Earth Surface Processes and Landforms
Earth Surf. Process. Landforms **28**, 1389–1409 (2003)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/esp.507

THE INFLUENCE OF PARENT MATERIAL ON TOPSOIL GEOCHEMISTRY IN EASTERN ENGLAND

B. G. RAWLINS,[1]* R. WEBSTER[2] AND T. R. LISTER[1]

[1] *British Geological Survey, Keyworth, Nottingham, NG12 5GG, UK*
[2] *Rothamsted Experimental Station, Harpenden, Hertfordshire, AL5 2JQ, UK*

Received 5 April 2002; Revised 4 February 2003; Accepted 7 February 2003

ABSTRACT

The topsoil of around 10 000 km^2 in eastern England has recently been sampled intensely at 4609 sites to characterize its geochemistry. The parent materials, which include both solid geology and Quaternary sediments, range in age from Permian to Holocene. The distributions of the concentrations of major and trace elements have been characterized geostatistically, and the role of parent material on their spatial structure (anisotropy) and their spatial relationships (coregionalization) have been investigated.

Analysis of variance with the sites grouped by major parent material type showed that this classification accounted for 14 to 48 per cent of the variance for the various elements. Global variograms of 13 elements (Al, As, Ca, Cr, Cu, Fe, Mg, Mo, Ni, P, Pb, Ti, and U) have been computed and modelled. Eleven of the variograms seem to comprise two structures, both of which we modelled with spherical functions, one of short range, 3·5 to 9 km, and the other with a range of 15 to 23 km. The models included a nugget variance, which varied from 27 per cent (for As, Fe, and Mg) to 63 per cent (for P) of the total. The long-range structures are related to the separations of the major parent materials.

The variograms of several elements showed appreciable anisotropy, most notably that of Mg. Anisotropy is evident at short ranges of less than 5 km. This accords with the geological structure of the beds which dip from west to east so that their outcrops are elongated from north to south. A linear model of coregionalization fitted to the data emphasized several important geochemical associations, which we interpret. Elements commonly associated with clay minerals (Mg, Al) and the clay size fraction (Ti) are dominated by the long-range structure of the coregionalization, whilst several trace elements (As, Cr, Ni and U) are spatially correlated with Fe over short distances, through adsorption of the former on the surfaces of Fe oxyhydroxides.

The topsoil around large urban areas is enriched in lead, but it is not clear whether anthropogenic sources are responsible for this metal's anomalous spatial relationships with other elements.

KEY WORDS: spatial distributions; geostatistics; geochemistry; soil; trace elements

INTRODUCTION

Accurate estimates of the natural, background concentration of major and trace elements in topsoil are important for several applications, including the assessment of soil fertility, mineral exploration, assessing contamination, the application of sewage sludge to agricultural land, and estimating diffuse pollution through atmospheric deposition. There are two main groups of factors that determine the natural geochemistry of soil: the parent material and its geochemical composition, and the weathering (pedogenic) processes that have transformed that material into the soil itself. These are well known qualitatively, but we understand rather little of their relative importance in determining the spatial distribution of major and trace elements in the soil.

Geostatistics is now much used for investigating spatial distributions in mining and mineral prospection. The techniques have been used successfully to investigate the spatial variation of metals in soil and its relations with regional geological structure (such as the distribution of parent material types). McBratney *et al*. (1982) and Goovaerts and Webster (1994), for example, demonstrated the significance of parent material on the contents of Cu and Co in the soil in southeast Scotland where a classification by parent material accounted for a large proportion of the variance of readily extractable cobalt and copper in their sample. By modelling experimental

* Correspondence to: B. G. Rawlins, British Geological Survey, Keyworth, Nottingham, NG12 5GG, UK. E-mail: bgr@bgs.ac.uk

variograms we can estimate the correlation ranges of different variables (their limit of spatial dependence), which may be related to the structure of the parent material. Two-dimensional variograms can show whether the spatial variation is anisotropic (different amounts, or rates, of variation in different directions) and whether this is related to the orientation of parent material. In addition, the spatial correlation of several variables can be investigated by studying their coregionalization, as demonstrated by Webster *et al*. (1994) in the soil in the Swiss Jura.

Soils throughout England and Wales have developed from many kinds of bedrock and overlying Quaternary deposits, or a combination of the two where the latter are thin. Our knowledge of the spatial variation of soil geochemistry at the national scale is derived largely from the data from the National Soil Inventory displayed in the *Soil Geochemical Atlas of England and Wales* of McGrath and Loveland (1992). Soil samples for the Inventory were collected on a regular square grid at a resolution of 1 per 25 km^2 and analysed for 17 elements (total and extractable contents). The data are being analysed by M. A. Oliver and colleagues. The results published so far (Oliver *et al*., 1996; Scholz *et al*., 1999; Webster *et al*., 2000) show variograms of the elements to be bounded, with nugget variances (that part of the variation occurring at distances shorter than the sampling interval or resulting from analytical error) of around 60 per cent and longer-range structures of 70–90 km, attributable to the geology. The influence of geological structure is clear in kriged maps made from these data. Much less clear is the relative importance of the factors that determine the spatial relations of the elements over shorter distances, i.e. at the local scale. Also, until very recently, we have not had the data at a fine enough resolution to investigate the nature and causes of spatial variation at these scales.

The lack of geochemical data has recently been made good by the British Geological Survey with its Geochemical Baseline Survey of the Environment project (G-BASE). This has recently embodied a regional soil survey of 10 000 km^2 in eastern England in which 4609 sites were sampled. In a preliminary analysis we made maps by kriging from the data and compared those maps with ones of parent material. The two had remarkably similar patterns, suggesting a strong influence of parent material on the geochemistry of the soil throughout the region. We have since analysed the data (for which the sampling points have been classified by parent material) more purposefully to investigate:

(1) the relative importance of parent material in determining the background concentrations of major and trace elements;
(2) the spatial relations between geochemical variables and the structure of the parent material; and
(3) the spatial relations of variables by modelling their coregionalization.

STUDY REGION AND METHODS

Figure 1 shows the location of the study area that extends over 10 260 km^2. Figure 2 is a map of the parent materials, derived from a 1 : 250 000 map of the solid geology (British Geological Survey, 1983), overlain by a 1 : 625 000 map of Quaternary deposits (Institute of Geological Sciences, 1977). The parent material comprises Quaternary deposits covering 6686 km^2 (which is 65 per cent of the region), whilst bedrock geology crops out over the remaining 3574 km^2 (35 per cent).

Bedrock in the region ranges in age from Permian to Cretaceous and comprises a variety of substrates including Magnesian Limestone (Cadeby Formation), Sherwood Sandstone, Mercia Mudstone, Jurassic Limestone, Cretaceous marls and ironstones, and Cretaceous Chalk. The outcrops of these formations are aligned predominantly from north to south as a result of the inherited basement structure and active subsidence in the North Sea. The major Quaternary deposits, described by Straw and Clayton (1979), include two types of glacial till along the northeast coast (the Skipsea, and overlying Withernsea till) deposited during the Devensian glaciation between 70 000 and 13 000 years BP. Eustatic rise in sea level associated with the post-glacial Flandrian stage resulted in the deposition of marine alluvium (clays, silts, sands and gravels) along the coast. Lacustrine deposits cover much of the Vale of York (where there is also blown sand), and alluvium has been deposited along the River Trent (to the south of the Humber).

Sample sites for the soil were chosen from every second kilometre square of the British National Grid by simple random selection within each square, subject to the avoidance of roads, tracks, railways, domestic and public and gardens, and other seriously disturbed ground. The 4609 samples (one sampling point per 2.2 km^2)

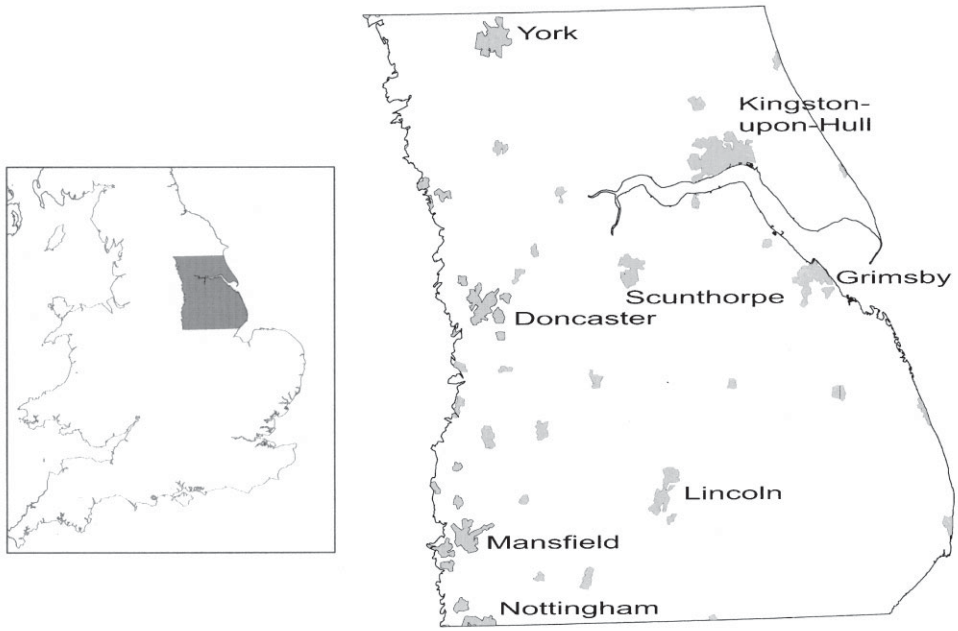

Figure 1. Location map of the study region and its urban areas in relation to part of the United Kingdom

were collected in the summers of 1994, 1995 and 1996 in rural and peri-urban areas throughout the region. At each site topsoil (0–15 cm) was taken from five holes augered at the corners and centre of a square with a side of length 20 m with a hand auger and combined to form a bulked sample. Information on the location, catchment geology, contamination, land use and other features required for interpretation was recorded in the field and later transferred to an electronic database.

All samples of soil were dried, disaggregated, sieved to pass 2 mm, coned and quartered. From each, a 50 g sub-sample was ground in an agate planetary ball mill. The total concentrations of major and trace elements were determined in each sample by wavelength dispersive XRFS (X-ray fluorescence spectrometry). The elements included Al, As, Ca, Cr, Cu, Fe, Mg, Mo, Ni, P, Pb, Ti, and U, and these are the ones that concern us here. Sub-samples were taken from 84 of the samples and analysed to provide an estimate of the analytical error for each element from an analysis of variance.

STATISTICAL METHODS

There is now a well-established and well-tried suite of statistical techniques for analysing the spatial distributions of continuous variables over the land surface. They include classical analysis of variance and more recent geostatistical methods, all of which are described in detail in, for example, Webster and Oliver (2001). Here we describe very briefly those used.

Statistical summaries and transformations

The data for all elements were summarized in terms of their means, variances, standard deviations and skewness coefficients (Table I). Histograms and box-plots identified outliers and were used to help us decide what, if any, transformations were needed to obtain approximately normal distributions. Strongly and moderately skewed data were then transformed.

Analysis of variance

The total variance for each variable was apportioned to that between classes, either of lithostratigraphy or parent material, and that within the classes by a one-way analysis. Table II is an example.

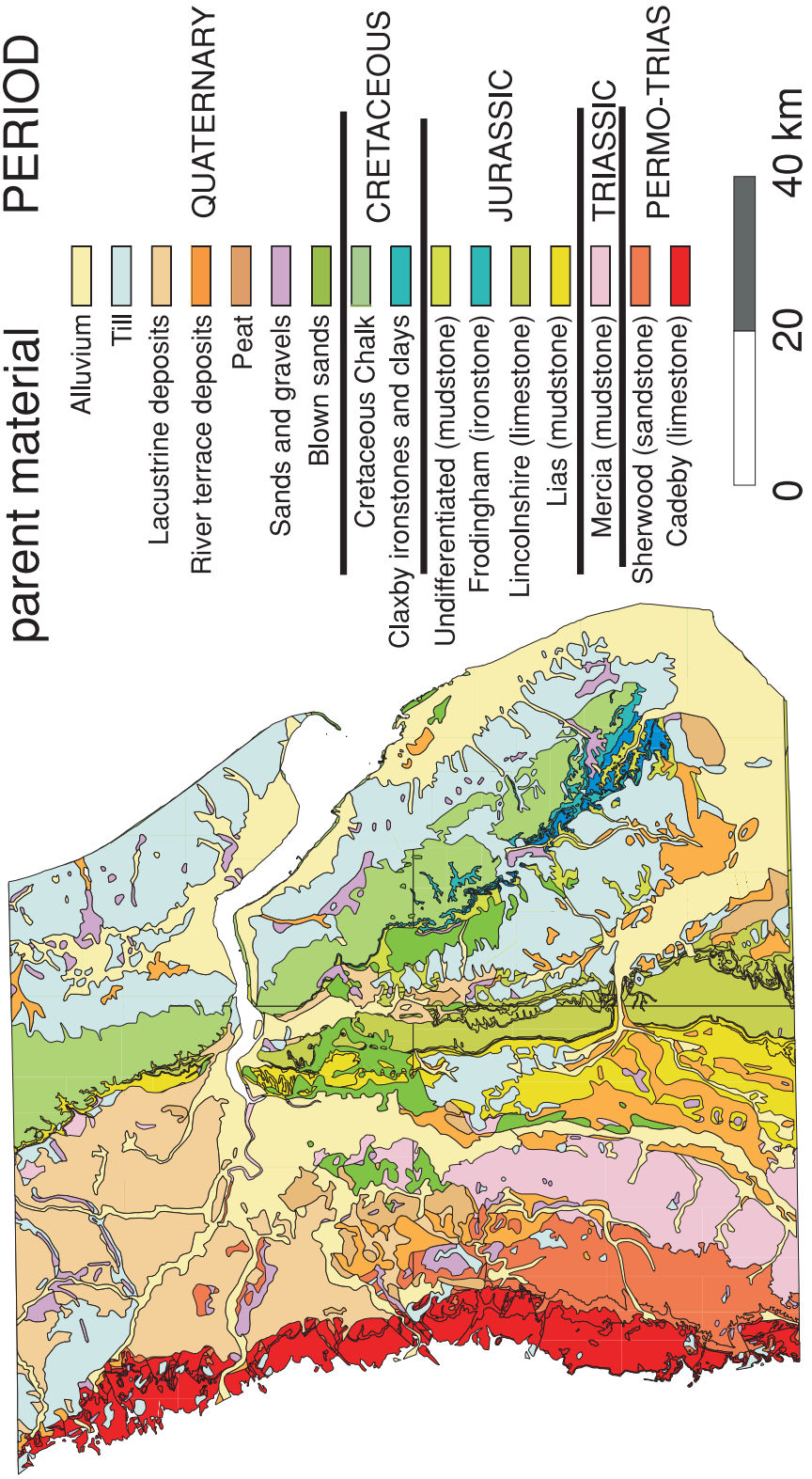

Figure 2. Parent materials of the study region

Table I. Statistical summary of major and trace element concentrations in topsoil samples ($n = 4609$). Data in mg kg^{-1} (unless otherwise stated)

	Al_2O_3 (%)	As	CaO (%)	Cr	Cu	Fe_2O_3t* (%)	MgO (%)	Mo	Ni	P_2O_5 (%)	Pb	TiO_2 (%)	U
Analytical variance	−0·38	212·800	48·100	27·30	−0.430	33·000	−0·010	−0·640	138·4	−0·590	<0·001	<0·001	10·013
Detection limit	0·1	1	0·05	1	1	0·01	0·1	1	1	0·05	1	0·005	1
Minimum	0·1	2	0·05	1	1	0·32	0·1	1	1	0·025	2	0·02	1
Maximum	22·2	342	48·1	254	655	33·0	13·4	14·2	97	2·22	1402	1·38	11·6
Mean	10·2	15·0	2·59	68·5	17·9	4·51	1·25	2·13	22·5	0·28	47·0	0·60	2·03
Median	10·6	12	1·01	67	16	4·31	0·9	2	22	0·25	36	0·62	2
Variance	20·7	212·8	19·8	745	238	5·41	1·17	0·97	138·4	0·02	2480	0·05	0·68
Standard deviation	4·55	14·6	4·44	27·3	15·42	2·33	1·08	0·99	11·77	0·13	49·8	0·22	0·83
Skewness	−0·38	9·56	4·05	0·67	18·1	2·72	3·30	2·04	0·75	2·57	12·05	−0·05	1·73
Analytical variance	0·2			29					1·0			<0·001	
Transformed data		$\log_{10}$	$\log_{10}$		$\log_{10}$	sqrt†	$\log_{10}$	$\log_{10}$		$\log_{10}$	$\log_{10}$		sqrt
Mean		1·11	0·10		1.18	2·06	−0·01	0·30		−0·59	1·60		1·40
Variance		0·05	0·23		0.06	0·26	0·09	0·04		0·03	0·04		0·08
Standard deviation		0·22	0·48		0.25	0·51	0·30	0·20		0·17	0·21		0·28
Skewness		0·91	0·56		−0.43	0·58	0·17	−0·64		0·30	1·61		0·28
Analytical variance		0·001	0·003		0.001	0·001	0·002	0·003		0·001	<0·001		0·013

* t following Fe_2O_3 means total.
† sqrt, square root.

Table II. Variance (%) accounted for due to parent material (Figure 4a) and lithostratigraphic (Figure 4b) classification

	Al_2O_3	As	CaO	Cr	Cu	Fe_2O_3t*	MgO	Mo	Ni	P_2O_5	Pb	TiO_2	U
Lithostratigraphic	9·1	28·7	21·2	13	11	19·8	28·7	13·6	18·3	6·6	18·5	13·5	7·1
Parent material	23·6	36·4	31·6	27·7	22·0	33·6	47·6	28·2	31·5	14	17	31·7	22·9

* t following Fe_2O_3 means total.

Spatial effects appear only insofar as the classes themselves occupy distinct portions of the region. As a byproduct of the analysis, we obtain the residuals from the class means.

Variograms

The variogram is now the most widely used function for describing quantitatively the way in which variance changes with separation in space. It is needed for geostatistical estimation. For an intrinsically stationary regionalized variable, $Z(\mathbf{x})$, it is

$$\gamma(\mathbf{h}) = \frac{1}{2}\mathrm{E}\left[\left\{Z(\mathbf{x}) - Z(\mathbf{x}+\mathbf{h})\right\}^2\right] \qquad \text{for all } \mathbf{h} \tag{1}$$

In this formalism $\mathbf{x}$ and $\mathbf{x}+\mathbf{h}$ are two places separated by the lag vector $\mathbf{h}$ in two dimensions, and E denotes the expectation. We estimated the function from the measured concentrations, $z(\mathbf{x}_1)$, $z(\mathbf{x}_2)$, . . . , $z(\mathbf{x}_N)$, for each element in turn using the usual computing formula:

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2M(\mathbf{h})}\sum_{i=1}^{M(\mathbf{h})}\left\{z(\mathbf{x}_i) - z(\mathbf{x}_i+\mathbf{h})\right\}^2 \tag{2}$$

where $M(\mathbf{h})$ is the number of paired comparisons at that lag.

Where variation appears to be isotropic the vector $\mathbf{h}$ can be treated as a scalar $h = |\mathbf{h}|$ in distance only, and Equation 2 is computed without regard to direction. For anisotropically varying constituents we computed the variogram in narrow bands in several directions.

The residuals from the class means can also be treated as spatially distributed random variables with their variograms, which can be estimated in the same way from the residuals deriving from the analysis of variance. We call these the pooled within-class variograms; they are pooled over all classes.

A sample variogram is an ordered set of values, and when plotted as $\hat{\gamma}(h)$ against h appears more or less erratic. The underlying variogram is continuous and smooth, and a plausible function is fitted to it, both for description and for estimation by kriging. We fitted several of the common conditional negative semi-definite functions by weighted least squares approximation using the FITNONLINEAR directive in GenStat (Payne, 2002).

Kriging

We mapped the regional distributions of the major and trace elements by first estimating their concentrations at the nodes of a fine grid by ordinary punctual kriging from the data and the model variograms. For each grid node we took a maximum of 20 points in its neighbourhood. For variables that we had earlier transformed we back-transformed the estimates to their original scales.

Coregionalization

Two or more variables may vary in space more or less coherently in the sense that they are spatially correlated with one another. Each pair of variables then has a cross-variogram in addition to the two auto-variograms. The cross-variogram for two regionalized variables, $Z_u(\mathbf{x})$ and $Z_v(\mathbf{x})$, is

$$\gamma_{uv}(\mathbf{h}) = \frac{1}{2}\mathrm{E}\left[\left\{Z_u(\mathbf{x}) - Z_u(\mathbf{x}+\mathbf{h})\right\}\left\{Z_v(\mathbf{x}) - Z_v(\mathbf{x}+\mathbf{h})\right\}\right] \qquad \text{for all } \mathbf{h} \tag{3}$$

and it can be estimated in a way analogous to the estimation of the auto-variogram by

$$\hat{\gamma}_{uv}(\mathbf{h}) = \frac{1}{2M(\mathbf{h})} \sum_{i=1}^{M(\mathrm{h})} \left\{ z_u(\mathbf{x}_i) - z_u(\mathbf{x}_i + \mathbf{h}) \right\} \left\{ z_v(\mathbf{x}_i) - z_v(\mathbf{x}_i + \mathbf{h}) \right\} \tag{4}$$

For V variables there are $V(V+1)/2$ variograms, V auto-variograms when $u = v$ and $V(V-1)/2$ cross-variograms. Together they describe the coregionalization of all the variables of interest, and they may be modelled together in a linear model. For this each variable $Z_u(\mathbf{x})$ is assumed to be a linear sum of orthogonal random variables $Y_j^k(\mathbf{x})$, each with mean 0 and variance 1, thus:

$$Z_u(\mathbf{x}) = \sum_{k=1}^{K} \sum_{j=1}^{V} a_{uj}^k Y_j^k(\mathbf{x}) + \mu_u \tag{5}$$

where μ_u is the mean of $Z_u(\mathbf{x})$ and the superscript k is simply an index, not a power. The variogram for any pair of variables u and v is then

$$\gamma(\mathbf{h}) = \sum_{k=1}^{K} \sum_{j=1}^{V} q_{uj}^k q_{vj}^k g^k(\mathbf{h}) \tag{6}$$

in which the $g^k(\mathbf{h})$ are basic variogram functions. If the products in the second summation are replaced by b_{uv}^k then Equation 6 becomes

$$\gamma(\mathbf{h}) = \sum_{k=1}^{K} b_{uv}^k g^k(\mathbf{h}) \tag{7}$$

The b_{uv}^k are the variances and covariances, i.e. nugget and sill variances for the independent components of the linear model if they are bounded, and they and the distance parameters of the basic functions $g^k(\mathbf{h})$ must be estimated together such that for all pairs of variables u and v and all k $b_{uu}^k \geq 0$, $b_{vv}^k \geq 0$, and

$$\left| b_{uv}^k \right| = \left| b_{vu}^k \right| \leq \sqrt{b_{uu}^k b_{vv}^k}$$

We fitted the model by first computing and modelling the auto-variogram of the first principal component (see below) to obtain a sensible structure and estimates of the distance parameters. With these distances fixed we found the b_{uv}^k by iteration using the algorithm of Goulard and Voltz (1992).

For each k the b_{uv}^k form a variance–covariance matrix which can be decomposed into its latent roots, λ_i, $i = 1, 2, \ldots, V$, and corresponding latent vectors, a_{ij}, effectively a principal component transformation. The latter can be converted to the correlation coefficients between the original variables and the principal components:

$$c_{ij} = a_{ij} \sqrt{\lambda_j / \sigma_i^2} \tag{8}$$

where σ_i^2 is the variance of the original ith variable. When the c_{ij} are plotted in unit circles they show the relations between the variables at the spatial scales of the model.

The precise details of the functions fitted, both in the coregionalization and to the auto-variograms, are given in the section on results below.

RESULTS AND INTERPRETATION

Table I summarizes the data. The box-and-whisker plots of some of the trace elements (As, Cu, and Pb in Figure 3) show a few large outliers. Those of Cu and Pb almost certainly result from human-derived contamination as they occur around towns. The majority of the exceptionally large concentrations of As, however, are likely to be

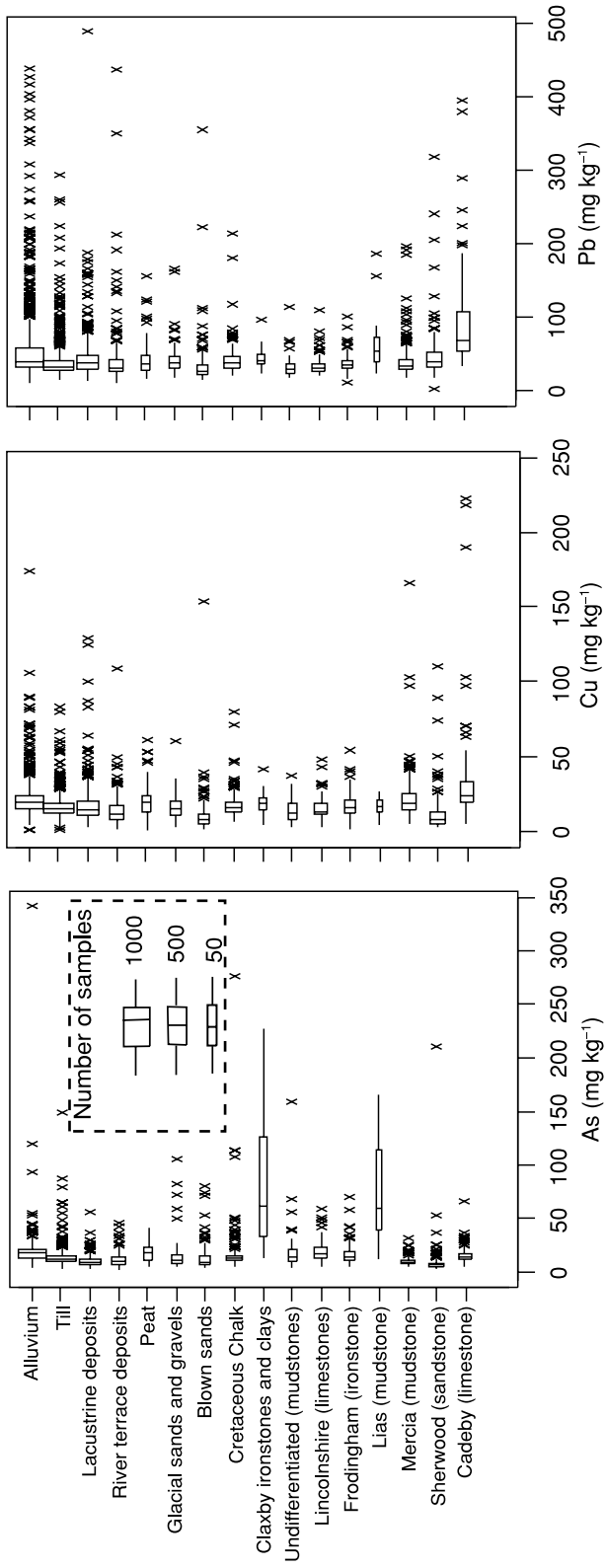

Figure 3. Box-and-whisker plots of As, Cu and Pb concentrations (mg kg^{-1}) classified by 16 parent material groups, with outlying values shown. The two vertical lines which form the left and right ends of each box represent the 25th and 75th percentiles (respectively) of the distribution in each class. The line across the box represents the median (50th percentile). The whiskers extend from each end of the box, up to a maximum of one and a half times its length, where larger and smaller values occur in each class. The crosses identify outliers beyond this range. Box width is proportional to the number of samples ($n = 4609$)

natural, as they occur in soil developed on the Cretaceous and Jurassic ironstones (Figures 2 and 3) which contain pyrite and other iron-bearing minerals with which As is commonly associated (Taylor and Curtis, 1995). These and several other elements were moderately or strongly positively skewed (skewness > 1), and we transformed them to their square roots or common logarithms to stabilize their variances for further statistical analysis. Table I lists the statistics of the transformed data in addition to those on the original scales of measurement.

Analysis of variance

Using digital versions of the map of parent materials (Figure 2), we classified the sampling sites in two ways, first by lithostratigraphy into 17 major types of bedrock lithology (Figure 4a), and second, into 16 types of parent material, on the assumption that the soil formed from Quaternary deposits where these are present, otherwise taking the bedrock (lithostratigraphic) classification (Figure 4b). We could not take into account the thicknesses of the outcrops, and so any effect of thickness does not appear in the classification.

An analysis of variance was done for each of the elements for these two classifications to determine the proportion of variance accounted for (Table II). The classification by parent material accounted for more of the variance of the elements than did that of lithostratigraphy, except for Pb. In order of decreasing explained variance, the elements most closely related to parent material were: Mg (48 per cent) > As (36 per cent) > Fe (34 per cent) > Ti = Ni = Ca (32 per cent). Phosphorus and Pb were the two elements with distributions that were least closely related to parent material (variances explained 14 and 17 per cent, respectively).

Isotropic variograms

Experimental variograms were computed for each of the 13 elements with the standard estimator for the semi-variance, Equation 2, at 1 km intervals regardless of direction. Valid models were then fitted to them. Figure 5 shows the results. The horizontal lines at the bottoms of the graphs show the sizes of the analytical error from the sub-samples, i.e. the laboratory error. All the variograms seemed bounded; and all except the global variograms of Mg and Ca were best fitted by double spherical functions with nugget:

$$\begin{aligned}\gamma(h) &= c_0 + c_1\left\{\frac{3h}{2a_1} - \frac{1}{2}\left(\frac{h}{a_1}\right)^3\right\} + c_2\left\{\frac{3h}{2a_2} - \frac{1}{2}\left(\frac{h}{a_2}\right)^3\right\} \quad \text{for} \quad 0 < h \le a_1 \\ &= c_0 + c_1 + c_2\left\{\frac{3h}{2a_2} - \frac{1}{2}\left(\frac{h}{a_2}\right)^3\right\} \quad \text{for} \quad a_1 < h \le a_2 \\ &= c_0 + c_1 + c_2 \quad \text{for} \quad h > a_2 \\ &= 0 \quad \text{for} \quad h = 0 \end{aligned} \tag{9}$$

The nugget variances of the global variograms ranged between 27 per cent (Mg and Fe) and 63 per cent (P). The nugget variance represents the combination of variation at distances less than the shortest sampling interval (approximately 800 m) plus laboratory error. We know from the analysis of sub-samples that the latter is only a small component of the nugget variance, except for Mo and U in which it represents about a fifth and a third of the nugget variance, respectively (Figure 5).

Phosphorus has the largest proportion of nugget variance. This is almost certainly because farmers have been applying phosphorus fertilizer in this predominantly arable region for many years and created a large field-to-field variation. Those elements that we might expect to be most closely related to parent material also had the smallest nugget variances (Mg, As, Fe).

The elements with the largest relative variance over shorter lag distances (steep gradients near the ordinate) include As, Cr, Fe, Ni and U, whereas Ca and Mg have smaller proportions of their variance over short distances (gentler gradients near the ordinate, Figure 5). The short-range structures of the global variograms (Figure 5) have ranges or effective ranges of between 3·5 (Cu and Fe) and 9·5 km (Al). The ranges of the long-range structures (which mark the limits of spatial dependence) vary between 15·4 (P) and 23·6 km (Pb).

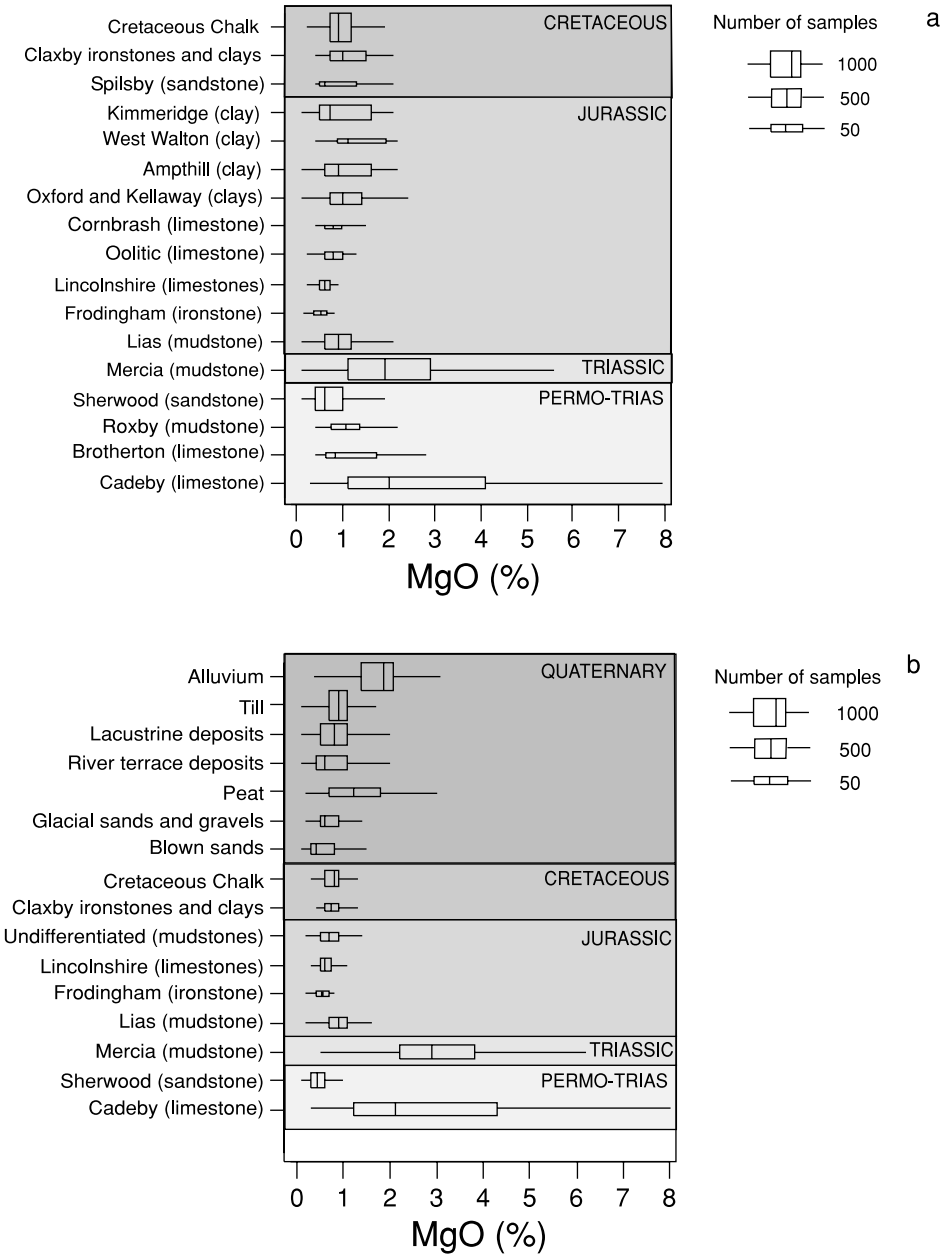

Figure 4. Box-and-whisker plot of MgO concentrations in soil for (a) 17 lithostratigraphic and (b) 16 parent material groups. See caption to Figure 3 for an explanation of box-and-whisker plots. Box width is proportional to the number of samples ($n = 4609$)

Residuals from the means of the 16 parent material classes were calculated, and from them were computed experimental pooled within-class variograms for each element. Models were fitted to them, too. The variograms are plotted in Figure 5, and the estimated parameters of the models are in Table III. Their sill variances are less than those of the regional variograms by an amount approximately equal to the differences in their variances (Table II). Although estimated distance parameters of the short-range structures are very different from those of the global variograms, the estimated parameters of the long-range structures are little changed (Table III), except for Al. The long-range structure is most probably related to the distribution of major parent material types in the region.

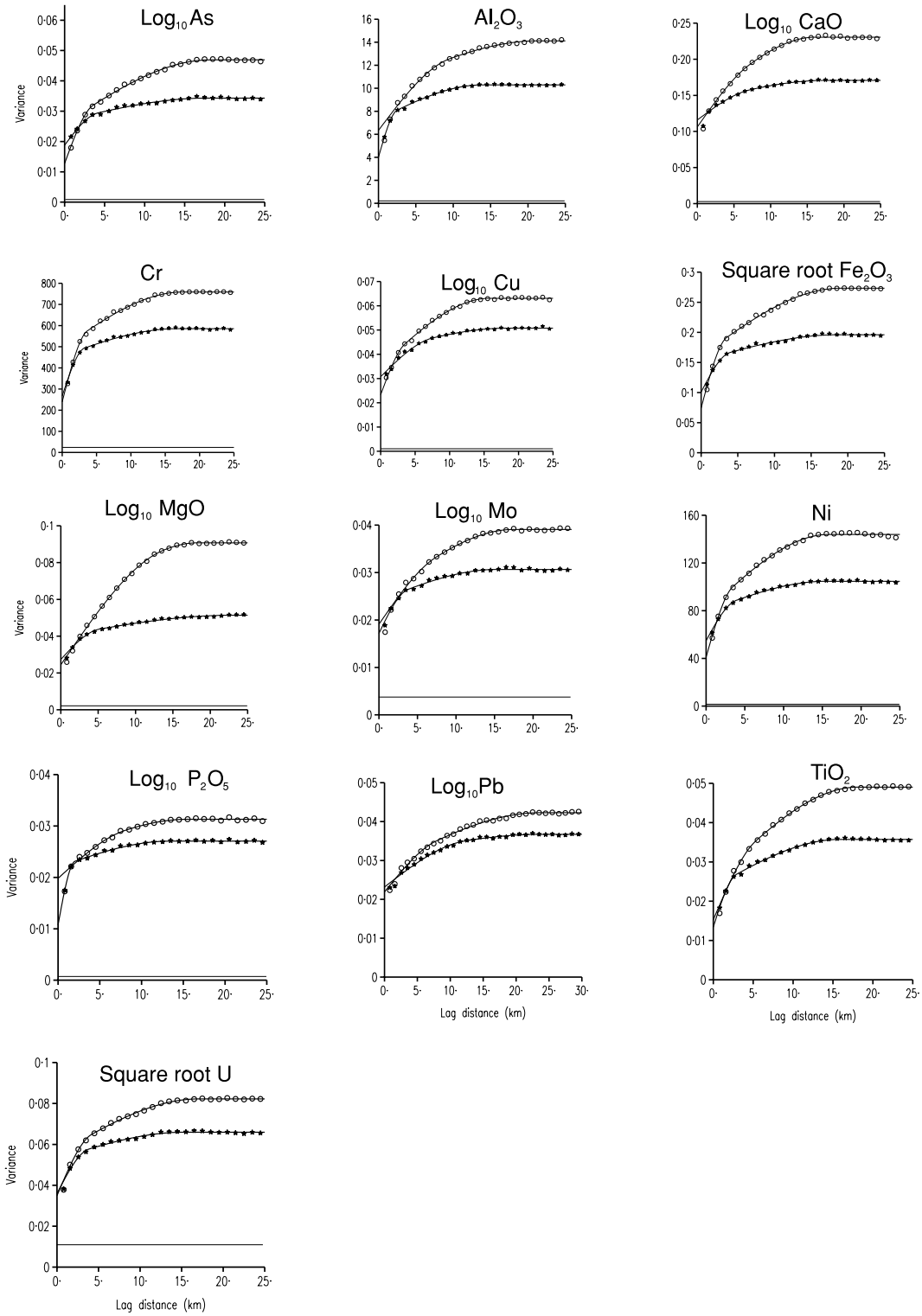

Figure 5. Global (o) and pooled within-class experimental variograms (*) for 13 elements and lines showing the models fitted to them. The horizontal line at the base of each plot shows the analytical error

Table III. Estimates of the nugget variances (%) and ranges of models fitted to the global and pooled within-class experimental variograms (Figure 5)

	Global variogram				Pooled-within class variogram			
	Model	Nugget (%)	Range 1 (km)	Range 2 (km)	Model	Nugget (%)	Range 1 (km)	Range 2 (km)
Al_2O_3	doublesph*	45	9·5	21·3	doublesph	39	2·5	13·9
As	doublesph	27	3·6	18·2	doublesph	55	4·2	17·7
CaO	spherical	46		16·1	doublesph	68	7·2	17·3
Cr	doublesph	31	3·7	17·5	doublesph	46	3·3	16·6
Cu	doublesph	37	3·5	15·8	doublesph	61	7·7	18·4
Fe_2O_3	doublesph	27	3·5	19·0	doublesph	51	3·9	18·6
MgO	spherical	27		18·3	doublesph	53	5·2	22·9
Mo	doublesph	49	7·4	17·5	doublesph	56	3·8	15·2
Ni	doublesph	28	3·5	15·8	doublesph	52	3·9	15·4
P_2O_5	doublesph	63	8·1	15·4	doublesph	40	2·0	13·3
Pb	doublesph	52	7·9	23·6	doublesph	63	12·3	23·5
TiO_2	doublesph	31	5·4	18·1	doublesph	38	2·9	15·9
U	doublesph	42	4·2	17·0	doublesph	54	3·9	15·3

* doublesph, double spherical.

ANISOTROPY

Experimental variograms were computed with the standard estimator, Equation 2, in four directions, 0, 45, 90, and 135°, following mathematical convention, counter-clockwise from east, in bands 800 m wide. Variation in several elements (Ca, Cr, Fe, Mg, Mo, Ni, Ti, U) appeared anisotropic; that of Mg was the most obviously so. Experimental variograms typically reached their maxima in the west–east dimension (direction 0) within 7–11 km. The variograms in the north–south dimension (direction 90) appeared still to be increasing after more than 24 km, and this made their modelling with coherent spherical functions somewhat difficult.

Figure 6 shows the example of Mg. The sample semivariances for the four directions are distinguished by the four symbols, and it is evident that variances of pairs of sample points in a north–south direction are substantially smaller than the others. We could fit a geometric anisotropic exponential model:

$$\gamma(h, \theta) = c_0 + c\left\{1 - \exp\left(-\frac{|\mathbf{h}|}{\Omega(\theta)}\right)\right\} \tag{10}$$

In this equation

$$\Omega(\theta) = \sqrt{A^2\cos^2(\theta - \phi) + B^2\sin^2(\theta - \phi)}$$

where A is the maximal distance parameter of the exponential model in direction ϕ, and B is the (minimum) distance parameter in the perpendicular dimension. This model (Figure 6b) provides the best fit to 24 km, but it is not very good. The curves are the envelope of the model. If we restrict the fitting to 15 km then the anisotropic power function

$$\gamma(h, \theta) = c_0 + \Omega(\theta)\,|\mathrm{h}|^{\alpha} \tag{11}$$

fits reasonably well (Figure 6a), and this is the model we have chosen for kriging. It is entirely satisfactory for this purpose because the data for any one estimate all lie well within a 15 km radius. The parameters of the model were as follows: $\phi = -0{\cdot}02046$ radians, $\alpha = 0{\cdot}7167$, the minimum gradient $B = 0{\cdot}007935$ km^{-1}, the maximum gradient $A = 0{\cdot}01337$ km^{-1}, and the nugget $c_0 = 0{\cdot}01840$.

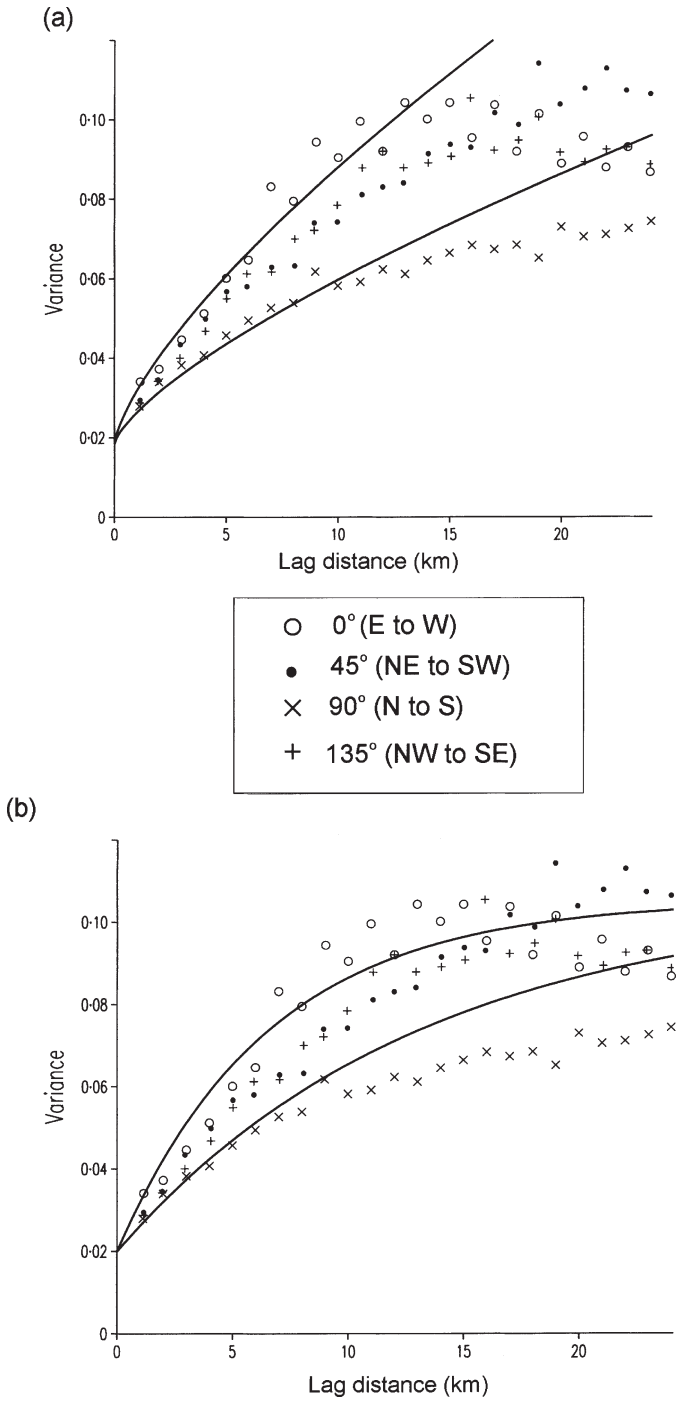

Figure 6. Semivariances of MgO in four directions and bounds of the anisotropic models fitted to them: (a) power function; and (b) exponential function

MAPPING

To demonstrate the influence of parent material on the soil's geochemistry at the regional scale we kriged the concentrations of MgO on a fine grid at intervals of 500 m from the data and the anisotropic model of the

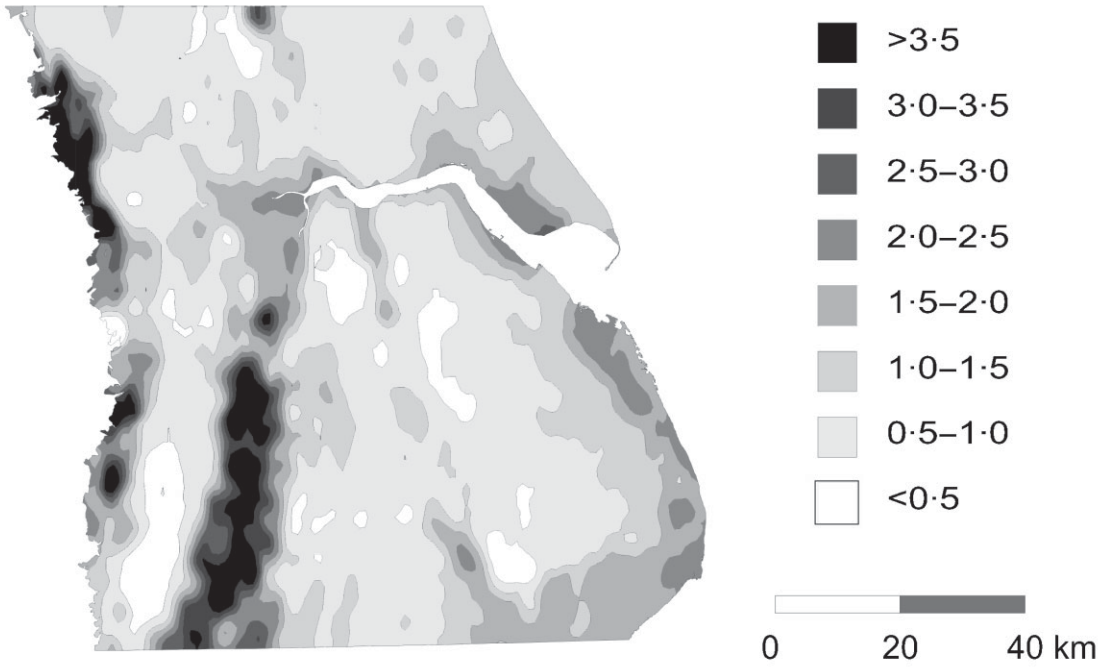

Figure 7. Kriged map of MgO concentration (%) after back transformation

variogram, Equation 11, and these were mapped (Figure 7). Notable spatial features include the large concentrations (>3 per cent) in soil over the Cadeby (Magnesian) Limestone and the Mercia Mudstone Group, although they have a broad range (0·5–8 per cent; Figure 4b). The large concentrations on the Magnesian Limestone are attributable to dolomite, $CaMg(CO_3)_2$, whereas on the Mercia Mudstone clays (mica and chlorite) are the main source, though there is dolomitic cement in some parts of the region (Bloodworth and Prior, 1993). One of the most striking features of the spatial pattern is that associated with the marine alluvium along the coast of Lincolnshire and Humberside. There soil overlying it has moderate median concentrations of MgO (2 per cent), although large values up to 3·5 per cent are attributable to the abundant Mg-bearing micaceous and chloritic clay mineral assemblages (Loveland, 1984). The spatial distribution of Mg concentrations in soil to the west of the Humber estuary also correlates closely with the underlying alluvium.

COREGIONALIZATION

The degree of correlation between major and trace elements over varying spatial scales reflects both the composition of the parent material and the outcome of pedogenesis. The Quaternary sediments have been sorted according to the environment in which they were deposited, and this sorting has almost certainly altered the associations between elements in the source rock and those in the sediments today. The most resistant soil minerals (the residual fraction) reflect the composition of the parent material. Weathering of the minerals and chemical and physical reactions between soil constituents (coprecipitation, cation exchange, adsorption and organic complexation) influence the associations between major and trace elements. We use our knowledge of geochemistry and pedogenic processes to interpret these associations. Although complexation reactions between trace elements and organic matter (and to a lesser extent manganese oxides) may significantly influence the spatial distribution of the former, we cannot comment on the significance of its role without data.

Only one element, namely P, was not closely correlated with the others, and we explained above that this is almost certainly because farmers have added it as fertilizer, field by field. This additional P is not of natural origin. We therefore excluded the P data from our model of coregionalization, which embraced the 12 elements in our list.

Table IV. Eigenvalues of the principal components analysis

Order	Eigenvalue	Percentage	Accumulated percentage
1	6·663	55·1	55·1
2	1·467	12·1	67·2
3	1·02	8·4	75·6
4	0·815	6·7	82·3
5	0·483	4·0	86·3
6	0·394	3·3	89·6
7	0·293	2·4	92·0

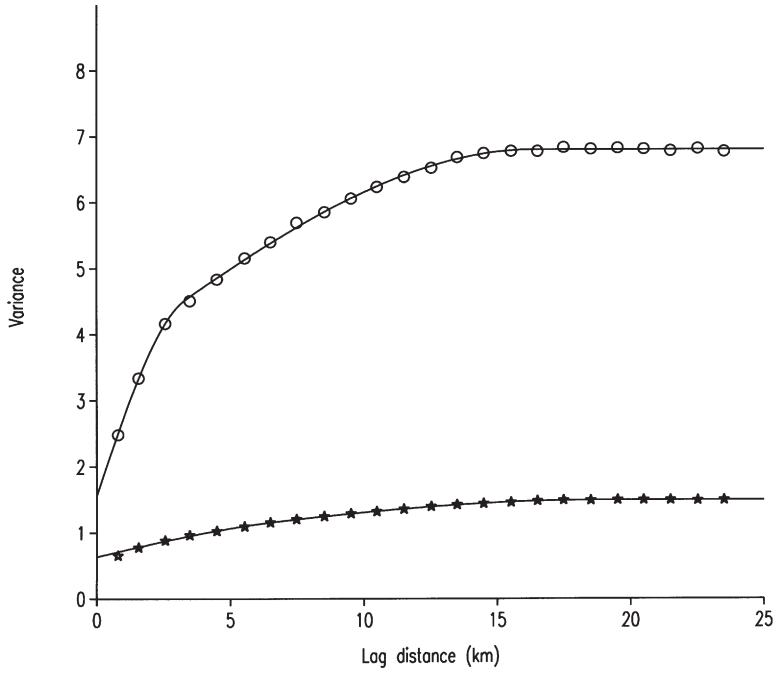

Figure 8. Experimental variograms of the first (o) and second (*) principal components and models fitted to them

We proceeded as described above. We first computed the principal components on the correlation matrix of the variables (transformed as appropriate) and the variograms of their scores. Table IV lists the eigenvalues. The first component accounted for 55 per cent of the variance, and we took its variogram as a reasonable summary of the spatial characteristics of all 12 original variables. We modelled it with a double spherical function plus nugget (Figure 8). The distance parameters were $a_1 = 3000$ m and $a_2 = 16\,000$ m, and we used these in the model of coregionalization. We thereby obtained three variance–covariance matrices, $\mathbf{B}^k$, of coefficients b^k_{uv} . Matrix $\mathbf{B}^1$ contains the nugget variances, $\mathbf{B}^2$ contains the sills of short-range structure with $a_1 = 3000$ m, and $\mathbf{B}^3$ those of the long-range structure with $a_2 = 16\,000$ m. They are listed in Table V. The linear model of coregionalization proved to fit remarkably well to each of the 12 global autovariograms (Figure 9).

Finally, we did principal component analyses on the three matrices of b coefficients in addition to that of the global correlation matrix. Table VI lists the eigenvalues and their cumulants for the three matrices, and in Figure 10 we plot the correlations between the principal component scores and the original variates (after transformation as described) in the planes of the first two axes. We have added the unit circles to aid interpretation.

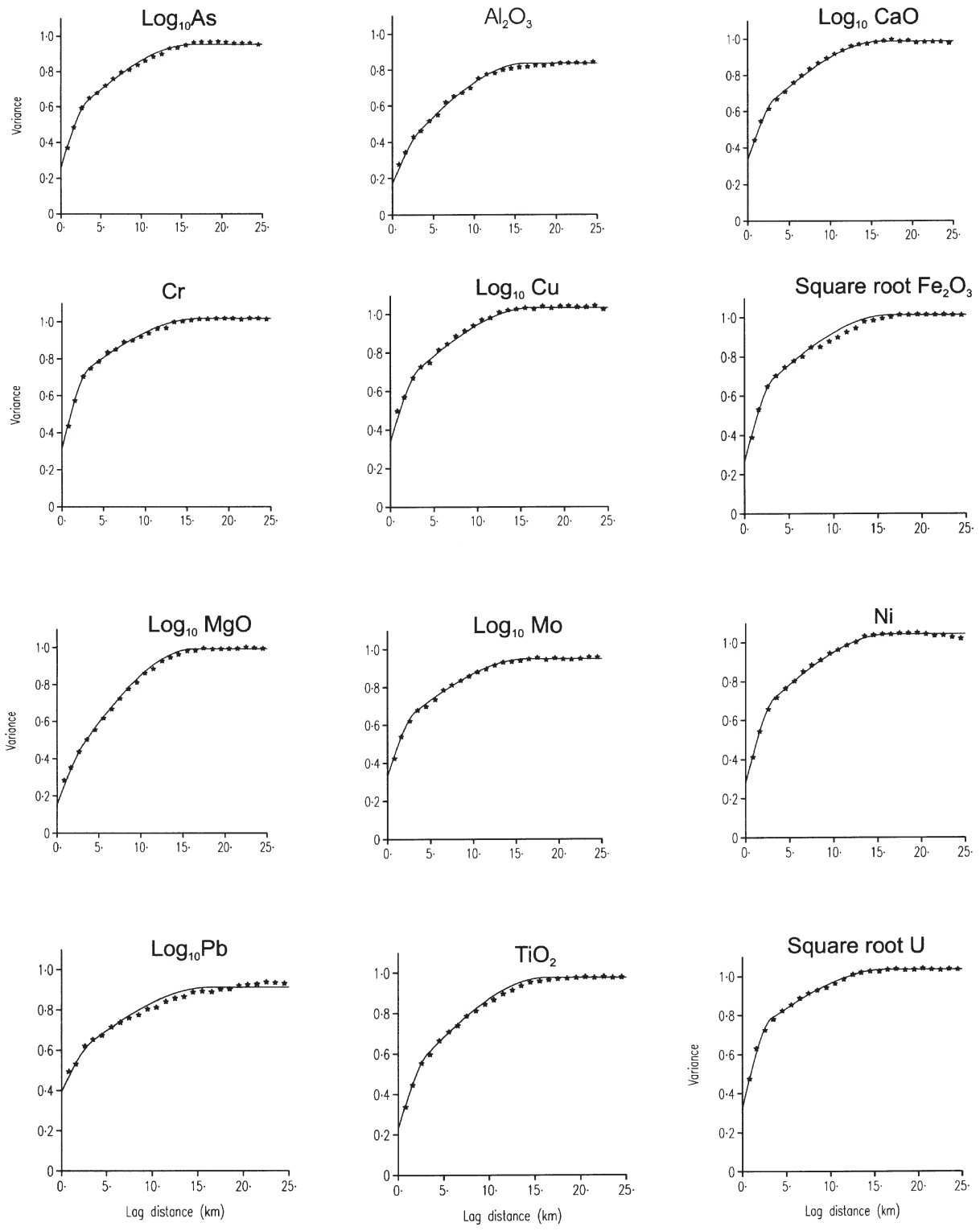

Figure 9. Experimental variograms (*) and lines showing the model of linear coregionalization for 12 elements

Table V. Coefficients (b^k) of double spherical coregionalization model for standardized variograms

	Nugget (b^1)	Short range (b^2)	Long range (b^3)
As	0·261	0·236	0·455
Al_2O_3	0·171	0·133	0·534
CaO	0·338	0·200	0·449
Cr	0·310	0·328	0·380
Cu	0·338	0·247	0·450
Fe_2O_3t	0·268	0·290	0·455
MgO	0·163	0·117	0·712
Mo	0·332	0·227	0·392
Ni	0·281	0·299	0·464
Pb	0·394	0·131	0·387
TiO_2	0·231	0·215	0·532
U	0·323	0·356	0·356

Table VI. Eigenvalues and cumulants of the structural variance-covariance matrices

	Nugget		Short-range		Long-range	
Order	Eigenvalue	Accumulated percentage	Eigenvalue	Accumulated percentage	Eigenvalue	Accumulated percentage
1	1·6471	48·3	1·9554	70·4	3·3507	59·8
2	0·5623	64·8	0·2897	80·8	0·8950	75·8
3	0·3054	73·7	0·2014	88·0	0·4324	83·5
4	0·2058	79·8	0·1195	92·3	0·2896	88·7
5	0·1874	85·3	0·0804	95·2	0·1810	91·9
6	0·1469	89·6	0·0529	97·1	0·1458	94·5
7	0·1306	93·4	0·0386	98·5	0·0997	96·3
8	0·0855	95·9	0·0269	99·5	0·0674	97·5
9	0·0583	97·6	0·0147	100·0	0·0398	98·2
10	0·0497	99·1	0	100·0	0·0292	98·8
11	0·0230	99·8	0	100·0	0·0210	99·1
12	0·0065	99·9	0	100·0	0·0145	99·4

Each graph has two features to interpret, as follows.

(1) The closer the elements lie to the circumference of the circle, the better they are represented by the two dimensions. It is a manifestation of the proportion of the total variance in the corresponding two eigenvalues. This attribute of graphs for the different spatial structures allows us to compare the degrees to which the components of the model represent the regional variation.
(2) The closeness with which the elements lie to one another in any one graph is a measure of their correlation in those dimensions at that scale. By comparing the configurations of points in the unit circles we can see how correlations between the elements change with changing scale.

The Pearson product–moment correlation coefficients indicate several geochemical associations with many exceeding 0·5; for example, Fe is associated with As and Cr as a result of chemical adsorption, and Fe with Ti are in the mineral ilmenite (Table VII). The close correlation between these elements may also be due to the association of Al-bearing micas with Ti oxides in clay-size fractions (<2 μm) and the adsorption of iron oxides on clays. The association between As and Mo decreases as the separating distance increases. At very short distances (nugget correlations) it may be attributable to their adsorption on oxyhydroxides, whereas Mo is associated with the clay signature (Al and Mg) in the short-range structure. The close correlations of U with Ti and Fe over short distances (Figure 10c) is most probably due to a combination of both primary mineral content

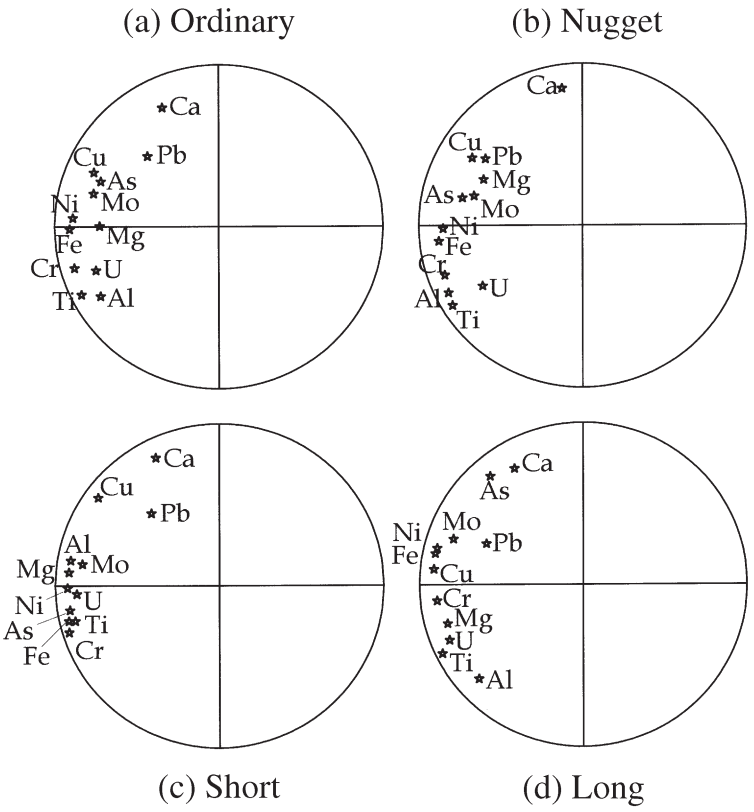

Figure 10. Projections of the correlation between the global (standardized) variables and the principal component scores into unit circles in the plane of the first two components: (a) general correlation matrix; (b) nugget matrix; (c) short-range correlation matrix; (d) long-range correlation matrix. Axis 1 horizontal, axis 2 vertical

Table VII· Correlation matrix

	Al_2O_3	As	MgO	CaO	Cr	Cu	Fe_2O_3	Mo	Ni	P_2O5	TiO_2	Pb	U
Al_2O_3	1·00												
As	0·15	1·00											
MgO	0·29	0·03	1·00										
CaO	−0·20	0·06	0·16	1·00									
Cr	0·69	0·50	0·29	−0·20	1·00								
Cu	0·21	0·14	0·28	0·09	0·29	1·00							
Fe_2O_3	0·51	0·73	0·23	−0·07	0·81	0·28	1·00						
Mo	0·39	0·46	0·22	0·12	0·57	0·34	0·66	1·00					
Ni	0·57	0·45	0·41	0·12	0·81	0·38	0·77	0·57	1·00				
P_2O_5	−0·25	0·22	−0·12	0·30	−0·12	0·13	0·07	0·15	−0·03	1·00			
TiO_2	0·75	0·17	0·35	−0·26	0·83	0·27	0·59	0·41	0·72	−0·32	1·00		
Pb	0·06	0·11	0·18	0·04	0·17	0·62	0·17	0·21	0·21	0·10	0·12	1·00	
U	0·52	0·16	0·35	−0·12	0·59	0·26	0·50	0·50	0·55	−0·13	0·66	0·12	1·00

(ilmenite and magnetite; Milodowski *et al*., 2000) and weathering leading to U adsorption by Fe oxyhydroxides (Gueniot *et al*., 1988; Flexser *et al*., 1993) with which they remain during subsequent transport. The chemical adsorption of As and Cr by oxyhydroxides may also explain their close associations with Fe (Figure 10c).

Calcium and Mg are typically closely associated geochemically in soil because they occur together in carbonate and certain clay minerals, and behave similarly in cation exchange reactions. However, throughout the region,

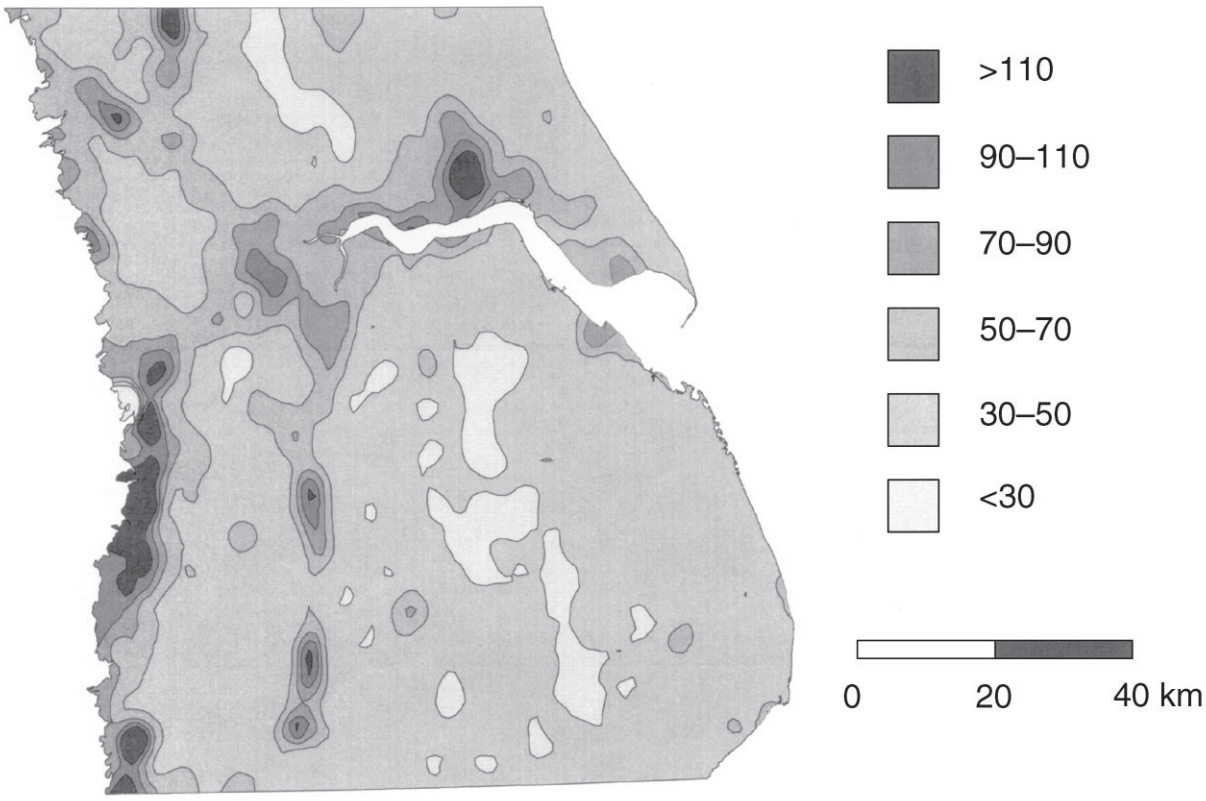

Figure 11. Kriged map of Pb concentrations (mg kg^{-1}) after back transformation

both at individual sites and spatially, their correlations were weak. This may be explained by the parent materials and their distributions. The largest and smallest concentrations of each element tend to occur (except for the Permo-Triassic limestone) in different parts of the region.

Magnesium is most abundant in the soil over the Cadeby (Magnesian) Limestone, marine alluvium and Mercia Mudstone, whereas the largest concentrations of Ca are in the soil on the Cretaceous Chalk (98 per cent pure $CaCO_3$) and Jurassic Limestone. Calcium is spatially correlated with As only at long distances, and this seems to be related to the ironstones and clays of the Cretaceous and Jurassic, and to a lesser extent to the other kinds of Jurassic rock.

Lead is somewhat anomalous; it plots closest to the centres of the projections and away from all the other elements, except for a close correlation with Cu in the nugget projection (very short distance). Webster *et al.* (1994) made a similar observation in a study of soils of the Swiss Jura. A map of Pb made by kriging at 500 m intervals using the variogram parameters listed in Table III (and illustrated in Figure 11) shows 'haloes' around several large towns (including Kingston-upon-Hull, York, Doncaster, Linclon and Nottingham). More widespread diffuse pollution might have modified its natural distribution throughout the region, as it seems to have done in the topsoil on the Carboniferous Coal Measures immediately to the west (Rawlins *et al.*, 2002). Copper, Ni and to a lesser extent Cr, are closely correlated with Fe in the long-range structure. Nickel and Cu have similar electronegativity, and they bind strongly to Fe oxyhydroxides and organic matter, which may account for their close correlation.

DISCUSSION

We have demonstrated that parent material is a significant factor in determining major and trace element concentrations in topsoil for a large region of England, and that the magnitude of its influence varies considerably

for the 13 elements studied. We have also shown an example of the anisotropic nature of the variation (over distances as short as 5 km) which we found in several of the elements, and which matches the orientation of the underlying geological structure.

The ratios of $b^1 : b^2 : b^3$ (the coefficients in Equation 7 shown in Table V) of several elements appear to form distinct groups; Al_2O_3, MgO and TiO_2 are dominated by the long-range structure (b^3), whilst As, Cr, Fe_2O_3t, Ni and U have larger relative contributions from the short-range structure (b^2). Our knowledge of soil geochemistry suggests that the first group is related to the spatial distribution of aluminosilicate, clay and Ti oxide minerals (of which the latter typically occur in the clay size fraction) over the longer-range structure. Elements in the second group are likely to be related to the distribution of Fe oxyhydroxides (to which they are adsorbed) over shorter distances, although reactions with the surfaces of organic matter and manganese oxides are also likely to be significant.

The anisotropy which we demonstrated in the topsoil geochemistry of MgO, is related to the spacing between geochemically (and mineralogically) distinct parent material groups and their clear orientation with the underlying geological structure. Without further data at the same resolution we cannot determine whether soil geochemistry throughout England and Wales exhibits anisotropy at this scale. However, the spacing between parent material groups (and their clear orientation with the underlying geological structure) in the study region appears to be representative of much of England and Wales (Institute of Geological Sciences 1977, 1979), and we would therefore expect similar relationships at the national scale.

There are practical implications of these findings for estimating natural background concentrations of major and trace elements in soil. Where more detailed soil maps are not available for an area in which geochemical sampling has been sparse, a map of parent material types would be an effective means of classifying samples for the purpose of estimating regional mean values.

ACKNOWLEDGEMENTS

We thank all the BGS staff and volunteer workers involved in the collection and analysis of samples in the G-BASE project. This paper is published with the permission of the Director of the British Geological Survey (NERC).

REFERENCES

Bloodworth AJ, Prior SV. 1993. *Clay mineral stratigraphy of the Mercia Mudstone Group in the Nottingham Area*. Technical Report WG/93/29. British Geological Survey: Keyworth.

British Geological Survey. 1983. *Humber–Trent Sheet 53N 02W*. British Geological Survey: Keyworth.

Flexser S, Wollenberg HA, Smith AR. 1993. Distribution of radon sources and effects on radon emanation in granitic soil at Ben-Lomond, California. *Environmental Geology* **22**: 162–177.

Goovaerts P, Webster R. 1994. Scale-dependent correlation between topsoil copper and cobalt concentrations in Scotland. *European Journal of Soil Science* **45**: 79–95.

Goulard M, Voltz M. 1992. Linear coregionalization model: tools for estimation and choice of cross-variogram matrix. *Mathematical Geology* **24**: 269–286.

Gueniot B, Munierlamy C, Berthelin J. 1988. Geochemical behavior of uranium in soils. 1. Influence of pedogenetic processes on the distribution of uranium in aerated soils. *Journal of Geochemical Exploration* **31**: 21–37.

Institute of Geological Sciences. 1977. *Quaternary Map of the United Kingdom*. Ordnance Survey (for Institute of Geological Sciences): Southampton.

Institute of Geological Sciences 1979. *Geological Map of the United Kingdom (South)*. Ordnance Survey (for Institute of Geological Sciences): Southampton.

Loveland PJ. 1984. The soil and clays of Great Britain: 1. England and Wales. *Clay Minerals* **19**: 681–707.

McBratney AB, Webster R, McLaren RG, Spiers RB. 1982. Regional variation in extractable copper and cobalt in the topsoil of south-east Scotland. *Agronomie* **2**: 969–982.

McGrath SP, Loveland PJ. 1992. *The Soil Geochemical Atlas of England and Wales*. Blackie Academic and Professional: Glasgow.

Milodowski AE, Hyslop EK, Kemp SJ, Gillespie MR, Chenery SR, Bland DJ, Prior SV, Strong GE, Wetton PD. 2000. *The mineralogical, petrographical and uranium and thorium distribution characteristics of Sellafield borehole drillcore samples selected for the Nirex safety assessment research programme*. Report WG/94/34. British Geological Survey: Keyworth.

Oliver MA, Webster R, McGrath SP. 1996. Disjunctive kriging for environmental management. *Environmetrics* **7**: 333–358.

Payne RW. 2002. *The Guide to GenStat. Part 2 Statistics*. VSN International: Oxford.

Rawlins BG, Lister TR, Mackenzie A. 2002. Trace metal pollution of soils in northern England. *Environmental Geology* **42**: 612–620.

Scholz M, Oliver MA, Webster R, Loveland PJ, McGrath SP. 1999. Designing a soil monitoring network. In *GeoENV II–Geostatistics for Environmental Applications*, Gómez-Hernández J, Soares A, Froidevaux R (eds). Kluwer Academic Publishers: Dordrecht; 465–476.

Straw A, Clayton KM. 1979. *Eastern and Central England*. Methuen: London.
Taylor KG, Curtis CD. 1995. Stability and facies association of early diagenetic mineral assemblages – an example from a Jurassic ironstone–mudstone succession, UK. *Sedimentary Petrology and Processes, Section A* **65**: 358–368.
Webster R, Oliver MA. 2001. *Geostatistics for Environmental Scientists*. John Wiley & Sons: Chichester.
Webster R, Atteia O, Dubois J-P. 1994. Coregionalization of trace-metals in the soil in the Swiss Jura. *European Journal of Soil Science* **45**: 205–218.
Webster R, Oliver M, Loveland P, Frogbrook Z, McGrath S. 2000. Factorial kriging of a soil inventory. In *Geostatistics 2000 Cape Town*, volume 2, Kleingeld WJ, Krige DG (eds). Geostatistical Association of Southern Africa: South Africa; 623–632.