mathbf{X} \right| + \mathbf{z}^T \mathbf{P} \mathbf{z} \right), \quad (3)$$

206 where \mathbf{z} is the vector of data values and \mathbf{P} is

$$\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1} \mathbf{X} \left(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbf{V}^{-1}. \quad (4)$$

207 We estimate the variance components numerically by finding the values which, when
 208 used to compute \mathbf{V} by means of Equation (??), maximize Equation (??).

209 We examined summary statistics of the data; the three metals and organic carbon
 210 data had skewness coefficients > 1 and so we transformed them by taking natural
 211 logarithms to stabilize their variances, as is recommended in standard texts such as
 212 Webster & Oliver (2007). In practice we might subsequently need to back-transform
 213 estimates to the original units. However, for our present purpose, to investigate the
 214 relative magnitudes of different sources of uncertainty, the results on appropriately-
 215 transformed data are most reliable. The pH data were not transformed. We then
 216 found the REML estimates of the variance components using the VCOMPONENTS and

217 REML directives in GenStat (Payne *et al.*, 2008). This also returns standard errors for
218 the variance components (root mean-square errors of the estimates). It also computes
219 Wald tests on the fixed effects, which allows us to test the null hypotheses that, for
220 example, the mean values do not differ between regions, landuses or parent materials.

221 **Results and their interpretation**

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

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

237 The variance components for the three random effects are shown in Table 4. In
238 each case the standard errors are an order of magnitude smaller than their respective
239 variance components. The variance components for each indicator diminish by about
240 an order of magnitude from medium-scale, to sampling, to analytical plus subsampling
241 variance. In the case of Cu, the magnitude of the variance components are somewhat
242 different to this general rule. The sampling variance of Cu is only three times smaller
243 than the medium-scale variance, whilst the analytical plus subsampling variance is
244 about 30-times smaller than the sampling variance. However, in common with the other

245 indicators, the difference between the medium-scale and analytical plus subsampling
246 variance for Cu is around two orders of magnitude. Further work would be required to
247 explain the larger sampling variances for Cu in relation to those of Ni and Zn.

248 **Discussion**

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

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

268 **Conclusion**

269 Our findings show that sampling error variances for five of the eight minimum
270 soil indicators identified for England and Wales (Environment Agency, 2006), across
271 contrasting regions of the United Kingdom, are about an order of magnitude larger than
272 the combined analytical and subsampling variance, and an order of magnitude smaller

273 than the variance within the regions, parent materials and land uses. This shows that
274 the largest reductions in estimation variance for means within these groups would be
275 obtained by increasing the number of sample sites, and the next largest reductions by
276 increasing the number of cores that are aggregated at each sample site.

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

279 **Acknowledgements**

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

293 **References**

- 294 Broadbent, F.E. 1953. The soil organic fraction. *Advances in Agronomy*, **5**, 153–183.
- 295 British Geological Survey, 2006. Digital Geological Map of Great Britain 1:50 000 scale
296 (DiGMapGB-50) data [CD-ROMom] Version 3.14. British Geological Survey,
297 Keyworth, Nottingham.

- 298 Cruickshank, J.G. 1997. *Soil and Environment: Northern Ireland*. Queen's University
299 of Belfast, Belfast.
- 300 Defra, 2003. *Sampling strategies and soil monitoring: Report SP0514*. Defra, London.
301 <http://randd.defra.gov.uk/Document.aspx?Document=SP0514847FRP.doc>. Ac-
302 cessed 21 May 2008.
- 303 Environment Agency, 2006. *The development and use of soil quality indicators for as-*
304 *sessing the role of soil in environmental interactions: Science Report SCO30265*.
305 Environment Agency, Bristol.
- 306 Lark, R. M. this volume. Estimating the mean status and change of soil properties
307 for regions: two distinct objectives for soil inventory. *European Journal of Soil*
308 *Science*.
- 309 Payne, R.W. (ed) 2008. *GenStat Release 11 Reference Manual. Part 2 Directives*.
310 VSN International, Hemel Hempstead.
- 311 Rawlins, B.G., Lister, T.R. & Mackenzie, A. 2002. Trace metal pollution of soils in
312 northern England. *Environmental Geology*, **7**, 612–620.
- 313 Ramsey, M. H. 1998. Sampling as a source of measurement uncertainty: techniques
314 for quantification and comparison with analytical sources. *Journal of Atomic*
315 *Spectrometry*, **13**, 97-104.
- 316 Soil Survey of England and Wales, 1983a. *Soils and their use in Midland and Western*
317 *England*. Ordnance Survey for the Soil Survey of England and Wales, Southamp-
318 ton.
- 319 Soil Survey of England and Wales, 1983b. *Soils and their use in Eastern England*.
320 Ordnance Survey for the Soil Survey of England and Wales, Southampton.
- 321 Soil Survey of England and Wales, 1983c. *Soils and their use in Northern England*.
322 Ordnance Survey for the Soil Survey of England and Wales, Southampton.

323 Webster, R. & Oliver, M.A. 2007. *Geostatistics for Environmental Scientists*, 2nd
324 Edition. John Wiley & Sons, Chichester.

325 Webster, R., Welham, S.J., Potts, J.M. & Oliver, M.A. 2006. Estimating the spa-
326 tial scales of regionalized variables by nested sampling, hierarchical analysis of
327 variance and residual maximum likelihood. *Computers and Geosciences*, **32**,
328 1320–1333.

329 **List of Figures and Captions**

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

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

340 **Figure 3** Location of the 196 duplicate sample sites across the three study regions
341 shown in Figure 1 classified by: a) land use, and b) parent material. The fre-
342 quency of each class is shown in parenthesis. Scale 1:2,000,000; grid coordinates
343 are Latitude-Longitude.

344 **Table 1** Major soil types across the three regions (%).

	East Anglia	Humber-Trent	Northern Ireland
Gleys Soils	29.5	52.7	54
Brown Soils	36.9	35.8	5.5
Lithomorphic Soils	6.2	4.4	0
Podzolic Soils	0.2	0	1.4
Peat Soils	2.6	3.1	14
Pelosols	12.2	2.9	17.8

345

346 **Table 2** Summary statistics of five topsoil indicators including transforms to natural
 347 logarithms for all regions combined and each of the regions. Metal concentrations are
 348 mg kg⁻¹ and SOC is % .

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

350 ^a sample size of 516 for two regions; nd – not determined

351 **Table 3** Wald test results from sequential addition of three fixed effects to the model:

352 a) regions, b) parent material and c) land use.

	Wald Statistic	<i>P</i> -value
a) region		
ln Cu	55.8	<0.001
ln Ni	27.0	<0.001
ln Zn	14.2	0.001
ln SOC	256.3	<0.001
pH	353.4	<0.001
b) parent material		
ln Cu	26.2	<0.001
ln Ni	20.5	0.003
ln Zn	13.5	0.04
ln SOC	100.7	<0.001
pH	33.3	<0.001
c) land use		
ln Cu	54.1	<0.001
ln Ni	53.5	<0.001
ln Zn	17.1	0.056
ln SOC	63.9	<0.001
pH	95.9	<0.001

353

354 **Table 4** Variance components and standard errors (in parenthesis) for the three ran-
 355 dom effects for each of the five soil indicators.

	Medium-scale, spatial	Sampling	Analytical + Subsampling
ln Cu	0.317 (0.04)	0.099 (0.010)	0.003 (22.2×10^{-5})
ln Ni	0.531 (0.057)	0.017 (0.002)	0.003 (20.3×10^{-5})
ln Zn	0.348 (0.039)	0.029 (0.003)	0.001 (80.0×10^{-6})
ln SOC	0.181 (0.027)	0.024 (0.003)	0.002 (21.5×10^{-5})
pH	0.326 (0.048)	0.048 (0.007)	0.004 (34.0×10^{-5})

Figure 1:

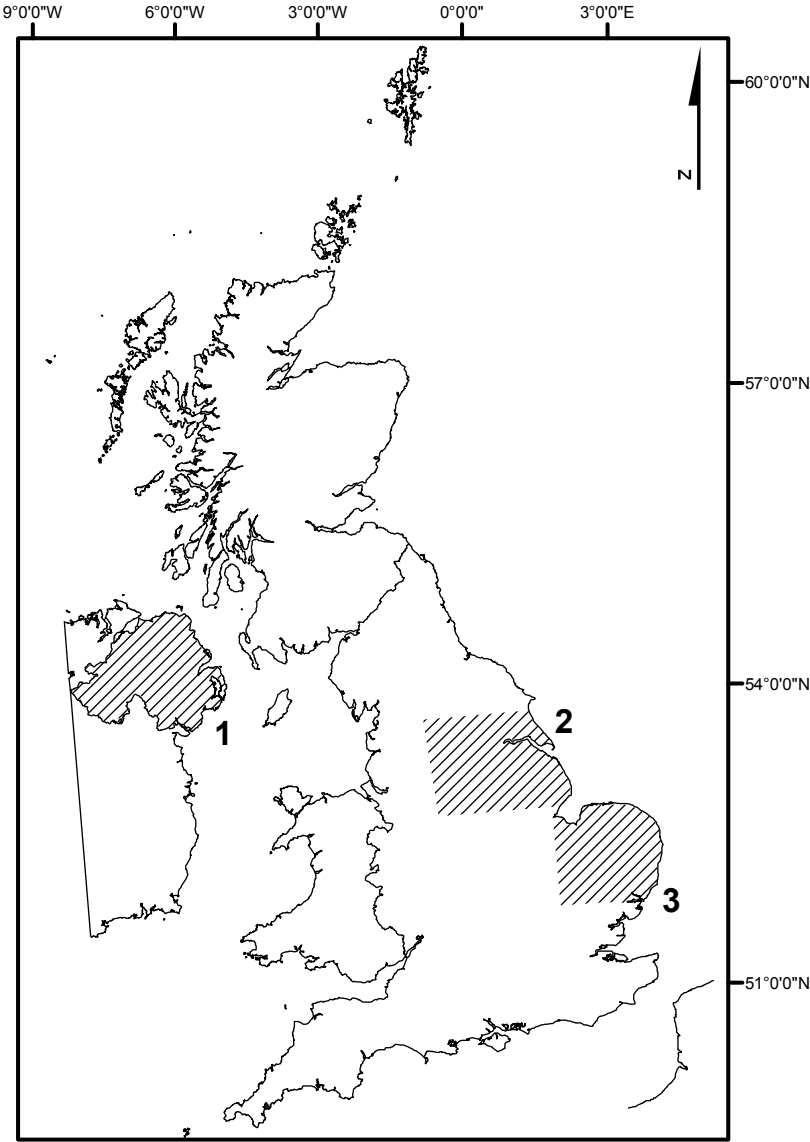
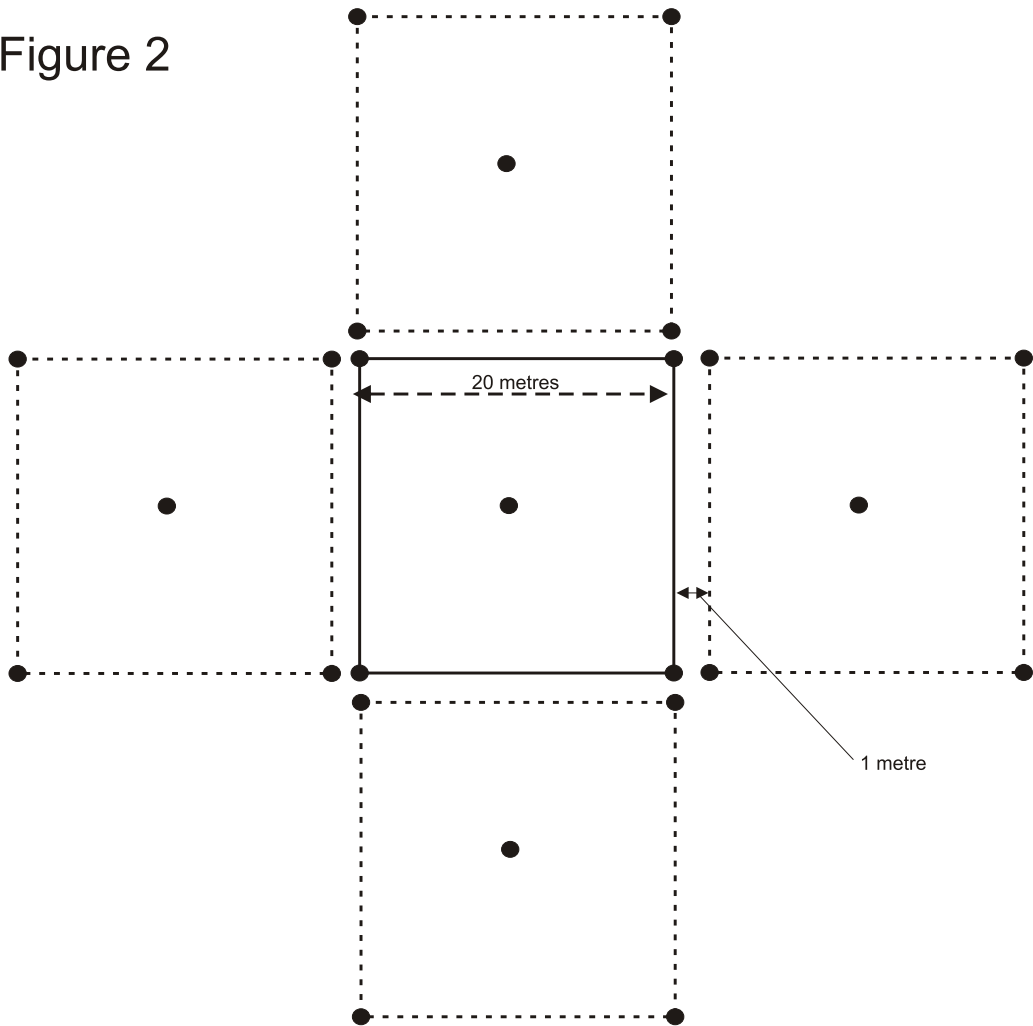


Figure 2



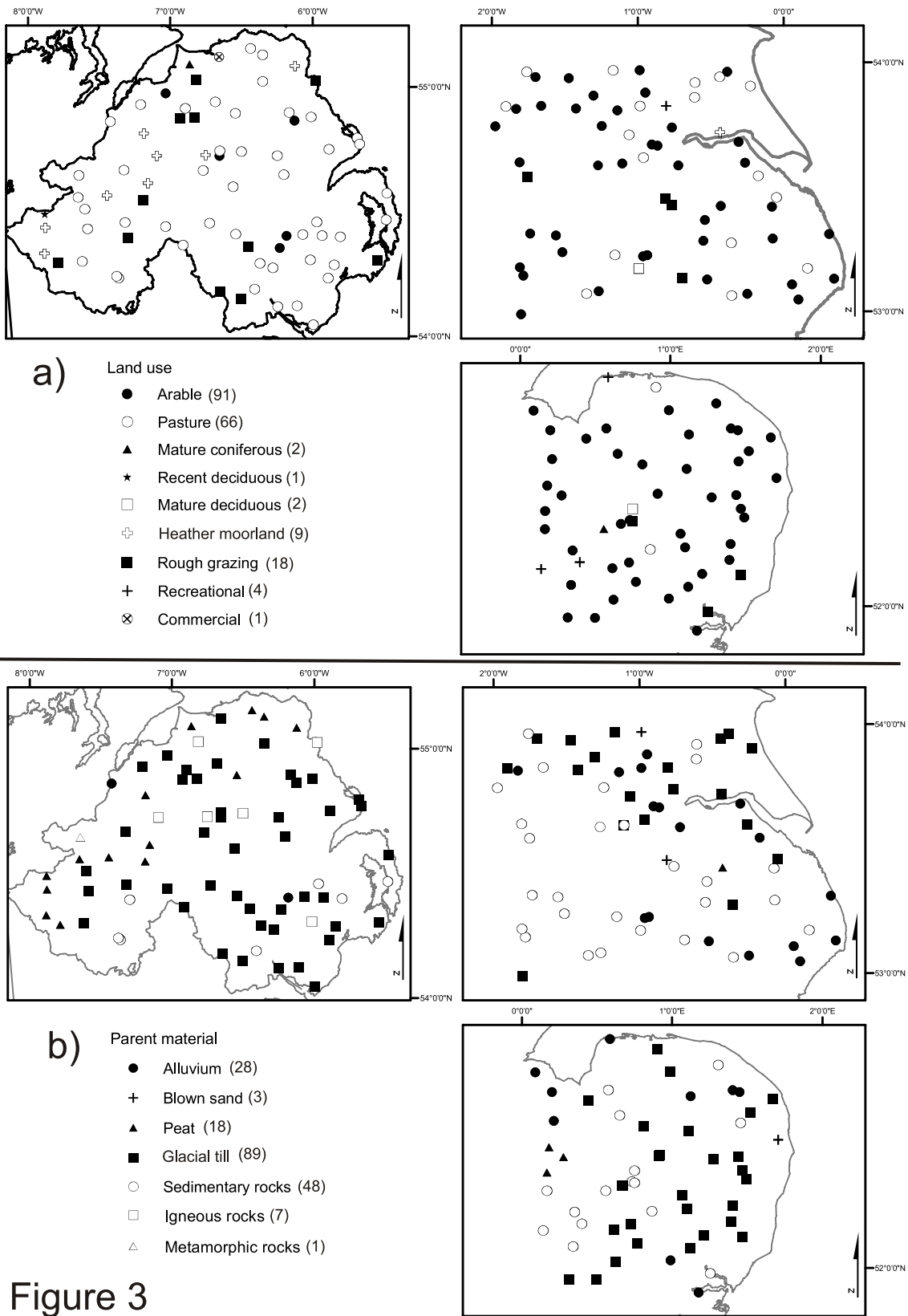


Figure 3