The singularity index for soil geochemical variables, and a mixture model for its interpretation

R.M. Lark[1*], M. Patton[2], E.L. Ander[1], D.M. Reay[2]

[1] *British Geological Survey, Keyworth, Nottinghamshire NG12 5GG, U.K.* [2] *Geological Survey of Northern Ireland, Dundonald House, Upper Newtownards Road, Ballymiscaw, Belfast BT4 3SB, U.K.*

Abstract

A geochemical anomaly is a concentration of an element or other constituent in a medium (soil, sediment or surface water) which is unusual in its local setting. Geochemical anomalies may be interesting as indicators of processes such as point contamination or mineralizations. They may therefore be practically useful, indicating sources of pollution or mineral deposits which may be of economic value. As defined, a geochemical anomaly is not merely a large (or small) concentration of a constituent as compared to the marginal distribution. To detect anomalies we must therefore do more than simply map the spatial distribution of the constituent. One proposed approach makes use of a singularity index based on fractal representation of spatial variation. The singularity index can be computed from local concentration measures in nested windows. In this paper we propose an approach to compute threshold values for the index to identify enrichment and depletion anomalies, separate from background information. The approach is based on a mixture model for the singularity index, and it can be supported by computing a distribution for background values of the index by parametric bootstrapping from a robustly-estimated variogram model for the target constituent. This approach is illustrated here using data on elements in the soil in four settings in Great Britain and Ireland.

Keywords. Geochemistry; Anomalies; Singularity; Fractal; Mixture model

[*] Now at University of Nottingham, Sutton Bonington Campus, Sutton Bonington, LE12 5RD, U.K. *E-mail address*: murray.lark@nottingham.ac.uk (R.M. Lark).

1. Introduction

1.1 The problem

Soil geochemical data comprises information on the concentration of elements in soil (e.g. heavy metals, micronutrients such as selenium and potentially harmful elements such as As), compounds (e.g. specific organic pollutants, ions such as nitrate or phosphate) and other constituents such as organic carbon. The soil may be a convenient medium for geochemical survey (e.g. Breward, 2007) focussed on mineral exploration or to support geological mapping. Soil geochemical data may also support the management of agricultural soils (e.g. Lark et al., 2014) or the assessment of particular threats to soil quality (e.g. Rawlins et al., 2006). In all cases a common objective in the analysis of soil geochemical data (as with data in other media such as stream sediments), is the identification of anomalies. A geochemical anomaly is a measurement, or local cluster of measurements, which have markedly large or small concentrations in local context. Anomalies may be important as indicators of mineralizations which could be economically important, or they may reflect point pollution processes which must be understood for environmental protection.

The detection of anomalies requires more than the mapping of large or small concentrations. Rather it is the identification of local accumulation or depletion which is anomalous in context. One method that has been used to tackle this problem invokes a multifractal model of spatial variation under which variation may include local singularities (e.g. Chen et al., 2007). This paper proposes an approach to the detection of anomalies in data on soil which is based on this method. The next section outlines the approach based on singularities in more detail. The methods used in this paper are then described (section 2.1) and then applied in four case studies on concentrations of elements in four contrasting settings in the United Kingdom and Ireland.

1.2 Anomalies and singularities

In the approach to anomaly detection based on a multifractal model the local anomalous accumulation of material (or equivalently, depletion), arising from local heterogeneities or cascade processes is treated as a *singularity* characterized by the local singularity index. A fuller account of the underlying theory is given by Cheng (2007; 2012) and Agterberg (2012), but we summarize here.

We denote a local support in d dimensions (e.g. a square or circle when $d = 2$) centred at location $\mathbf{x}$ and of (linear) size ε by $\mathcal{B}_{\mathbf{x}}(\varepsilon)$. The amount of some material within the support, $\mu(\mathcal{B}_{\mathbf{x}}(\varepsilon))$ depends on the local background concentration $c(\mathbf{x})$ scaled according to a local singularity index, $\alpha(\mathbf{x})$:

$$\mu(\mathcal{B}_{\mathbf{x}}(\varepsilon)) \;=\; c(\mathbf{x})\varepsilon^{\alpha(\mathbf{x})}. \tag{1}$$

The equivalent expression for the mean concentration over the support is

$$\rho(\mathcal{B}_{\mathbf{x}}(\varepsilon)) \;=\; c(\mathbf{x})\varepsilon^{\alpha(\mathbf{x})-d}. \tag{2}$$

Allégre and Lewin (1995) reviewed the range of processes which give rise to observed distributions of geochemical variables. In many cases a normal or log-normal distribution may be expected under which the expected value of $\alpha(\mathbf{x})$ over a domain of interest is equal to d. In the presence of local anomalies, however, the variation is multifractal with local values of $\alpha(\mathbf{x}) < d$ where there is local enrichment of the material of interest and $\alpha(\mathbf{x}) > d$ where there is depletion.

For a multifractal process the set of points with a particular singularity index value itself constitutes a fractal set. This provides the basis for the practical approach taken to the identification of anomalies from the singularity index by the concentration-area model (Cheng, 2012). Under this model the area over which the singularity index is larger than some value, α, $A[> \alpha]$, the survival function of α, follows a power-law,

$$A[> \alpha] \;\propto\; \alpha^{-\beta}, \tag{3}$$

although there may be several values of β over distinct sub-ranges of the value of α. When the survival function is plotted on double-log axes these ranges should be revealed as linear segments of the plot. Liu et al. (2014) fit such linear segments and, from the break-points between them, identify a range of values of α which correspond to the background process and limits which define the range for enrichment and depletion anomalies respectively.

In this paper we consider case studies in which the singularity index was computed for the concentration of different elements in the topsoil across four different areas. In no case did the double-log plot of the empirical survival function of α clearly resemble a limited number of linear segments, rather, like any non-linear plot, it could be approximated to some arbitrary degree of accuracy by increasing numbers of such segments (see Figure 26) which makes the outcome for the range of values of the index assumed to correspond predominantly to background normal or log-normal variation essentially arbitrary. This is unsatisfactory. For this reason we propose an alternative approach. The singularity index under the normal or log-normal monofractal background model is assumed to have a distribution conditional on the spatial correlation of the variable, the distribution of the sample points and the scales examined. It is assumed that the distribution of the index for the whole field can be represented as a mixture of normal distributions, of which the dominant component represents the background. The mixture also includes one or more additional components which introduce mass into one or both tails of the overall distribution, representing anomalies. Note that previous workers have used mixture models for the concentrations of elements in soil to represent background and anomalous concentrations (e.g. Liu et al., 2010). It is important to remember that, in this paper, we model the singularity index rather than the concentrations themselves as a mixture of components.

In the remainder of the paper we describe the methods used and outline the results for the case studies.

2. Materials and Methods

2.1 Computation

The data used in this paper are described in detail in sections 2.2–2.5. In all cases the data were total concentrations of an element in the topsoil (soil to a depth of 15 cm from the surface). As described for each section, summary statistics and histograms of the data were obtained, and a decision was made as to whether a transformation was required prior to geostatistical analysis to ensure the plausibility of an assumption of normality (although the computation of the singularity index was done on the data on their original units of measurements, mg kg^{-1}).

2.1.1 The singularity index. In all case studies the singularity index was computed on the nodes of a 100-m square grid. At any node the mean concentration of the variable of interest was calculated within four local supports, each circular areas of radius 1000, 2000, 4000 and 8000 m. The ordinary least squares regression coefficient for the regression of log-transformed mean concentration on log-transformed radius of the circular support centered at $\mathbf{x}$ provides an estimate of $\alpha(\mathbf{x}) - d$. Because d is a constant (2 in this case with the analysis in two dimensions) the estimate of $\alpha(\mathbf{x})$ is easily obtained.

The ordinary least squares estimate of the regression coefficient must be treated with some caution in these circumstances. This is because the circular supports are nested within each other and so the residuals from the fitted line cannot be treated as independent. However, we make no assumptions of independence in any subsequent inferences, and the parametric bootstrapping of the background distribution of the index, described in section 2.1.3, explicitly reproduces this dependence.

For comparability we used the same radii for the four windows in all case studies. In the first study (section 2.2 below) with the sparsest sampling the mean concentration could not be evaluated at all nodes for the smallest radius, but checks showed that a value could be obtained for all radii of 2000 m or more.

2.1.2 Geostatistics. To allow the bootstrapping of the distribution of the background values of the singularity index we required a variogram function for the target variable. This was computed for the variable after any transformation. Because the aim is to obtain the variogram for the background process we wanted to minimize the effects of any outlying observations, including spatial outliers which appear unusual only in their local context, since such values can have a disproportionate effect on estimates of the variogram (Lark, 2000). We therefore estimated the variogram using the standard estimator due to Matheron (1962), but also the alternative estimators due to Cressie and Hawkins (1980), Dowd (1984) and Genton (1998). Lark (2000) reviews these estimators and provides further detail. We applied them using the georob package for the R platform (Papritz, 2016; R core team, 2014). Initially we estimated the variograms for different directions, but there was no marked anisotropy, particularly at lags corresponding to the scales at which the singularity index was computed, and so isotropic variograms were used.

Each variogram model was assessed by cross-validation. Each observation was removed from the data set in turn and predicted from the remaining ones by ordinary kriging. The standardized squared prediction error (SSPE) was computed for each cross-validation prediction:

$$\theta(\mathbf{x}) = \frac{\left\{\tilde{Z}(\mathbf{x}) - z(\mathbf{x})\right\}^2}{\sigma^2_{\mathrm{OK}}(\mathbf{x})}, \tag{4}$$

where $\tilde{Z}(\mathbf{x})$ is the ordinary kriging prediction at location $\mathbf{x}$, $z(\mathbf{x})$ is the corresponding observed value and $\sigma^2_{\mathrm{OK}}(\mathbf{x})$ is the corresponding ordinary kriging variance. The expected value of this statistic over all observations is one if the variogram model is correct, but outlying observations will affect both the numerator and denominator. We therefore observed the histograms of the cross-validation errors, and if the assumption that these were normal seemed plausible we examined the median value of the standardized squared prediction error. This should be close to 0.455 if the variogram model is correct, and will tend to be smaller if outliers are affecting the estimated variogram. On this basis one of the proposed variogram models was selected for further work.

2.1.3 Mixture modelling and parametric bootstrapping. The histograms of the singularity index values were examined. In all cases an assumption of a normal distribution with some additional mass in one or both tails seemed plausible, and so a mixture of normal distributions was fitted using the mixtools package in R, (Benaglia et al., 2009). The boot.comp procedure in the package was applied to a simple random sample of 1000 observations to evaluate the evidence that two or more components should be included in the model. This procedure undertakes bootstrap resampling of the likelihood ratio statistic to evaluate the evidence to include $p + 1$ rather than p component distributions for $p = 1, 2 \ldots$. Once the number of components had been selected a normal mixture model with this number of constituent distributions was fitted with the normalmixEM procedure. This estimates p sets of parameters (mean and standard deviation) and the proportions of each constituent distribution: $\lambda_1, \ldots, \lambda_{p-1}$. From these estimates one may compute the posterior density for each component at any value of α. Lower and upper threshold values, α_{L} and α_{U}, were identified to define the range of values of α to be identified with the background process. The threshold values were those such that the posterior density of the component of the mixture model identified with the background process was larger than that for any other component for some α where $\alpha_{\mathrm{L}} < \alpha < \alpha_{\mathrm{U}}$.

The next question is how to identify which component of the mixture model represents the background. We would expect it to have a mean close to 2, and to be the dominant component (largest value of λ). To support this inference, and to provide evidence for the appropriateness of the mixture model approach to analysis of the singularity index, we undertook a parametric bootstrap estimation of the parameters of the index for a normal random variable (perhaps after transformation) with the same variogram parameters as the data and the same spatial distribution. The process was as follows.

a. For the n data used in the computation of the observed values of the singularity index, compute the distance matrix and from this the covariance matrix given the estimated variogram parameters from the model selected by cross-validation.

b. Compute 1000 realizations of the normal variable with this covariance matrix, and the mean of the (transformed) values of the original data using the mvnorm procedure from the MASS package in R (Venables and Ripley, 2002).

c. For each realization in turn, back-transform the simulated values to the scale of the original measurements.

d. Compute the singularity index from these data at the same grid locations as used in the analysis of the real data, and with for the same set of windows.

e. Compute the mean and standard deviation of the singularity index of the simulated data over all nodes and record these statistics.

f. Iterate steps (c)–(d) for each realization.

2.2 Case study 1: Zn east of Dartmoor, south-west England

This study area (see Figure 1) is a region within which anomalous concentrations of soil zinc might be expected. This is because of the presence of stratiform sulphide mineralizations in the region (Benham et al., 2005) which has a history of metal mining from prior to the Roman occupation of Britain. Benham et al. (2005) report assessments of geological information, geophysical surveys, information on known deposits and geochemical data to support the expectation of such mineralizations. The latter data include anomalous zinc concentrations in streambed sediment surveys and identification of zinc enrichment in soils associated with magnetic anomalies identified by geophysical surveys.

The soil data from south-west England analysed in this paper were collected in 2013 according to the protocols of the Geochemical Baseline Survey of the Environment (G-BASE) of the British Geological Survey (Johnson et al., 2005). The sample points were located at more or less regular intervals, subject to constraints in the field, at a density locally of about one sample per 8 km^2. Each sample was a composite formed from cores collected at the centre and vertices of a 20-m square. The cores were length 15 cm excluding surface litter. Material was subsequently air-dried, disaggregated and sieved to pass 2 mm and sub-sampled by coning and quartering. A 50-g sub-sample was ground in an agate planetary ball mill until 95% of the material was finer than 53 µm. Total concentrations of each of some 50 elements were determined for each sample by X-Ray Fluorescence Spectrometry.

2.3 Case study 2: As in Counties Monaghan and Armagh, north of Ireland

The Longford-Down terrane in the north-east of Ireland has a history of mineral exploration; lead and zinc vein deposits have been worked since the nineteenth century. In recent years the area has been recognised for its gold prospectivity. Dublin based junior exploration company, Conroy Gold and Natural Resource, have defined a 30 mile gold trend with a number of gold targets identified. These include the Clontibret gold target in County Monaghan and additional targets that are identified along trend to the north east at Clay Lake in County Armagh and to the south west in County Monaghan. The gold mineralisation is associated with arsenopyrite and as such arsenic is used as a 'pathfinder' element for precious metal mineralisation. We therefore examined soil geochemical data for anomalous As concentrations.

The soil data from Ireland analysed in this paper were collected in two surveys, the Tellus survey of Northern Ireland and the Tellus Border survey of the six border counties of the Republic of Ireland. The study area is shown in Figure 1. The geochemical survey activities in these two projects are described by Young et al. (2016).

The sampling of soil in both Tellus and Tellus Border followed the field protocols of G-BASE as described in section 2.2. In the Tellus survey soil sampling was undertaken at a density of one sample per 2 km^2 and in Tellus Border at a density of one sample per 3.6 km^2. Samples were dried and disaggregated by hand, then sieved to pass through 2 mm. The sub-2 mm fraction was milled using an agate planetary ball mill to produce a sample of predominantly < 53 µm fraction. A 1-g sub-sample of the milled material was treated by two-acid (ratio of 2:1 HNO_3:HCl *aqua regia* variant) sample digestion, and the digestate was analysed for concentrations of a range of elements including As by multi-element ICP-MS analysis (Knights, 2013).

2.4 Case study 3: Pb in the Trent valley, East Midlands of England

This study area (Figure 1) comprises land primarily in rural environments around the River Trent in the East Midlands of England, although the urban centres of Worksop and Doncaster lie respectively on the southern and northern corners of its western edge. It was selected for examination for anomalies in soil lead concentration. The area does not include any known geogenic sources of lead, being some way to the east of the nearest mineralization domains for this element (Ander et al. 2013). It does not include a major urbanised domain for lead, as identified by Ander et al. (2013) but anthropogenic sources of lead from industrial activity, paint, leaded petrol and other sources are possible in the west due to the populated areas of Worksop and Doncaster and associated small towns. A priori, a likely source of lead is the alluvium of the River Trent. It is known that the Trent alluvium may have instances of large concentrations of lead (Izquierdo et al., 2014). Isotopic studies of this lead (Izquierdo et al., 2014) show that it has multiple sources including geogenic ones — lead mineralization in the Peak District of Derbyshire in the East Midlands of England upstream and to the west of the selected study area. This region includes deposits of lead which have been mined since the Roman occupation, and are also subject to natural weathering. Lead also occurs in East Midlands coal deposits. There is also evidence for anthropogenic lead, at least some of which can be attributed to lead tetraethyl which was used as an additive in petrol in the United Kingdom until 1999, although the lead content of petrol in the UK was reduced from 1986 (Noble et al., 2008). Geogenic sources of lead contribute to alluvium through weathering and transport by water. The anthropogenic sources of lead contribute through discharge of wastes directly into rivers and onto soils.

The soil data used for this case study were collected as part of the G-BASE programme following the same protocols as described for the Zn data in section 2.2, but with sampling at a density of approximately one sample per 2 km^2 with samples located as close as possible to the centres of alternate 1-km^2 cells of the British National Grid.

2.5 Case study 4: Hf in north Norfolk, eastern England

This study area is in the north of Norfolk in eastern England (Figure 1). The element considered here is hafnium (Hf). This element is known to be relatively enriched (along with zirconium) in aeolian deposits (Taylor et al., 1983), and this association has been confirmed for the loess deposits and coversands of eastern England by Scheib and Lee (2010). These aeolian deposits are of late Pleistocene origin when the southern limit of the British ice sheet was just to the north of the study area. Loess is found in the north-eastern part of the selected area and coversands in the south-west.

The soil data used in this case study were collected according to the same field and analytical protocols as the lead data described in section 2.4, with sampling at a mean density of one sample per 2 km^2.

3. Results

3.2 Case study 1: Zn east of Dartmoor, south-west England

On the original scale of measurement (mg kg^{-1}) the data on Zn concentration are somewhat skew (Figure 2, Table 1) but this is reduced by transformation to logarithms. Most of the larger concentrations are in the east of the region in soil formed over sandstones, mudstones, metasandstones, metamudstones and slates, the latter formed by metamorphosis under the influence of the igneous granite intrusion in the east of the region which underlies Dartmoor (British Geological Survey, 1995).

There are some differences among the empirical variograms obtained with different estimators (Figure 4). The histograms of the cross-validation errors suggest that these may be assumed to be normal (Figure 5) and the median SSPE (Table 2) for the variogram fitted to estimates obtained by the robust estimator of Cressie and Hawkins (1980) is closest to the expected value of 0.455, so this was used for the parametric bootstrapping of the singularity index for the background random variable (Table 3).

Table 4 shows that the two-component mixture model was favoured for the values of the singularity index for the Zn data. The dominant component has a standard deviation of 0.304, which is close to the median value of the standard deviation over the 1000 parametric bootstrap resamples (Table 3) and within the interquartile range. This supports the interpretation of the mixture model as showing, in the dominant component, the singularity index values for the background process, with additional mass in the tails of the distribution, corresponding to anomalies, introduced by a minor component with a larger standard deviation (Figure 6). The posterior likelihood of this second component is larger for values of the singularity index smaller than 1.51 and larger than 2.85, so these values were used as thresholds (Figure 7).

Most of the enrichment anomalies in Figure 7 appear in the east of the region. Note in particular that there are three clusters which lie on a line approximately north north-east to south south west, aligned with the distribution of known sites with stratified mineralization in the area. This suggests that the geochemical anomalies identified by the analysis are, at least in part, the result of local mineralization, some of which may be of economic significance.

3.3 Case study 2: As in Counties Monaghan and Armagh, north of Ireland

The original data on As were markedly skew (Figure 8, Table 1) and this was reduced by transformation to logarithms although the histograms suggest that a contaminated normal model may be appropriate with additional observations in an upper tail. There are marked differences between the variograms estimated by different methods (Figure 10). The cross validation errors in Figure 11 suggest that a normal model of these is plausible. The median SSPE for the variogram model fitted to estimates obtained with the estimator proposed by Dowd (1984) is 0.41, closer to 0.455 than that for any other model. This was used to obtain the parametric bootstrap samples of the statistics for the singularity index in Table 3.

Table 4 shows that a two-component mixture model was favoured for the singularity index. The dominant component had a standard deviation of 0.182. This was somewhat smaller than the median value of the parametric bootstrap resamples of the singularity index (Table 3) and outside the 95 percentile range (0.193–0.280). On the assumption that the dominant component represents the background process the thresholds for anomalies are 1.67 and 2.54 (Figure 13).

The arsenic enrichment anomalies evident in Figure 13 correspond to known arsenic (and gold) mineralisation identified through mineral exploration in the area. Both occur as coincident discrete packages along the line of the Orlock Bridge Fault, which is believed to be a controlling feature. The anomaly at location 270 500E (Irish National Grid, ING), 327 000N(ING) (County Monaghan) is located on the northern edge of the previously-identified gold anomaly at the Glenish gold target. The Clay Lake target at 282 000E(ING), 335 100N(ING) shows good correlation with the singularity anomaly in that location. However of particular interest, the Clontibret gold deposit at location 275 800E(ING), 330 000N(ING) shows no corresponding singularity anomaly in the soil dataset, based on the thresholds used. A minor anomaly is evident along strike between Glenish and Clay Lake which sits in the general gold trend identified. A further anomaly is located along trend to the south west at 257000E(ING), 311000N(ING) which may be related to the same mineralisation event. This anomaly at 257 000E(ING), 311 000N(ING) is located to the south of the mapped location of the Orlock Bridge Fault, rather than to the north as is the case at Glenish and Clay Lake.

In addition to the major NE - SW basement orientation, structural mapping in the area has identified NNE and NW trending lineaments and faults. Further anomalies in the area could be interpreted to follow the discrete package style of occurrence which typifies the mineralisation along the Orlock Bridge fault, in these other orientations.

3.4 Case study 3: Pb in the Trent valley, East Midlands of England

The original data on lead concentrations were very strongly skewed (Figure 14, Table 1) and this was reduced by log transformation although the coefficient of skewness for the transformed data still exceeds 1. There are marked differences between the variograms estimated by different methods (Figure 16) and the median SSPE for the variogram model fitted to estimates obtained with the estimator proposed by Dowd (1984) is 0.46, which is very close to the expectation for normal kriging errors with a correct variogram model. This mode was used to obtain the parametric bootstrap samples of the statistics for the singularity index in Table 3.

The two-component mixture model was favoured for the singularity index (Table 4, Figure 18). The standard deviation of the dominant component was 0.140, which is just outside the 95 percentile range for the bootstrap resampled estimate of this statistic under the lognormal model (0.11–0.13). Under the two component mixture model the minor component has a mean somewhat less than 2.0 (1.85) and so introduces most additional mass into the lower tail, corresponding to enrichment anomalies. The thresholds for anomalies under the mixture model were 1.81 and 2.52.

Figure 19 shows the singularity index and the thresholded values. It is clear that much of the region corresponding to enrichment anomalies is close to the course of the River Trent, shown by a white line in Figre 19b. There are also other patches with enrichment anomalies, notably to the west of the river. The largest such patch is near the British National Grid coordinates 470 000E, 381 000N. This corresponds to the small town of Retford in Nottinghamshire. While this town does not have a heritage of heavy industry it is a significant point of convergence for transport routes with two railway lines, significant trunk roads and a canal historically used for transport of industrial goods. The River Idle also passes through the town. This river, a tributary of the Trent, has a low water quality rating due to the urban setting of its tributaries and its reception of significant discharges from sewage treatment works (Environment Agency, 2006). In summary, much of the spatial distribution of enrichment anomalies is accounted for by the course of the Trent, and other enrichment anomalies are likely, as with the patch near Retford, to reflect anthropogenic factors.

3.5 Case study 4: Hf in north Norfolk, eastern England

The data on hafnium concentrations on the original scale are mildly skewed (Table 1, Figure 20) and have a marked negative skew on the logarithmic scale. For this reason a transformation to square roots was used. The larger Hf concentrations are seen in the north east and south west of the region (Figure 21) corresponding to known loess and cover sand deposits respectively. There are differences between the variogram estimates obtained with different estimators (Figure 22) and the cross validation supports the use of the model fitted to estimates obtained with the estimator proposed by Dowd (1984).

A two-component mixture model was favoured (Table 4). The dominant component has a standard deviation of 0.082 which is close to the median value for the parametric bootstrap resample estimates (0.087) and is within the 95 percentile range (0.076–0.101). The mixture model introduces very little additional mass from the second component, for which $\lambda = 0.03$. The threshold values for identification of anomalies according to the mixture model are 1.69 and 2.24. Very few values fall outside these limits, as can be seen in Figure 25. This suggests that, while there are areas with larger concentrations of Hf than others, and these can be explained from the known distribution of aeolian deposits, this spatial variation is consistent with the (trans) normal distribution and there is no evidence for substantial local enrichment or depletion inconsistent with this background variation.

4. Discussion and Conclusions

In this study we examined data on concentrations of four elements in the soil in four contrasting settings. In three of these cases there was clear evidence for geochemical anomalies, predominantly enrichment. In these cases the pattern of anomalies was consistent with independent knowledge about sources, geogenic or anthropogenic, of enrichment — the stratiform mineralizations near Dartmoor, the geogenic gold/arsenic mineral occurrences in the Longford-Down terrane, the Trent alluvium and anthropogenic sources of lead in the East Midlands of England. There was no substantial evidence of anomalies in the data on Hf in north Norfolk. While elevated values were indeed found as expected in areas with known aeolian deposits, these are consistent with a simple (trans)normal process, with no evidence of locally anomalous behaviour at the scales examined.

In all cases the likelihood ratio tests favoured a mixture of normals model with two but no more components. In all the minor component (smaller λ) had the larger standard deviation and so introduced additional mass into both tails of the overall distribution, although in most cases this was not symmetrical, and in the case of lead in the East Midlands of England the mean of the minor component was markedly smaller than 2.0, accounting for a marked lower tail in the overall distribution of the singularity index. The double-log plots of the empirical survival function of α for each case study (Figure 26) do not, in any of the cases, clearly comprise linear segments. This suggests that the mixture model used here is a more suitable way to determine threshold values of the singularity index.

In the case of Zn and Hf the standard deviation of the major component in the mixture model fell within the 95 percentile range of the parametric bootstrapped values extracted from values simulated with the selected variogram. In the case of Pb and As the standard deviation fell just outside this range, but was much closer to the bootstrapped values than was the standard deviation for the minor component. This supports the mixture interpretation of the singularity index values, and the thresholds derived from this. It is possible that the differences reflect limitation of the robust estimation of the variogram. It may also be that, while the assumption of normality for the transformed data seemed reasonable, as judged by the histograms, there may be structure in the spatial variation with non-normal higher-order moments — features which require multiple-point geostatistics (Meerschman et al., 2013).

It is notable that the two-component mixture model for the singularity index is compatible with a wide range of behaviours by the variable of interest. The mass introduced into the tails of the distribution may be very small in both tails if the standard deviation of the minor (non-background) component is not much larger than the standard deviation of the major component. That was seen in the case of hafnium in this study. With a larger standard deviation for the minor component, more mass can be introduced into both tails, modelling the presence of both enrichment and depletion anomalies. A strong preponderance of enrichment anomalies can be modelled if the mean for the minor component in the model is markedly smaller than two (as with lead). In a case where there was a combination of locally intensive enrichment (e.g. from industrial contamination) and more diffuse enrichment by processes such as wind dispersal, this might be represented by a minor component in the mixture model with a large standard deviation creating a heavy lower tail for the overall distribution of the singularity index, although in some cases more than two components might be required in the mixture model.

If the mean of the background process is not stationary (e.g. there is a spatial trend) then this should not markedly affect the corresponding component of the distribution model for the singularity index as this depends on the local variability in much the same way that ordinary kriging is robust to trends (Goovaerts, 1997). However, if the background process is not stationary in the variance and autocorrelation, then a more complex model would be necessary, and this is a topic for further study. Note that if there was non-stationarity in the variance then the median value of SSPE, examined in the cross-validation of the variogram models (Table 2) would be expected to deviate markedly from 0.45 (Lark, 2009).

To conclude, it has been shown that anomalous values (or their absence) can be identified in soil geochemical data by means of the singularity index, and that the interpretation of this index can be facilitated with a mixture model. More work is needed on the statistics of the index, particularly for the robust characterization of the background distribution under a null (trans)normal distribution, and for the modelling of non-stationary background processes.

5. Acknowledgements

MP and DMR were funded by the Geological Survey of Ireland under Short Call project ‘Geochemical anomaly detection: spatial analysis for improved use of geochemical data’, reference 2015-sc-069. We are grateful to Fiona McEvoy, Paul Lusty, Chris Yeomans and Gerry Stanley for helpful discussions. This paper is published with the permission of the Executive Director of the British Geological Survey (NERC).

References

Agterberg, F.P. 2012. Multifractals and geostatistics. Journal of Geochemical Exploration, 122, 113–122.

Allègre, C.J., Lewin, E. 1995. Scaling laws and geochemical distributions. Earth and Planetary Science Letters. 132, 1–13.

Ander, E.L., Johnson, C.C., Cave, M.R., Palumbo-Roe, B., Nathanail, C.P., Lark, R.M. 2013. Methodology for the determination of normal background concentrations of contaminants in English soil. Science of the Total Environment, 454–455, 604–618.

Benaglia, T., Chauveau, D., Hunter, D.R., Young, D. 2009. mixtools: An R Package for Analyzing Finite Mixture Models. Journal of Statistical Software, 32, 1–29.

Benham, A.J., McEvoy, F.M., Rollin, K.E. 2005 Potential for stratiform massive sulphide mineralisation in south-west England. Transactions of the Institution of Mining and Metallurgy, Section B Applied Earth Science. 113, 227–246.

Breward, N., 2007. Arsenic and presumed resistate trace element geochemistry of the Lincolnshire (UK) sedimentary ironstones and revealed by a regional geochemical survey using soil, water and stream sediment sampling. Applied Geochemistry, 22, 1970–1993.

British Geological Survey. 1995. Dartmoor Forest. Solid and Drift Map. 1:63,360/1:50,000 geological map series, New Series Sheet 338. British Geological Survey, Keyworth, Nottingham.

Chen, Z., Cheng, Q., Chen, J., Xie, S. 2007. A novel iterative approach for mapping local singularities from geochemical data. Nonlinear Processes in Geophysics, 14, 317–324.

Cheng, Q. 2007. Mapping singularities with stream sediment geochemical data for prediction of undiscovered mineral deposits in Gejiu, Yunnan Province, China. Ore

Geology Reviews, 32, 314–324.

Cheng, Q. 2012. Singularity theory and methods for mapping geochemical anomalies caused by buried sources and for predicting undiscovered mineral deposits in covered areas. Journal of Geochemical Exploration, 122, 55–70.

Cressie, N., Hawkins, D. 1980. Robust estimation of the variogram. Journal of the International Association of Mathematical Geology 12, 115–125.

Dowd, P.A. 1984. The variogram and kriging: robust and resistant estimators. In: Geostatistics for Natural Resources Characterization (eds G. Verly, M. David, A.G. Journel and A. Marechal), Part 1. pp. 91–106. D. Reidel, Dordrecht.

Environment Agency 2006. The Idle and Torne Catchment Abstraction Management Strategy. Environment Agency, Bristol.

Genton, M.G. 1998. Highly robust variogram estimation. Mathematical Geology 30, 213–221.

Goovaerts, P. 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press, New York.

Izquierdo, M., Tye, A.M., Chenery, S.R. 2012. Sources, lability and solubility of Pb in alluvial soils of the River Trent catchment, U.K. Science of the Total Environment, 433, 110–122.

Johnson, C.C., Breward, N., Ander, E. L., Ault, L. 2005. G-BASE: Baseline geochemical mapping of Great Britain and Northern Ireland. Geochemistry: Exploration, Environment, Analysis 5, 1–13.

Knights, K.V. 2013. Quality control and statistical summaries of Tellus Border topsoil regional geochemical data. Report Version 1.0. Geological Survey of Ireland and Geological Survey of Northern Ireland joint report.

Lark, R.M. 2000. A comparison of some robust estimators of the variogram for use in soil survey. European Journal of Soil Science 51, 137–157.

Lark, R.M. 2009. Kriging a soil variable with a simple non-stationary variance model. Journal of Agricultural Biological and Environmental Statistics, 14, 301–321.

Lark, R.M., Ander, E.L., Cave, M.R., Knights, K.V., Glennon, M.M, Scanlon, R.P. 2014. Mapping trace element deficiency by cokriging from regional geochemical soil data: a case study on cobalt for grazing sheep in Ireland. Geoderma, 226–227, 64–78.

Lin, Y., Cheng, B., Shyu, G., Chang, T. 2010. Combining a finite mixture distribution model with indicator kriging to delineate and map the spatial patterns of soil heavy metal pollution in Chunghua County, central Taiwan. Environmental Pollution, 158, 235–244.

Liu, Y., Cheng, Q., Xia, Q., Wang, X. 2014. Identification of REE mineralization-related geochemical anomalies using fractal/multifractal methods in the Nanling belt, South China. Environmental Earth Sciences, 72, 5159–5169.

Matheron, G. 1962. Traité de Géostatistique Appliqué, Tome 1. Mémoires du Bureau de Recherches Géologiques et Minières, Paris.

Meerschman, E., Van Meirvenne, M., Van De Vijver, E., De Smedt, T.,Islam, M.M., Saey, T. 2013. Mapping complex soil patterns with multiple-point geostatistics. European Journal of Soil Science, 64, 183–191.

Noble, S.R., Horstwood, M.S.A., Davy, P., Pashley, V., Sprio, B., Smith, S. 2008. Evolving Pb isotope signatures of London airborne particulate matter (PM_{10}) — constraints from on-filter and solution-mode MC-ICP-MS. Journal of Environmental Monitoring, 10, 830–836.

Papritz, A. 2016. georob: Robust Geostatistical Analysis of Spatial Data. R package version 0.2-3. http://CRAN.R-project.org/package=georob

R Core Team 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rawlins, B.G., Lark, R.M., Webster, R., O'Donnell, K.E. 2006. The use of soil survey data to determine the magnitude and extent of historic metal deposition related to atmospheric smelter emissions across Humberside, UK. Environmental Pollution, 143, 416–426.

Scheib, A.J., Lee, J.R. 2010. The application of regional-scale geochemical data in defining the extent of aeolian sediments : the Late Pleistocene loess and coversand deposits of East Anglia, UK. Quaternary newsletter, 120, 5–14.

Taylor, S.R., McLennan, S.M., McCulloch, M.T. 1983. Geochemistry of loess, continental crustal composition and crustal model ages. Geochimica et Cosmochimica Acta, 47, 1897–1905.

Venables, W. N., Ripley, B. D. 2002. Modern Applied Statistics with S. Fourth Edition. Springer, New York.

Young, M.E., Knights, K.V., Smyth, D., Glennon, M.M., Scanlon, R.P., Gallagher, V. 2016. The Tellus geochemical surveys, results and applications. In Unearthed, impacts of the Tellus surveys of the north of Ireland. (ed. M.E. Young), pp 33–52. Royal Irish Academy, Dublin.

Table 1: Summary statistics for soil variables on original and transformed scale

Variable	Mean	Median	Standard deviation	Skewness	Quartile 1	Quartile 3
Zn mg kg^{-1}	96.64	78.05	77.70	3.35	45.42	126.93
Zn $\log\left(\mathrm{mg\ kg}^{-1}\right)$	4.34	4.36	0.68	0.07	3.82	4.84
As mg kg^{-1}	8.00	5.76	9.48	6.58	3.91	8.53
As $\log\left(\mathrm{mg\ kg}^{-1}\right)$	1.79	1.75	0.71	0.42	1.36	2.14
Pb mg kg^{-1}	49.52	35	62.98	11.16	28	49
Pb $\log\left(\mathrm{mg\ kg}^{-1}\right)$	3.68	3.56	0.56	1.44	3.33	3.89
Hf $\left(\mathrm{mg\ kg}^{-1}\right)$	7.59	7.3	2.37	1.40	6.1	8.6
Hf $\left(\mathrm{mg\ kg}^{-1}\right)^{0.5}$	2.72	2.70	0.41	0.46	2.47	2.93

Table 2: Mean and median standardized squared prediction error for cross-validation on each variable

Variable	Standardized squared prediction error							
	Variogram estimator							
	Matheron		Cressie–Hawkins		Dowd		Genton	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Zn $\log\left(\text{mg kg}^{-1}\right)$	0.90	0.32	1.27	0.43	1.71	0.51	1.17	0.39
As $\log\left(\text{mg kg}^{-1}\right)$	0.86	0.25	1.09	0.33	1.35	0.41	1.13	0.34
Pb $\log\left(\text{mg kg}^{-1}\right)$	1.02	0.23	1.58	0.34	2.15	0.46	1.71	0.37
Hf $\left(\text{mg kg}^{-1}\right)^{0.5}$	0.96	0.29	1.19	0.37	1.35	0.43	4.25	1.31

Table 3: Mean and quantiles of parameters of the parametric-bootstrapped singularity index under the (log)normal null model

	Mean of α					Standard deviation of α				
Quantile	0.025	0.25	0.50	0.75	0.975	0.025	0.25	0.50	0.75	0.975
Variable										
Zn $\log\left(\mathrm{mg\ kg}^{-1}\right)$	2.02	2.05	2.08	2.10	2.17	0.235	0.272	0.298	0.331	0.413
As $\log\left(\mathrm{mg\ kg}^{-1}\right)$	2.03	2.05	2.06	2.07	2.09	0.193	0.216	0.228	0.243	0.280
Pb $\log\left(\mathrm{mg\ kg}^{-1}\right)$	2.010	2.015	2.018	2.022	2.030	0.109	0.117	0.121	0.125	0.134
Hf $\left(\mathrm{mg\ kg}^{-1}\right)^{0.5}$	2.002	2.007	2.009	2.011	2.016	0.076	0.083	0.087	0.0912	0.101

Table 4: Mixture fitting for singularity index

	Log-likelihood ratio and P-values				Mixture-model parameters					
	2 components		3 components		Component 1			Component 2		
	L	P	L	P	λ	μ	σ	λ	μ	σ
Variable										
Zn $\log\left(\text{mg kg}^{-1}\right)$	17.56	0.010	9.65	0.1	0.73	2.12	0.304	0.27	2.03	0.487
As $\log\left(\text{mg kg}^{-1}\right)$	39.59	<0.001	5.05	0.28	0.81	2.08	0.182	0.19	2.00	0.364
Pb $\log\left(\text{mg kg}^{-1}\right)$	104.05	<0.001	6.78	0.16	0.81	2.10	0.140	0.19	1.85	0.319
Hf $\left(\text{mg kg}^{-1}\right)^{0.5}$	13.38	0.04	4.68	0.30	0.97	2.01	0.082	0.03	2.11	0.157

Figure captions

Figure 1. Map of the United Kingdom and Ireland showing the locations of the rectangular regions within which the singularity index was evaluated for some element in the topsoil. a). East of Dartmoor in the south-west of England (Zn); b). Longford-Down terrane in Counties Monaghan and Armagh in the north of Ireland (As); c). Part of the Trent valley in the East Midlands of England (Pb) d). Part of North Norfolk in eastern England (Hf).

Figure 2. Boxplots and histograms for (2a and 2b) soil Zn content from the south-west study area and (2c and 2d) soil Zn content transformed to natural logarithms.

Figure 3. Post-plot of the Zn data, the symbols distinguish sample points in the quartiles of the distribution. Coordinates are in metres relative to the origin of the British National Grid.

Figure 4. Empirical variograms for Zn content from the south-west study area with fitted models. Estimators are Matheron (4a), Cressie-Hawkins (4b), Dowd (4c) and Genton (4d).

Figure 5. Histograms of cross-validation errors for Zn content from the south-west study area. Estimators are Matheron (5a), Cressie-Hawkins (5b), Dowd (5c) and Genton (5d).

Figure 6. Histogram of the singularity index for Zn content from the south-west study area with PDFs for two components of the mixture superimposed.

Figure 7. (Top) values of the singularity index for Zn and (bottom) values thresholded on the basis of the mixture model: enrichment anomaly in black, background in grey, depletion anomaly hachured. The dark grey symbols indicate the locations of known stratiform mineral deposits in the area, the grey star indicating a deposit where Zn is a major ore element. Coordinates are in metres relative to the origin of the British National Grid.

Figure 8. Boxplots and histograms for (8a and 8b) soil As content from the Longford-Down terrane and (8c and 8d) soil As content transformed to natural logarithms.

Figure 9. Post-plot of the As data, the symbols distinguish sample points in the quartiles of the distribution. Coordinates are in metres relative to the origin of the Irish National Grid.

Figure 10. Empirical variograms for soil As content from the Longford-Down terrane with fitted models. Estimators are Matheron (10a), Cressie-Hawkins (10b), Dowd (10c) and Genton (10d).

Figure 11. Histograms of cross-validation errors for soil As content from the Longford-Down study area. Estimators are Matheron (11a), Cressie-Hawkins (11b), Dowd (11c) and Genton (11d).

Figure 12. Histogram of the singularity index for soil As content from the Longford-Down study area with PDFs for two components of the mixture superimposed.

Figure 13. (Top) values of the singularity index for As and (bottom) values thresholded on the basis of the mixture model: enrichment anomaly in black, background in grey, depletion anomaly hachured. Coordinates are in metres relative to the origin of the Irish National Grid.

Figure 14. Boxplots and histograms for (14a and 14b) soil Pb content from the East Midlands study area and (14c and 14d) soil Pb content transformed to natural logarithms.

Figure 15. Post-plot of the Pb data, the symbols distinguish sample points in the quartiles of the distribution. Coordinates are in metres relative to the origin of the British National Grid.

Figure 16. Empirical variograms for soil Pb content from the East Midlands study area with fitted models. Estimators are Matheron (16a), Cressie-Hawkins 16b), Dowd

(16c) and Genton (16d).

Figure 17. Histograms of cross-validation errors for soil Pb content from the East Midlands study area. Estimators are Matheron (17a), Cressie-Hawkins (17b), Dowd (17c) and Genton (17d).

Figure 18. Histogram of the singularity index for soil Pb content from the East Midlands study area with PDFs for two components of the mixture superimposed.

Figure 19. (Top) values of the singularity index for Pb and (bottom) values thresholded on the basis of the mixture model: enrichment anomaly in black, background in grey, depletion anomaly hachured. The white line shows the course of the River Trent. Coordinates are in metres relative to the origin of the British National Grid.

Figure 20. Boxplots and histograms for (20a and 20b) soil Hf content from the North Norfolk study area and (20c and 20d) soil Hf content transformed to natural logarithms.

Figure 21. Post-plot of the Hf data, the symbols distinguish sample points in the quartiles of the distribution. Coordinates are in metres relative to the origin of the British National Grid.

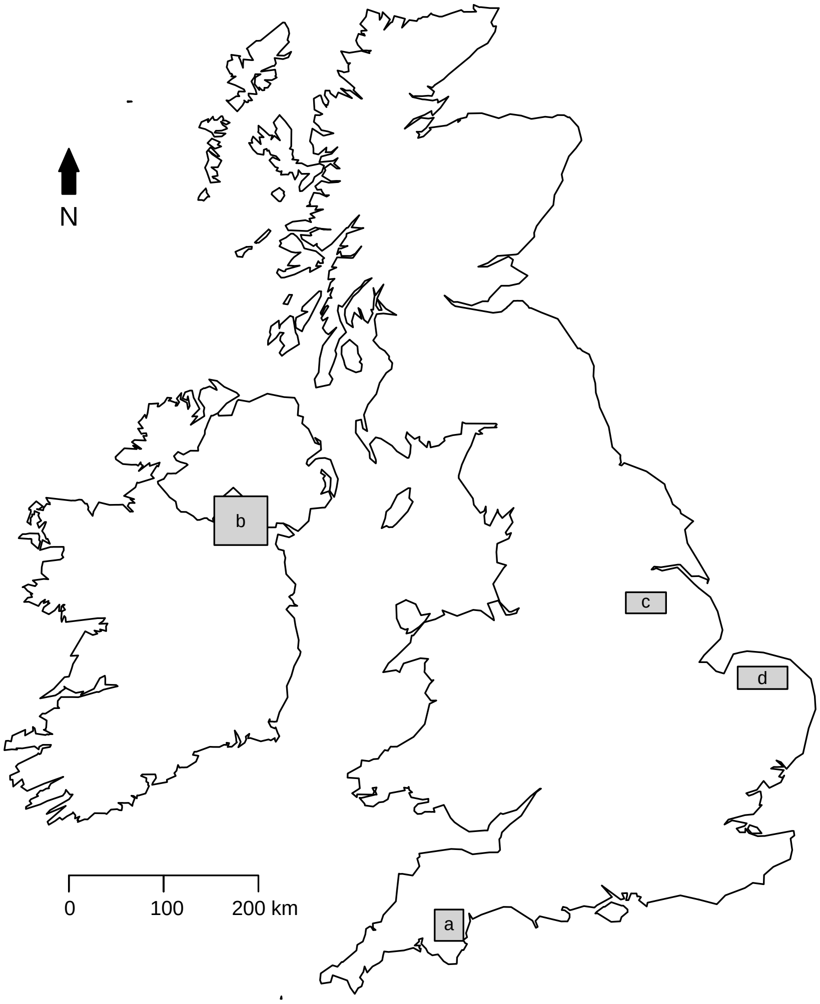
Figure 22. Empirical variograms for soil Hf content from the North Norfolk study area with fitted models. Estimators are Matheron (22a), Cressie-Hawkins (22b), Dowd (22c) and Genton (22d).

Figure 23. Histograms of cross-validation errors for soil Hf content from the North Norfolk study area. Estimators are Matheron (23a), Cressie-Hawkins (23b), Dowd (23c) and Genton (23d).

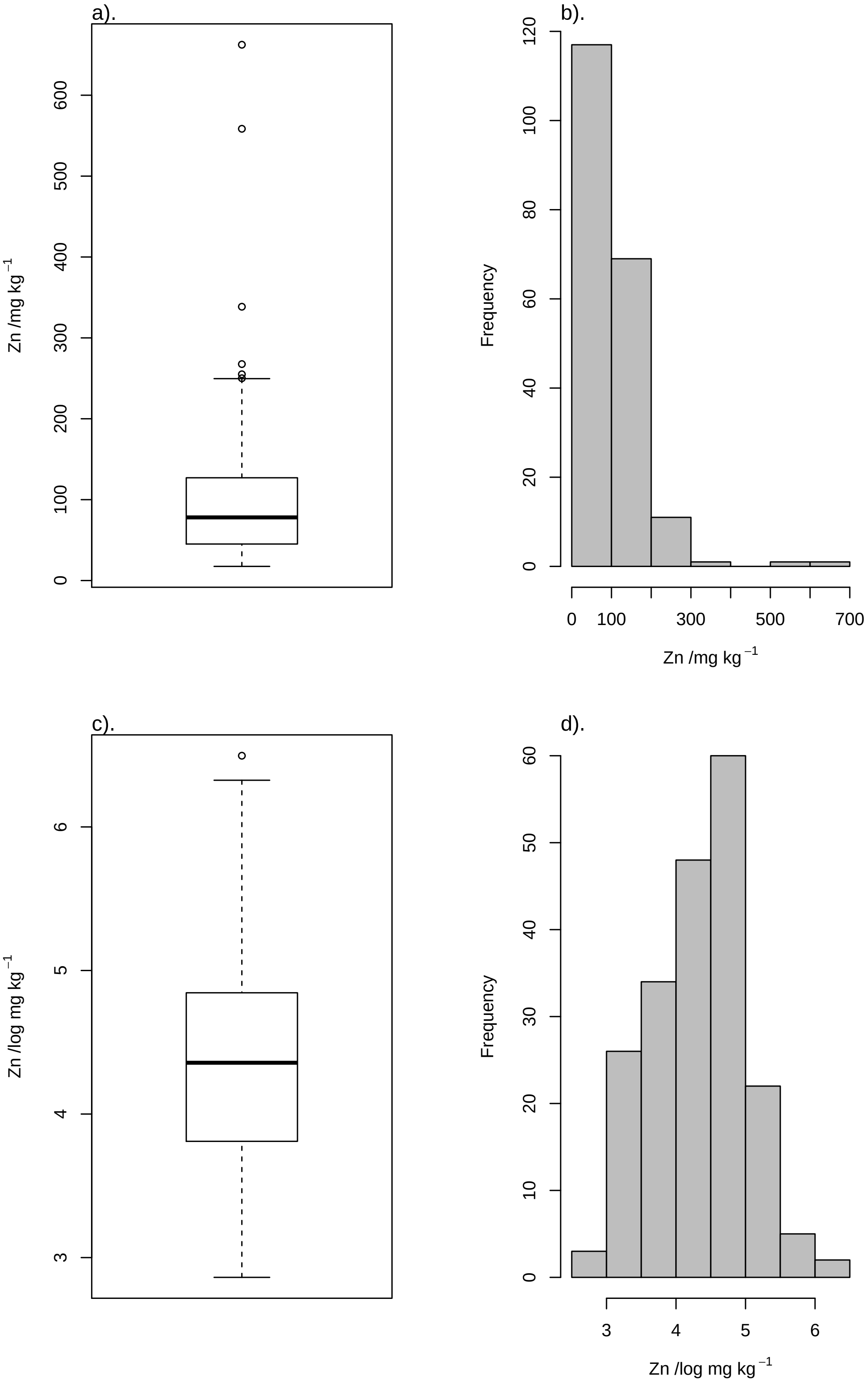
Figure 24. Histogram of the singularity index for soil Hf content from the North Norfolk study area with PDFs for two components of the mixture superimposed.

Figure 25. (Top) values of the singularity index for Hf and (bottom) values thresholded on the basis of the mixture model: enrichment anomaly in black, background in grey, depletion anomaly hachured. Coordinates are in metres relative to the origin of the British National Grid.

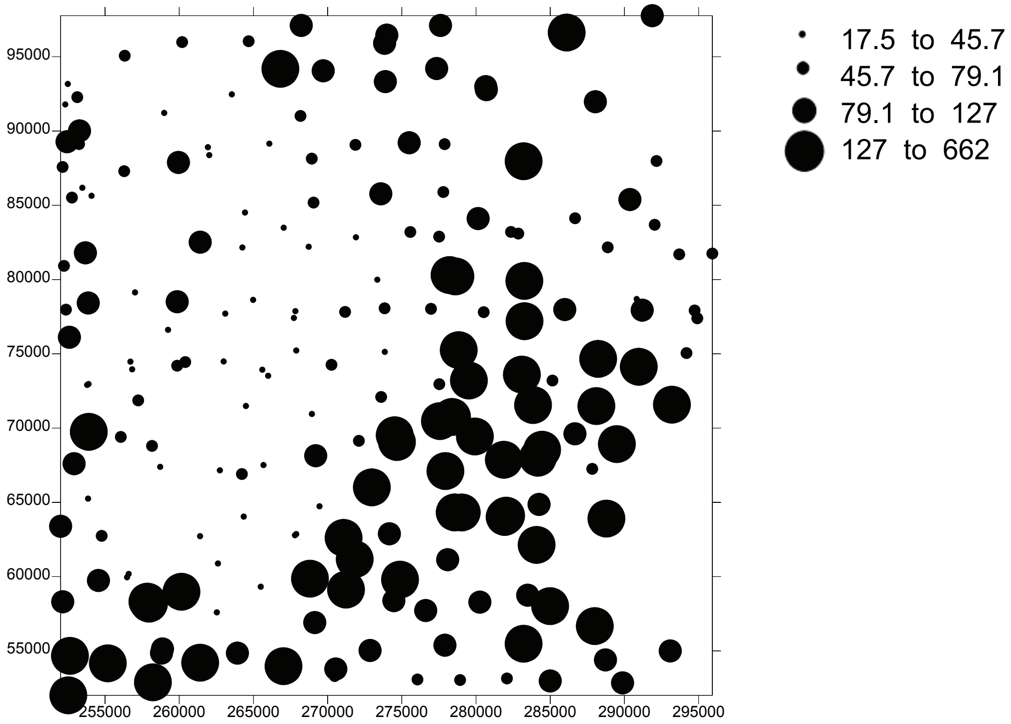
Figure 26. Plots of the empirical survival function of the singularity index, i.e. the area corresponding to values of the index less than α on double-log axes for a) Zn in the south-west study area; b) As at Longford-Down; c) Pb in the East Midlands and d) Hf in North Norfolk.



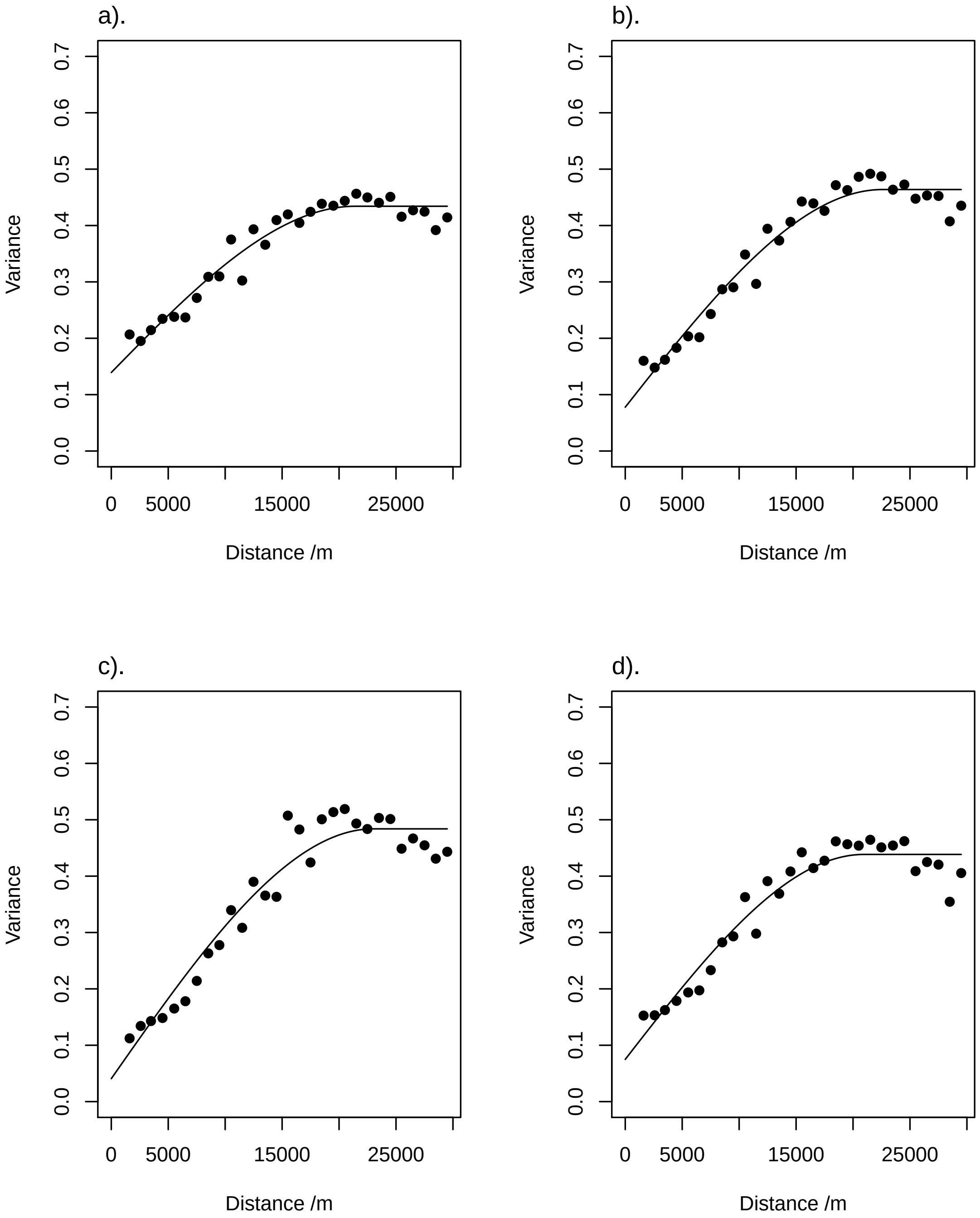
1: Fig 1


a).
Zn /mg kg^{-1}
0
100
200
300
400
500
600
b).
Frequency
0
20
40
60
80
100
120
0
100
300
500
700
Zn /mg kg^{-1}
c).
Zn /log mg kg^{-1}
3
4
5
6
d).
Frequency
0
10
20
30
40
50
60
3
4
5
6
Zn /log mg kg^{-1}

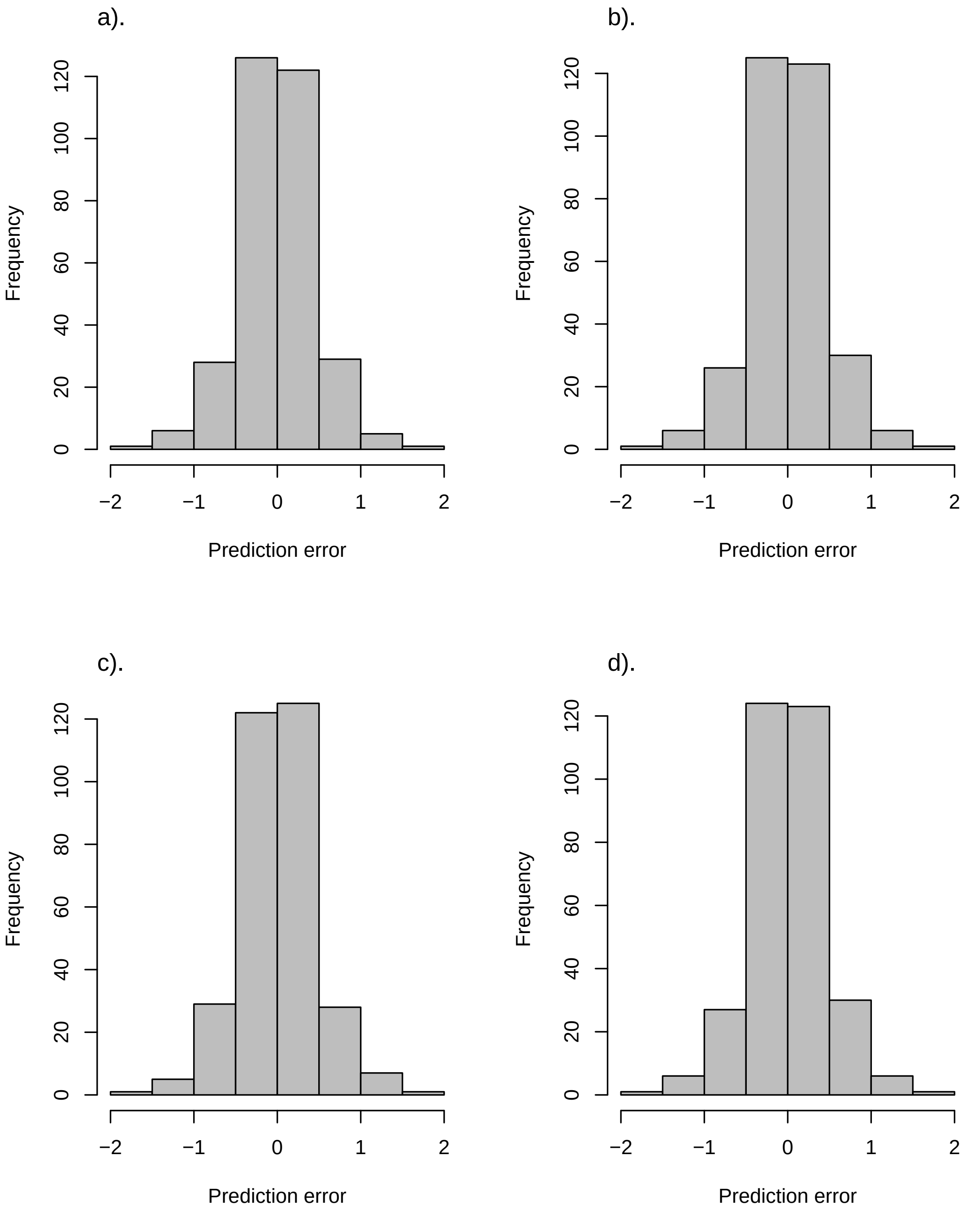
2: Fig 1


95000
90000
85000
80000
75000
70000
65000
60000
55000
255000
260000
265000
270000
275000
280000
285000
290000
295000
17.5 to 45.7
45.7 to 79.1
79.1 to 127
127 to 662

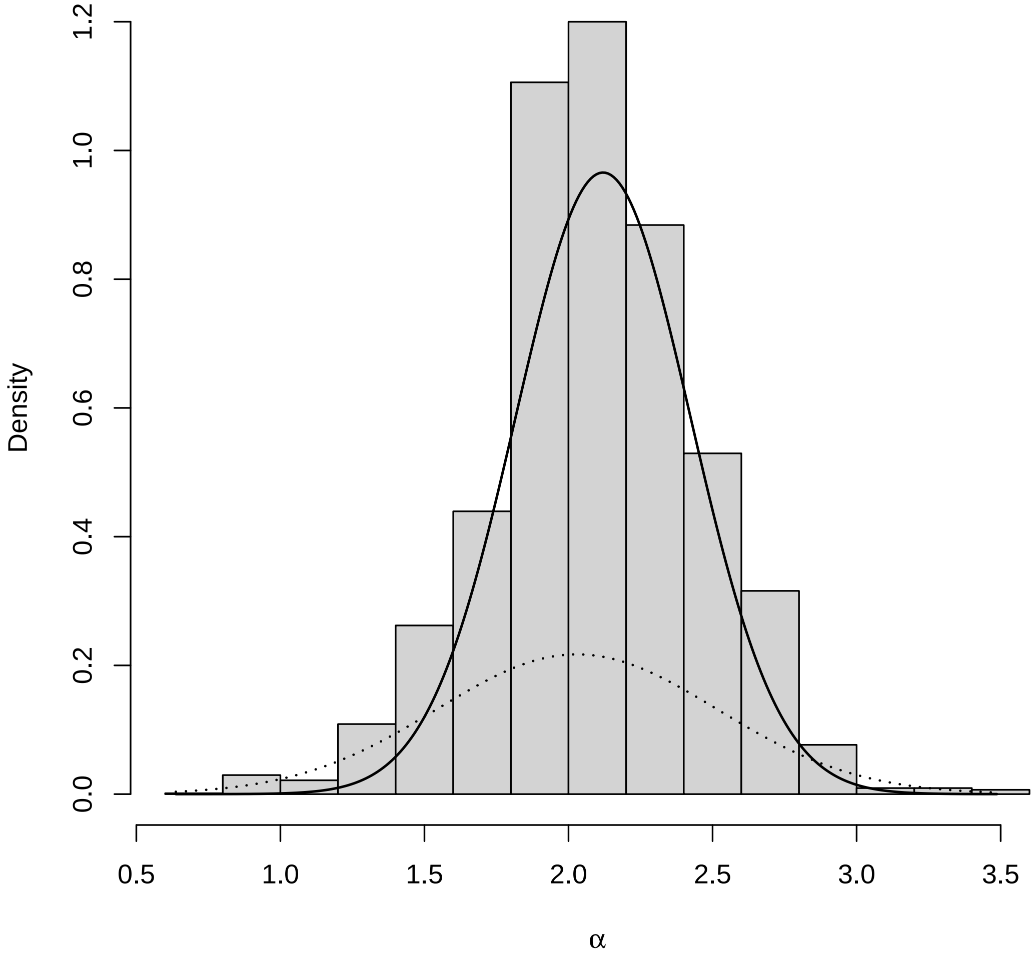
3: Fig 3


a).
b).
c).
d).
Variance
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
5000
15000
25000
Distance /m

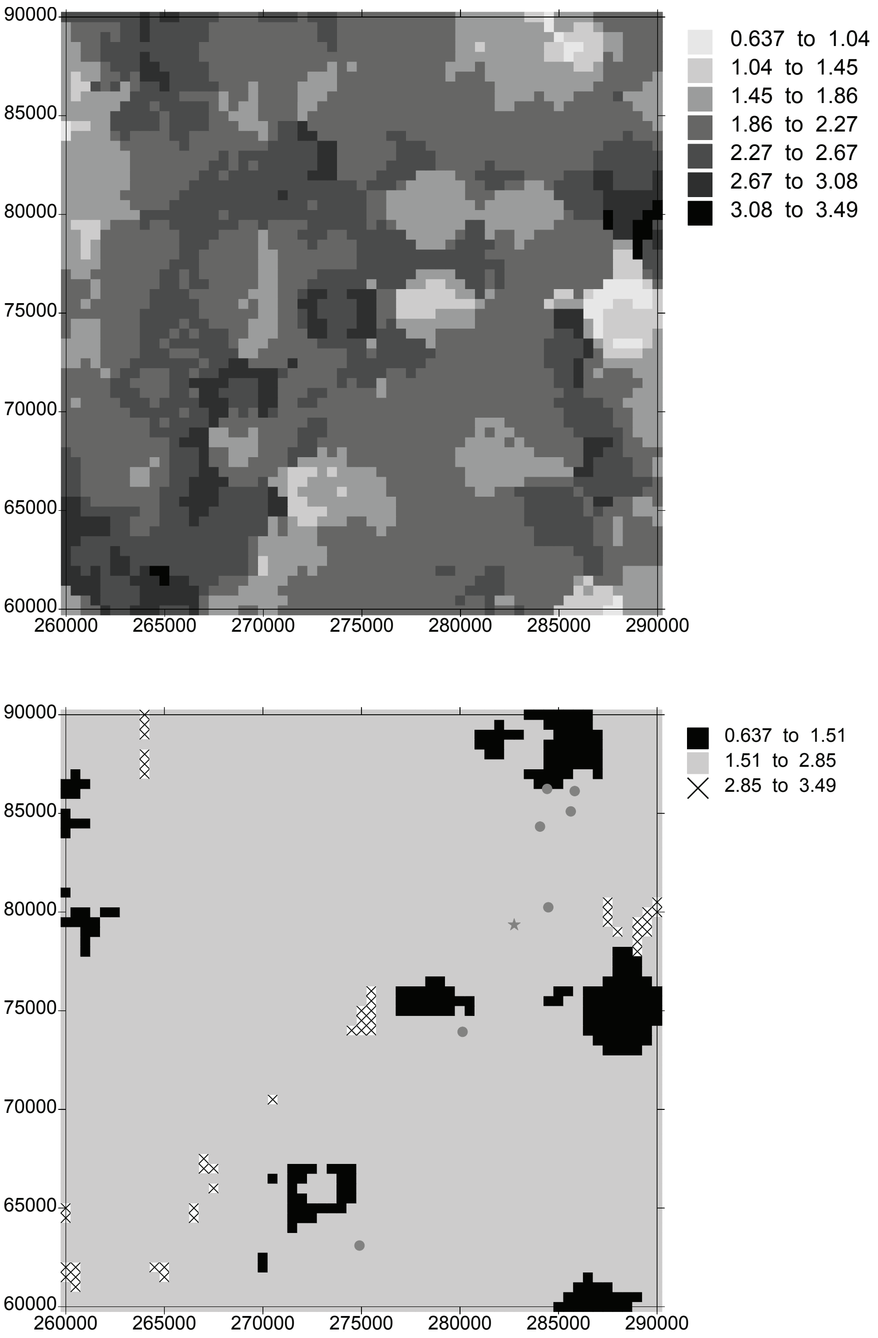
4: Fig 4



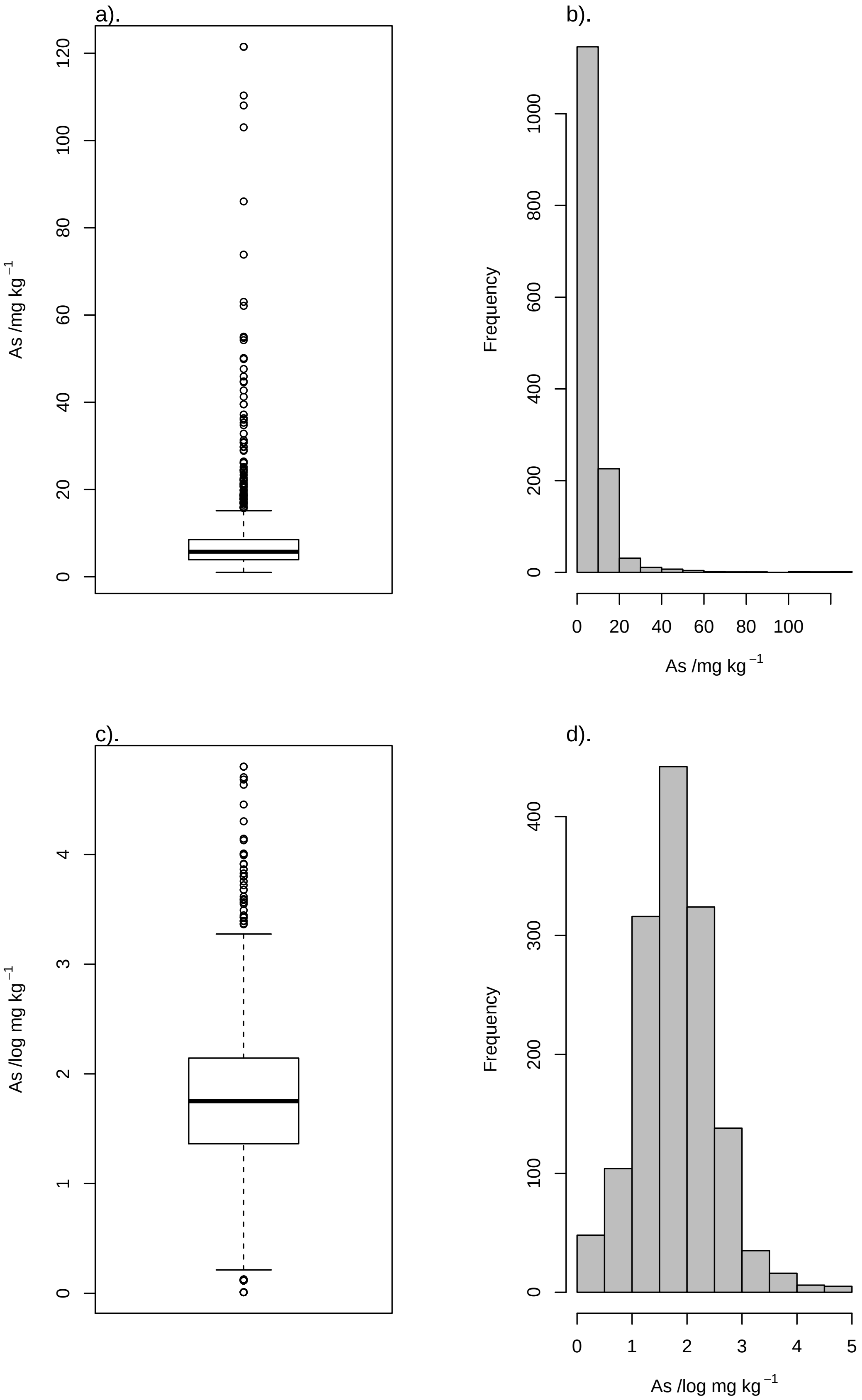
5: Fig 5



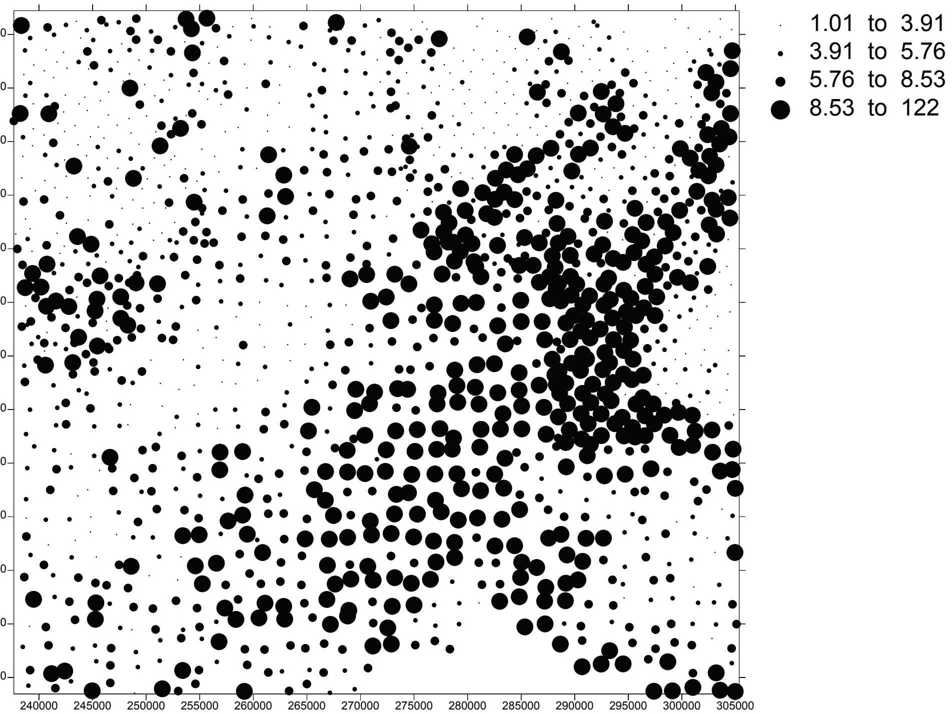
6: Fig 6


90000
85000
80000
75000
70000
65000
60000
260000
265000
270000
275000
280000
285000
290000
0.637 to 1.04
1.04 to 1.45
1.45 to 1.86
1.86 to 2.27
2.27 to 2.67
2.67 to 3.08
3.08 to 3.49
90000
85000
80000
75000
70000
65000
60000
260000
265000
270000
275000
280000
285000
290000
0.637 to 1.51
1.51 to 2.85
2.85 to 3.49

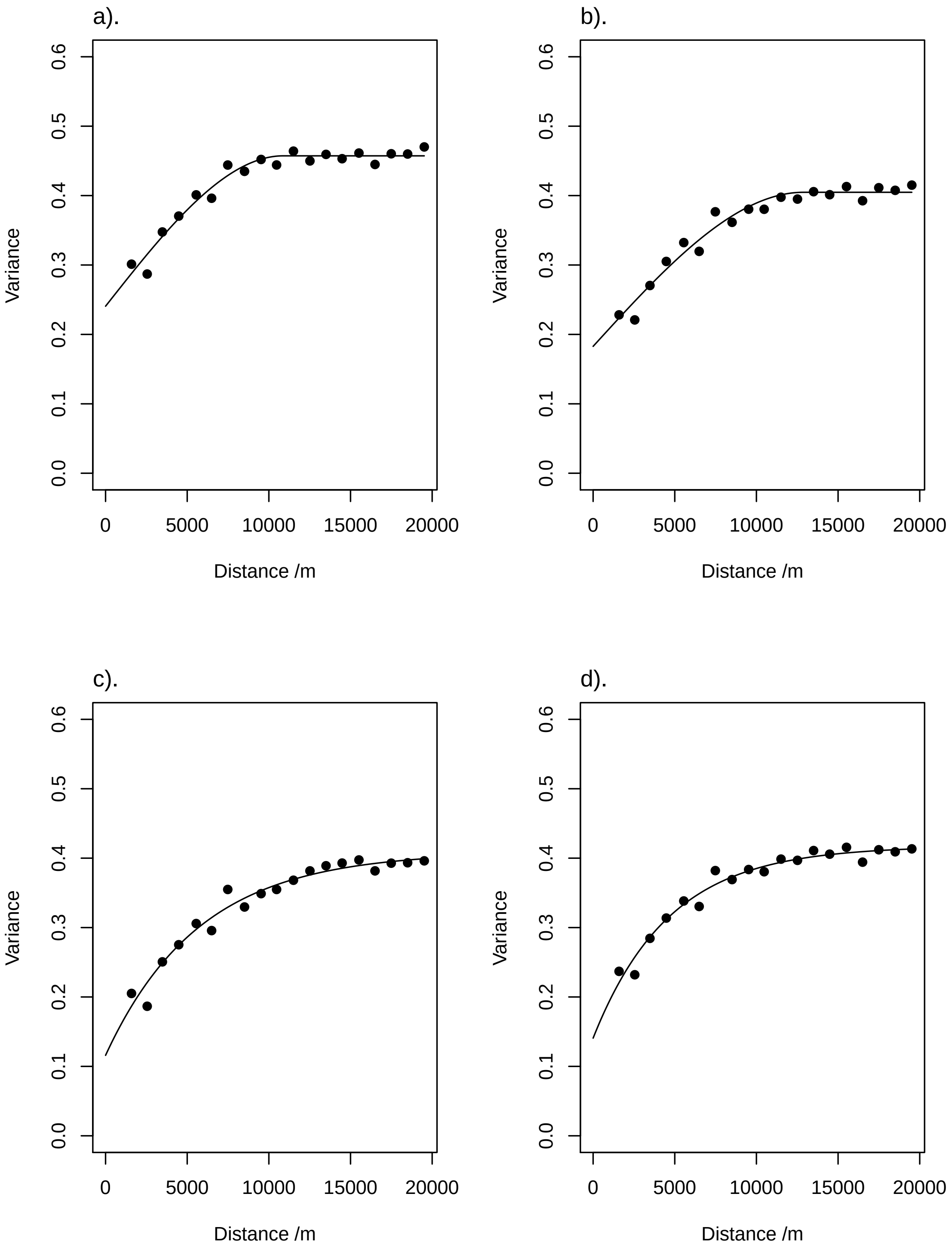
7: Fig 7


a).
As /mg kg^{-1}
0
20
40
60
80
100
120
b).
Frequency
0
200
400
600
800
1000
0
20
40
60
80
100
As /mg kg^{-1}
c).
As /log mg kg^{-1}
0
1
2
3
4
d).
Frequency
0
100
200
300
400
0
1
2
3
4
5
As /log mg kg^{-1}

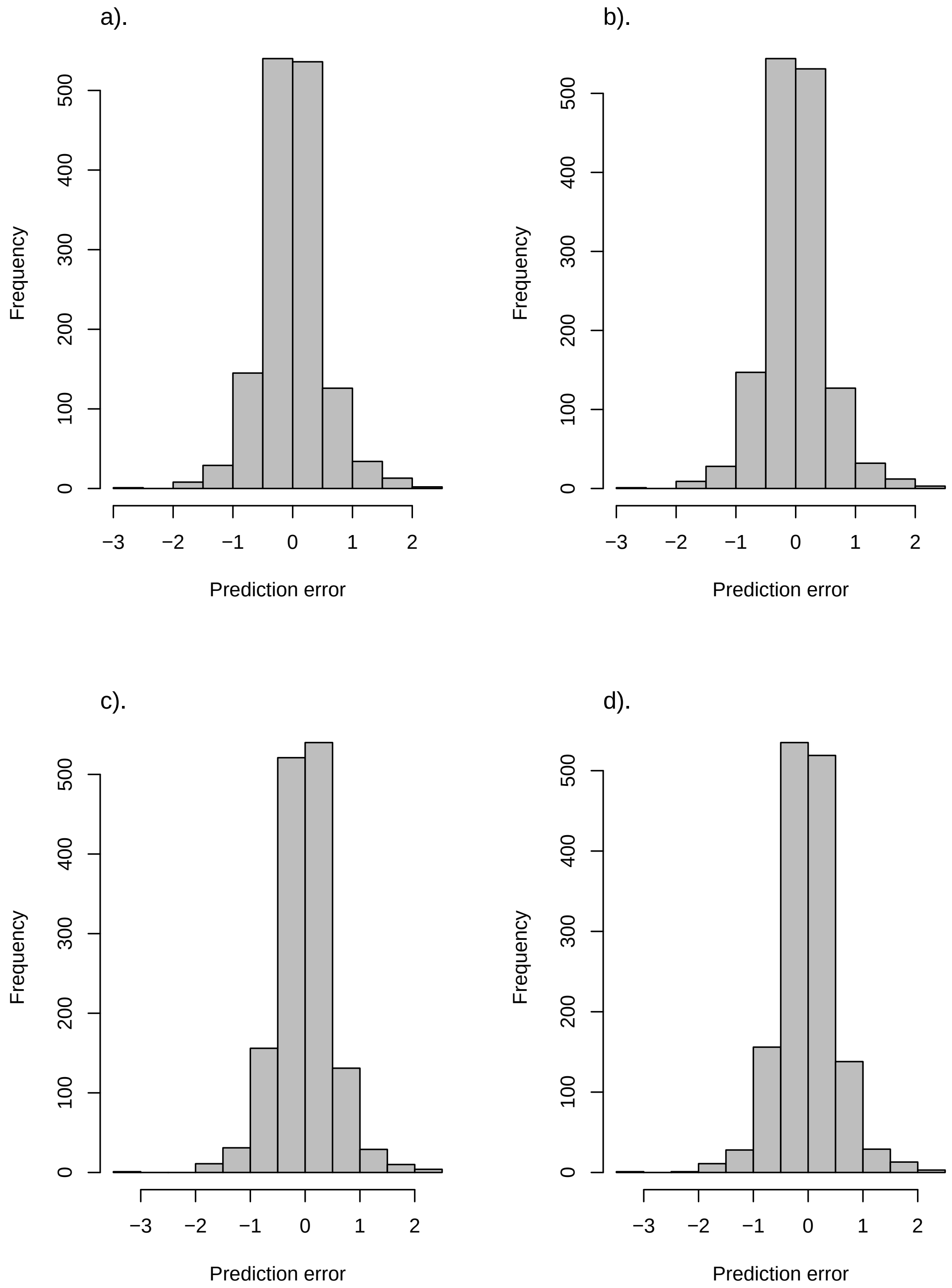
8: Fig 8


350000
345000
340000
335000
330000
325000
320000
315000
310000
305000
300000
295000
290000
240000
245000
250000
255000
260000
265000
270000
275000
280000
285000
290000
295000
300000
305000
1.01 to 3.91
3.91 to 5.76
5.76 to 8.53
8.53 to 122

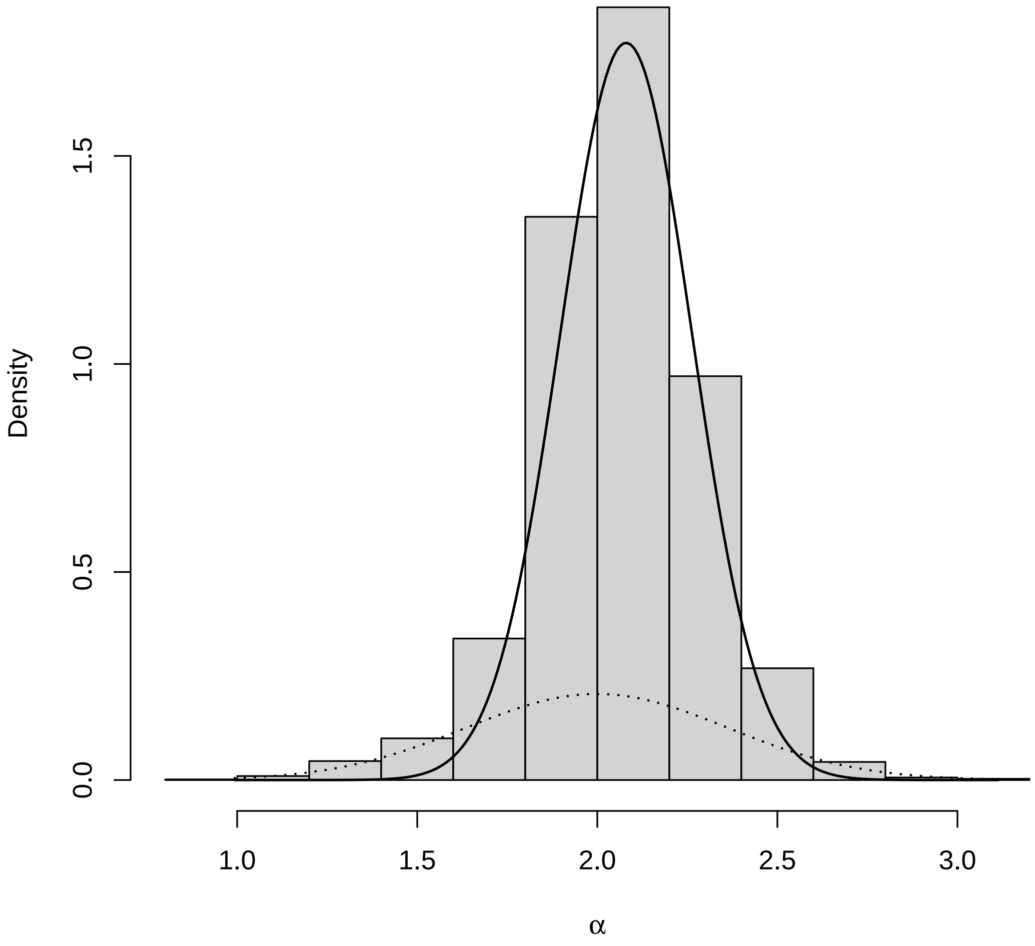
9: Fig 9


a).
b).
c).
d).
Variance
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
5000
10000
15000
20000
Distance /m

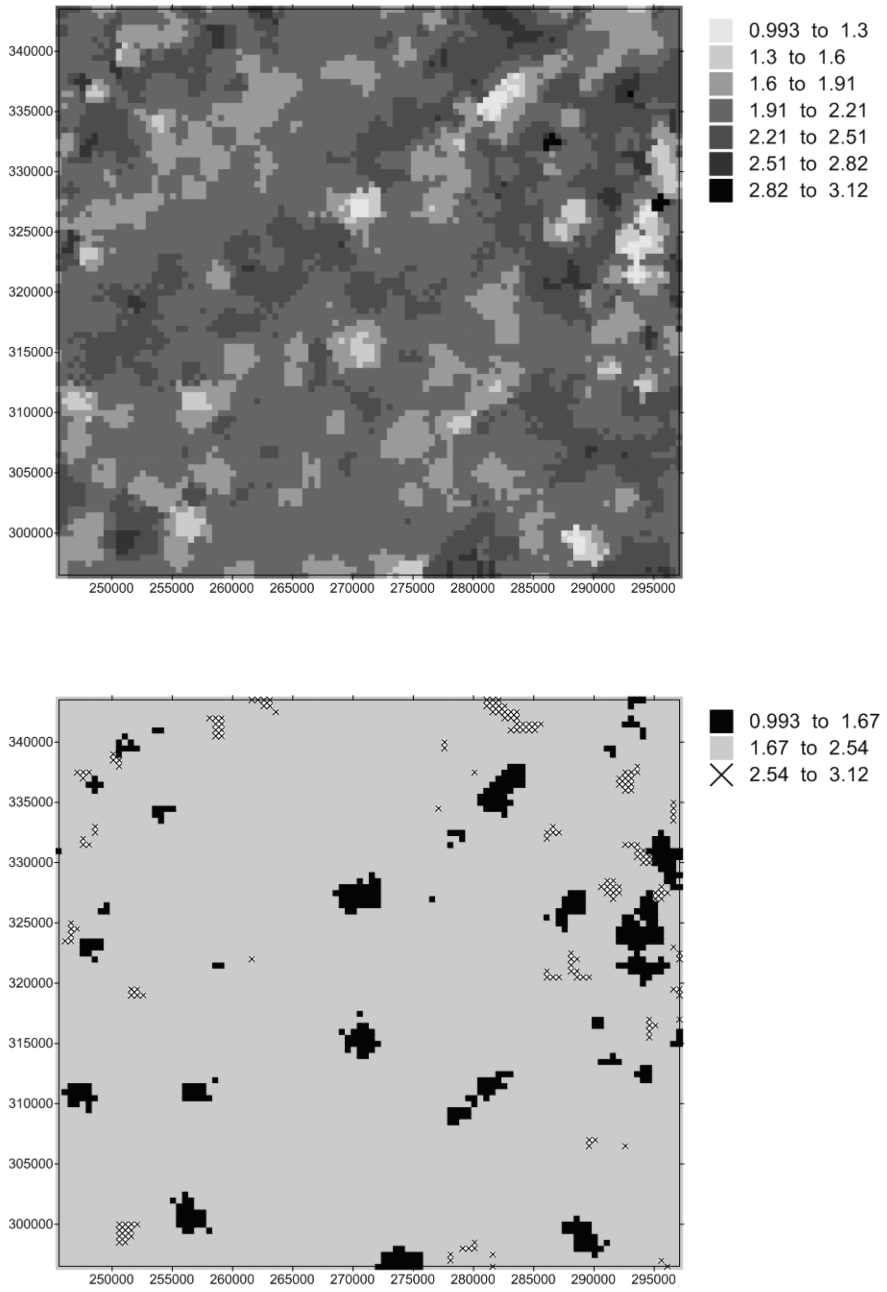
10: Fig 10


a).
b).
c).
d).
Frequency
0
100
200
300
400
500
−3
−2
−1
0
1
2
Prediction error

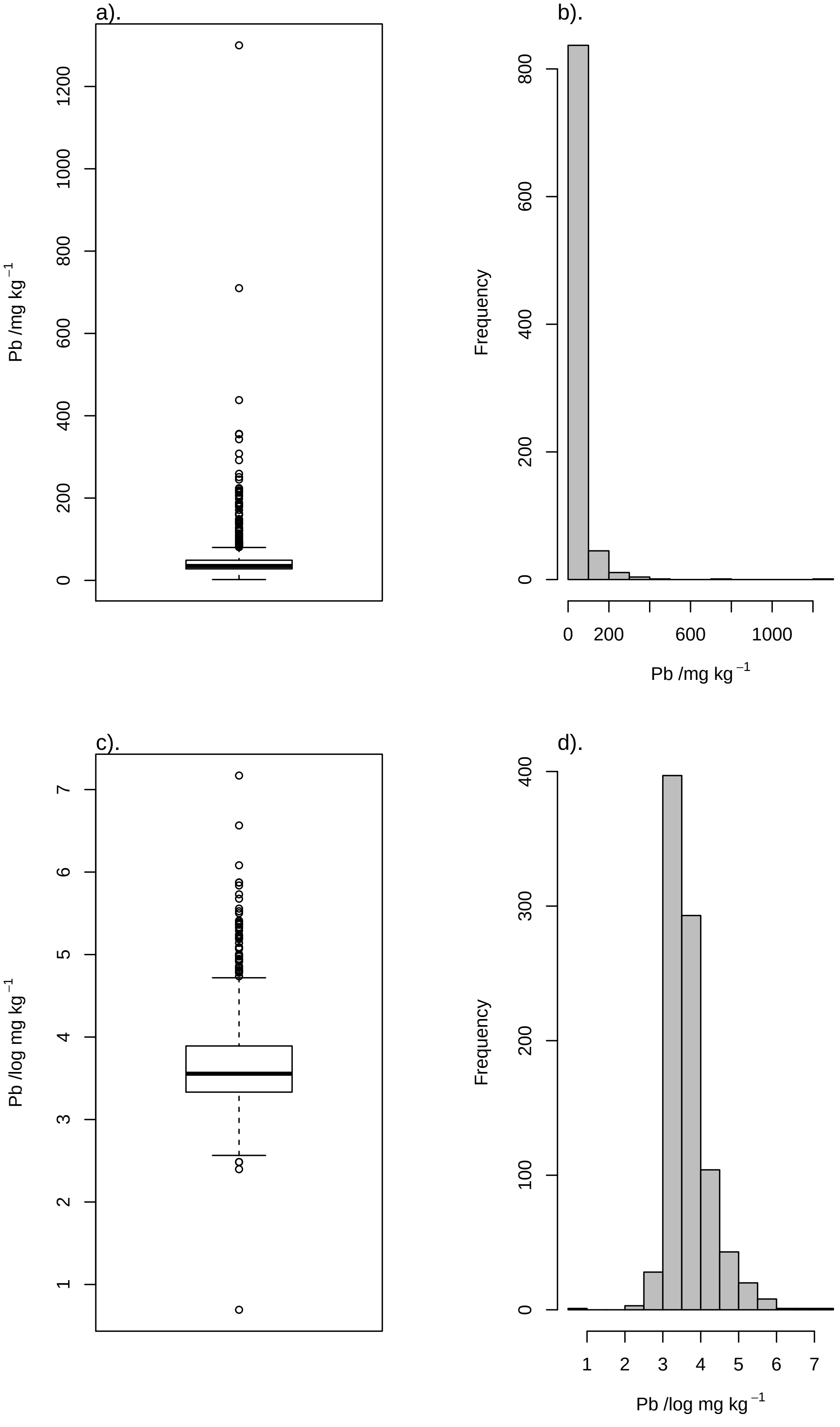
11: Fig11



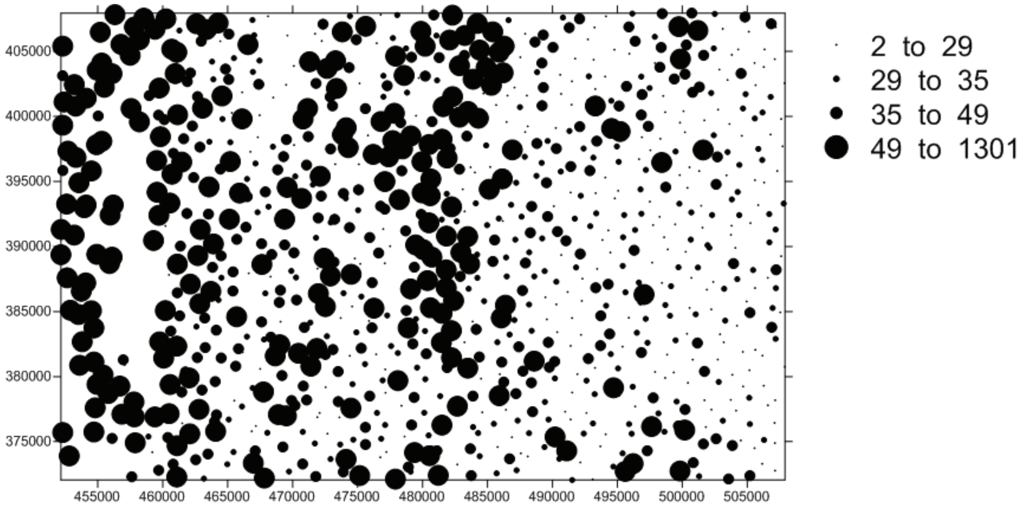
12: Fig 12


0.993 to 1.3
1.3 to 1.6
1.6 to 1.91
1.91 to 2.21
2.21 to 2.51
2.51 to 2.82
2.82 to 3.12
340000
335000
330000
325000
320000
315000
310000
305000
300000
250000
255000
260000
265000
270000
275000
280000
285000
290000
295000
0.993 to 1.67
1.67 to 2.54
2.54 to 3.12
340000
335000
330000
325000
320000
315000
310000
305000
300000
250000
255000
260000
265000
270000
275000
280000
285000
290000
295000

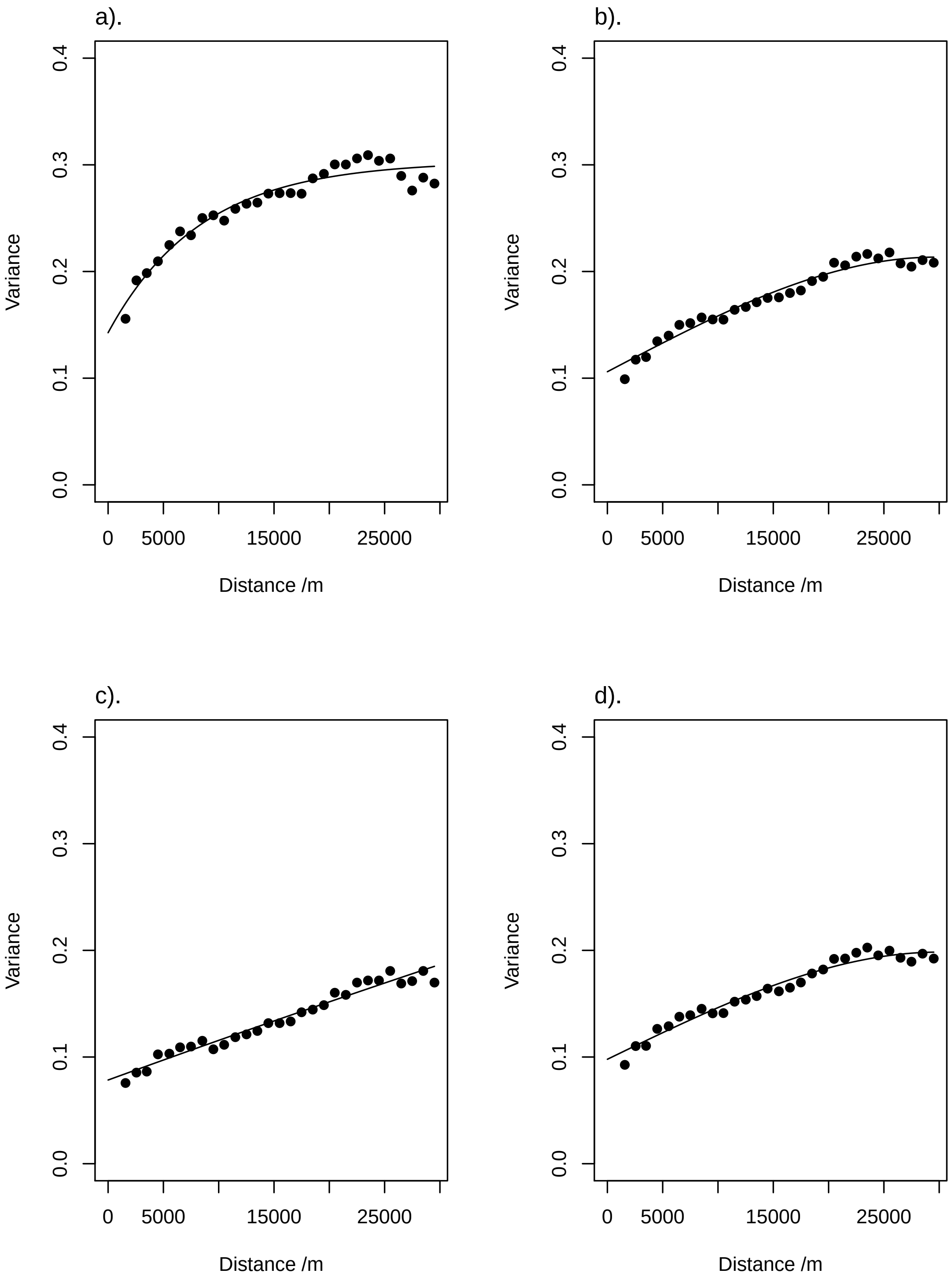
13: Fig 13


a).
Pb /mg kg^{-1}
0
200
400
600
800
1000
1200
b).
Frequency
0
200
400
600
800
0
200
600
1000
Pb /mg kg^{-1}
c).
Pb /log mg kg^{-1}
1
2
3
4
5
6
7
d).
Frequency
0
100
200
300
400
1
2
3
4
5
6
7
Pb /log mg kg^{-1}

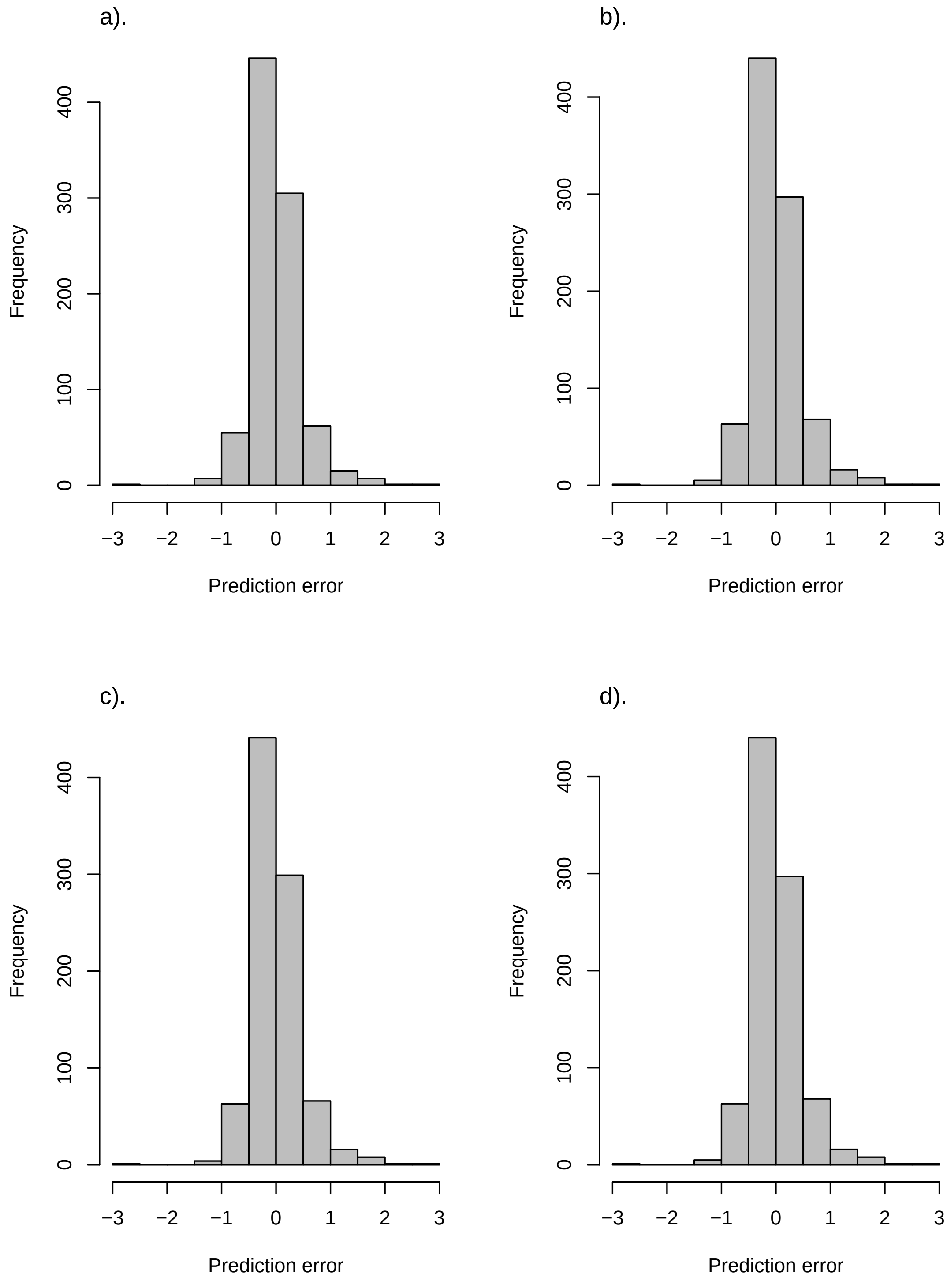
14: 43Fig 14


405000
400000
395000
390000
385000
380000
375000
455000
460000
465000
470000
475000
480000
485000
490000
495000
500000
505000
2 to 29
29 to 35
35 to 49
49 to 1301

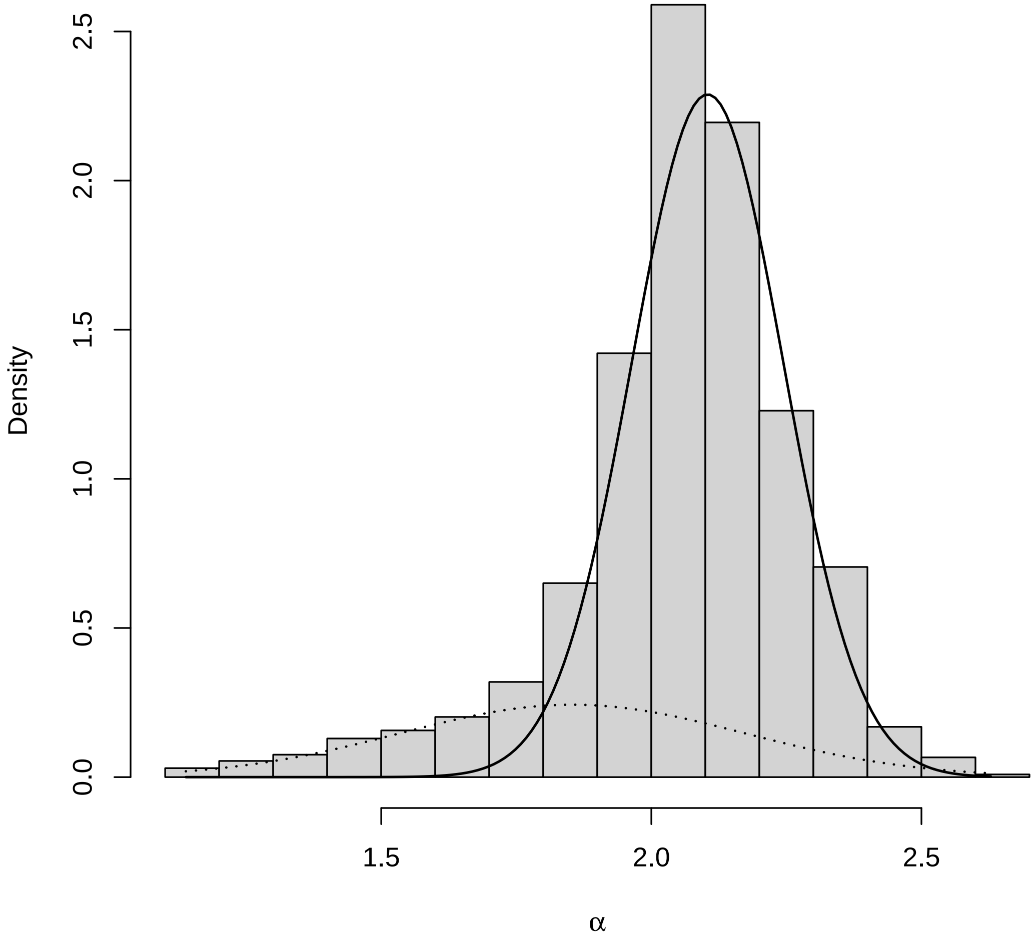
15: Fig 15



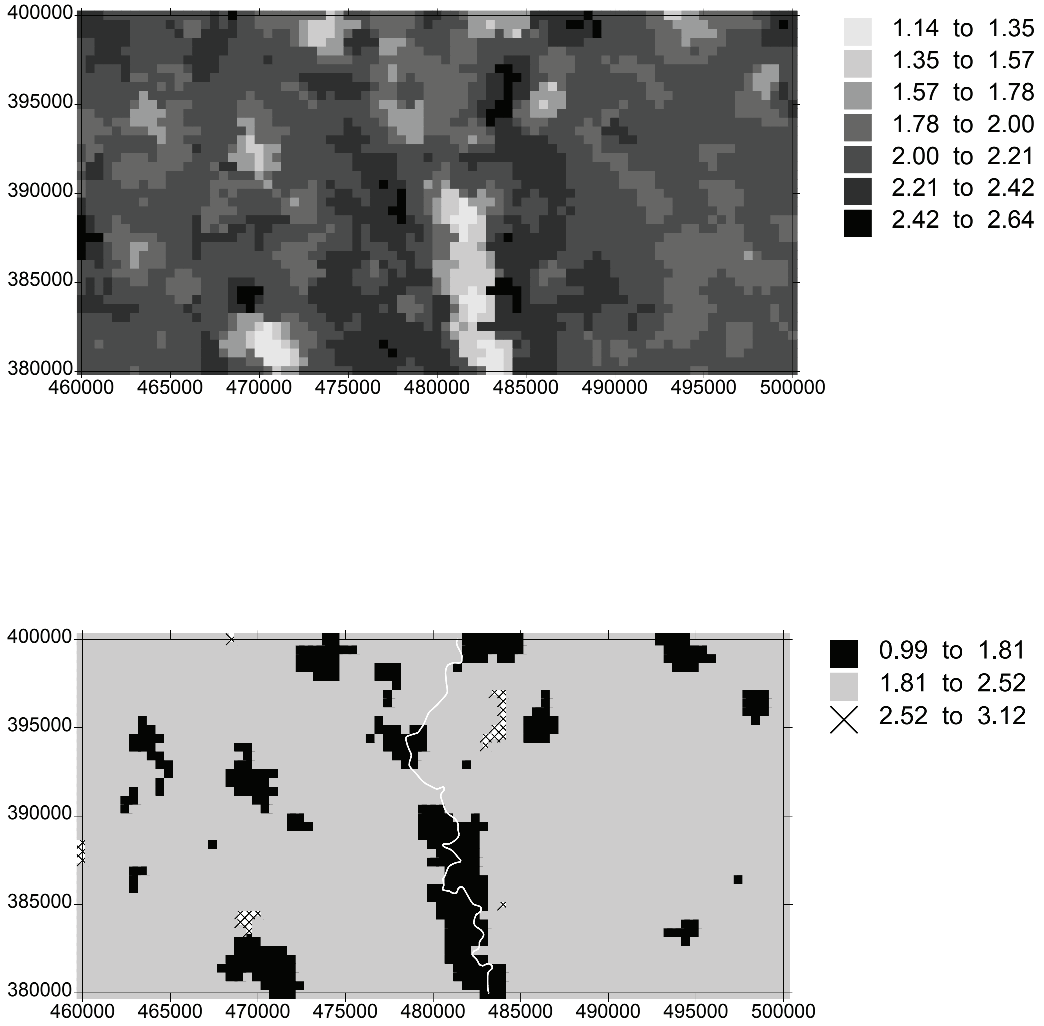
16: Fig 16



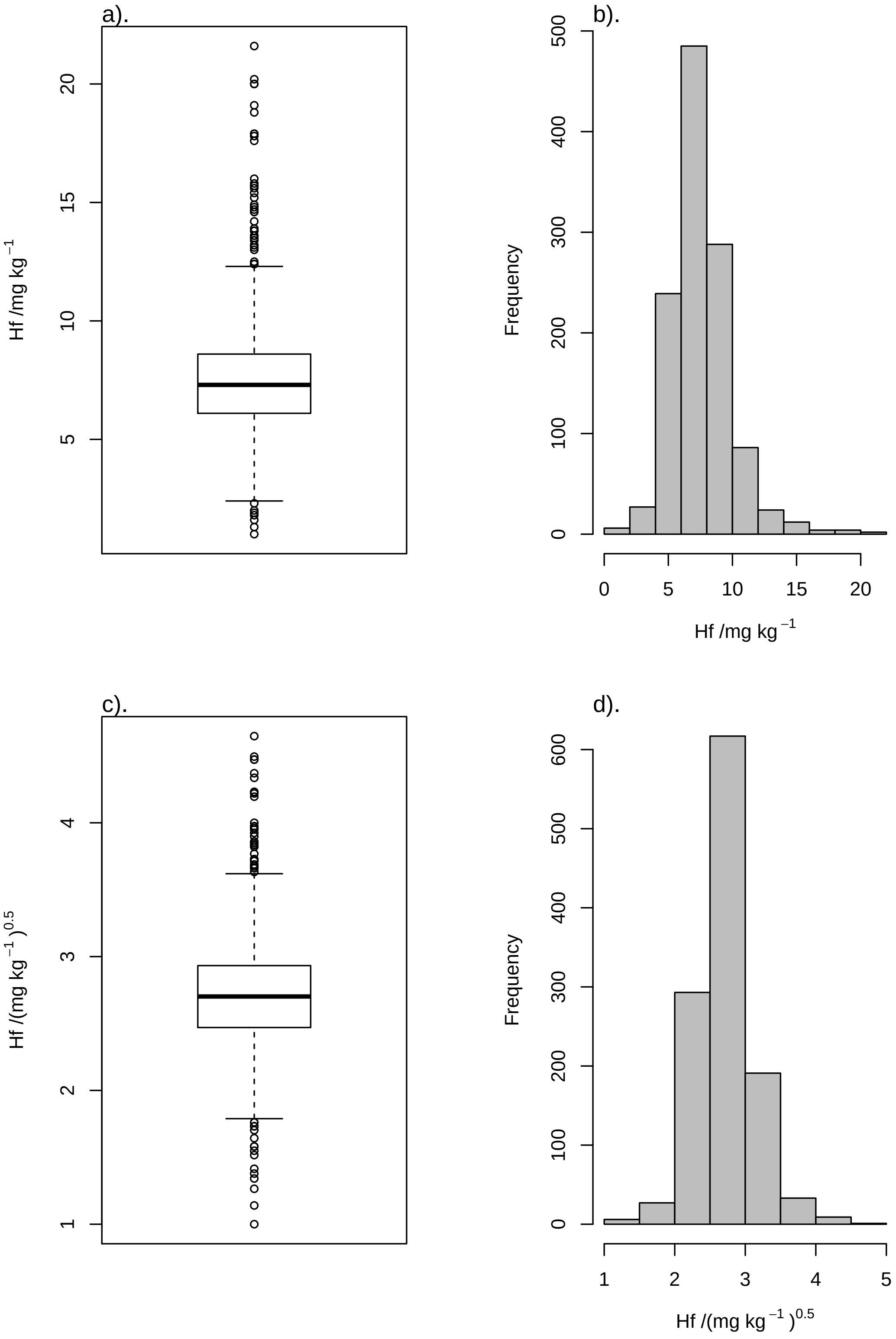
17: Fig 17



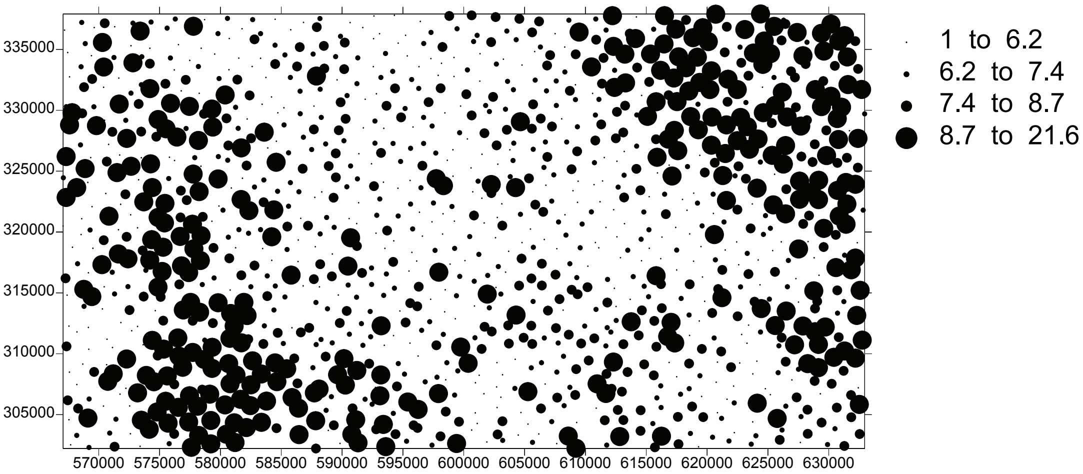
18: Fig 18


400000
395000
390000
385000
380000
460000
465000
470000
475000
480000
485000
490000
495000
500000
1.14 to 1.35
1.35 to 1.57
1.57 to 1.78
1.78 to 2.00
2.00 to 2.21
2.21 to 2.42
2.42 to 2.64
400000
395000
390000
385000
380000
460000
465000
470000
475000
480000
485000
490000
495000
500000
0.99 to 1.81
1.81 to 2.52
2.52 to 3.12

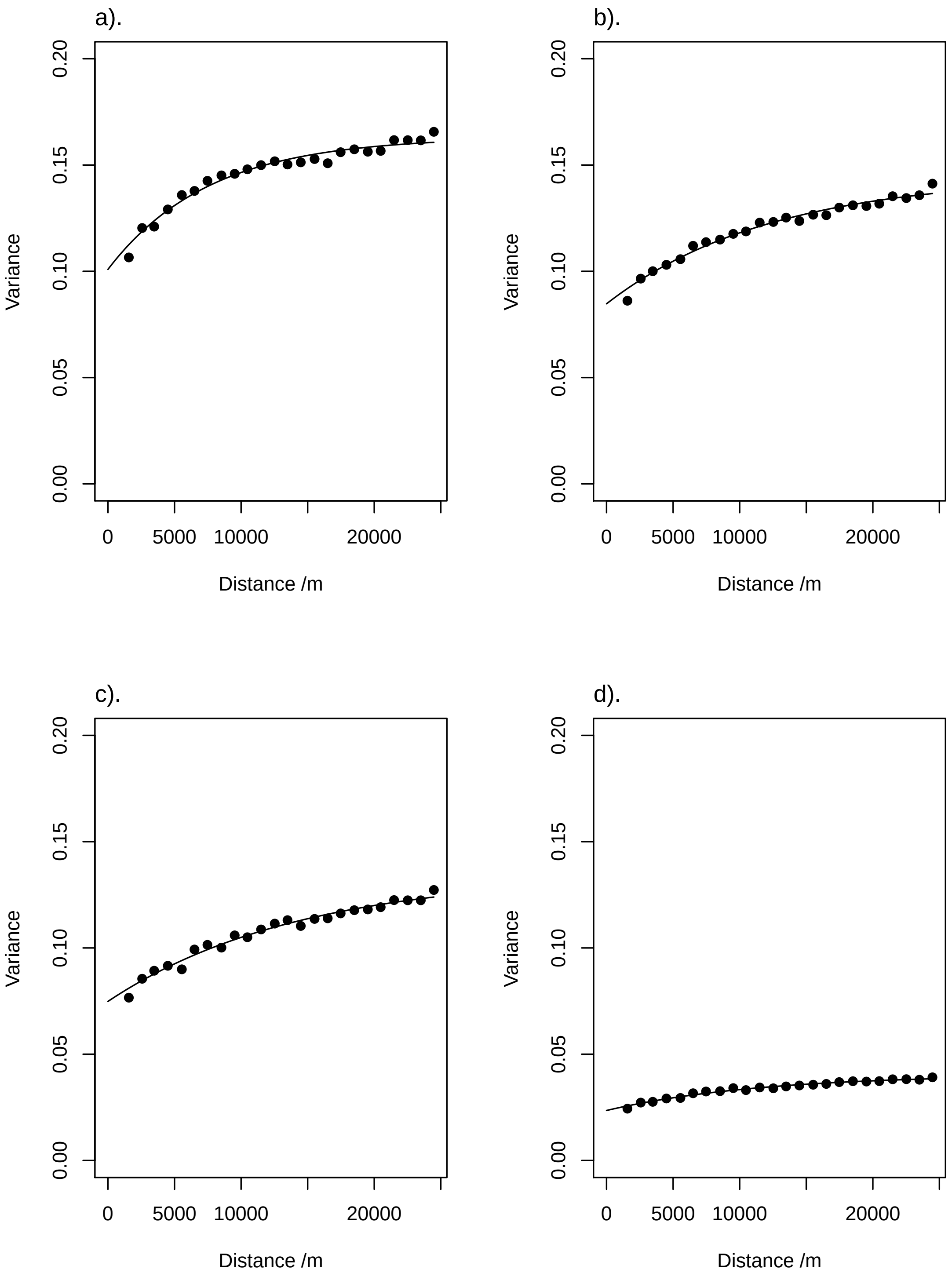
19: Fig 19


a).
b).
c).
d).
Hf /mg kg^{-1}
5
10
15
20
Frequency
0
100
200
300
400
500
0
5
10
15
20
Hf /mg kg^{-1}
Hf /(mg kg^{-1})$^{0.5}$
1
2
3
4
Frequency
0
100
200
300
400
500
600
1
2
3
4
5
Hf /(mg kg^{-1})$^{0.5}$

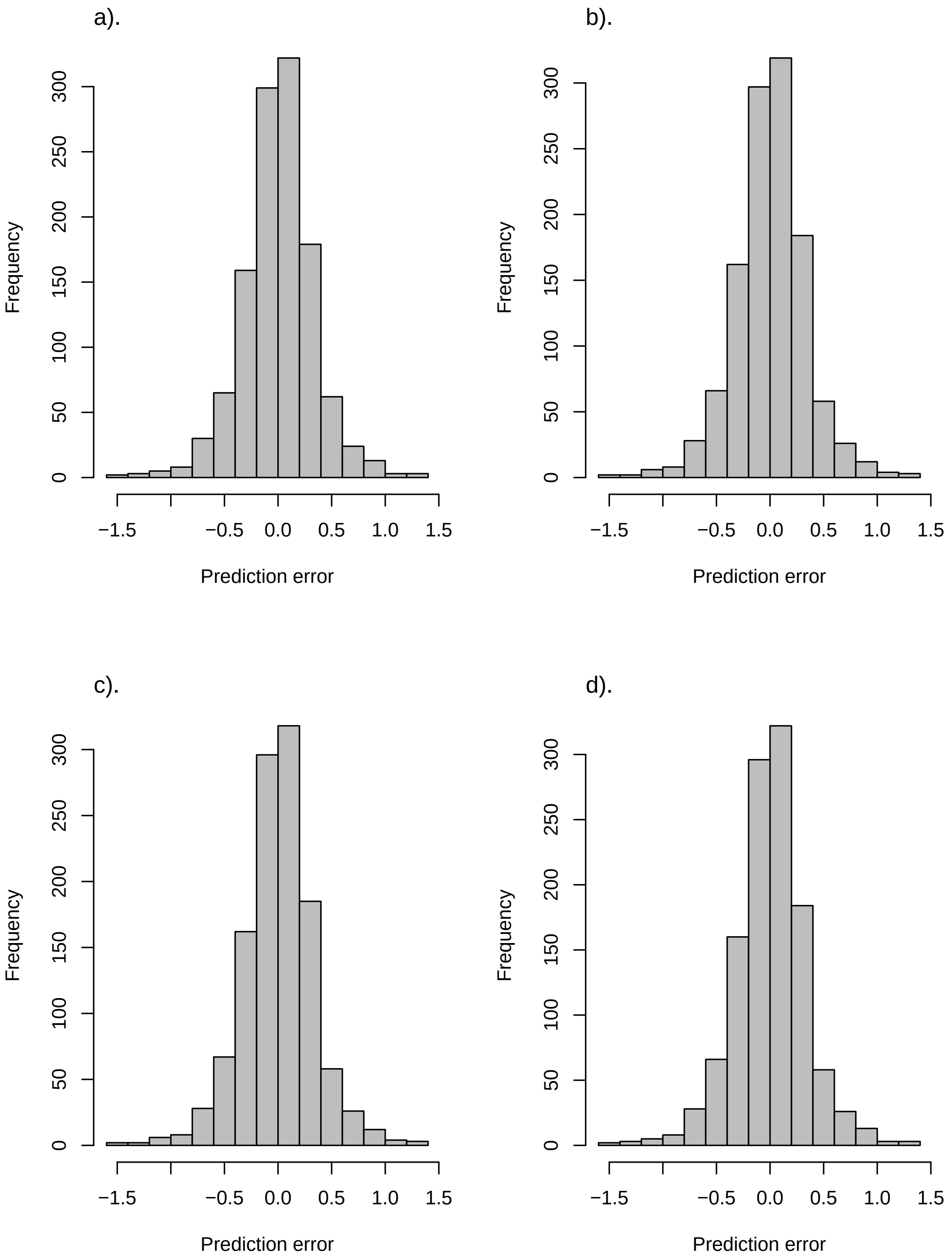
20: Fig 20


335000
330000
325000
320000
315000
310000
305000
570000
575000
580000
585000
590000
595000
600000
605000
610000
615000
620000
625000
630000
1 to 6.2
6.2 to 7.4
7.4 to 8.7
8.7 to 21.6

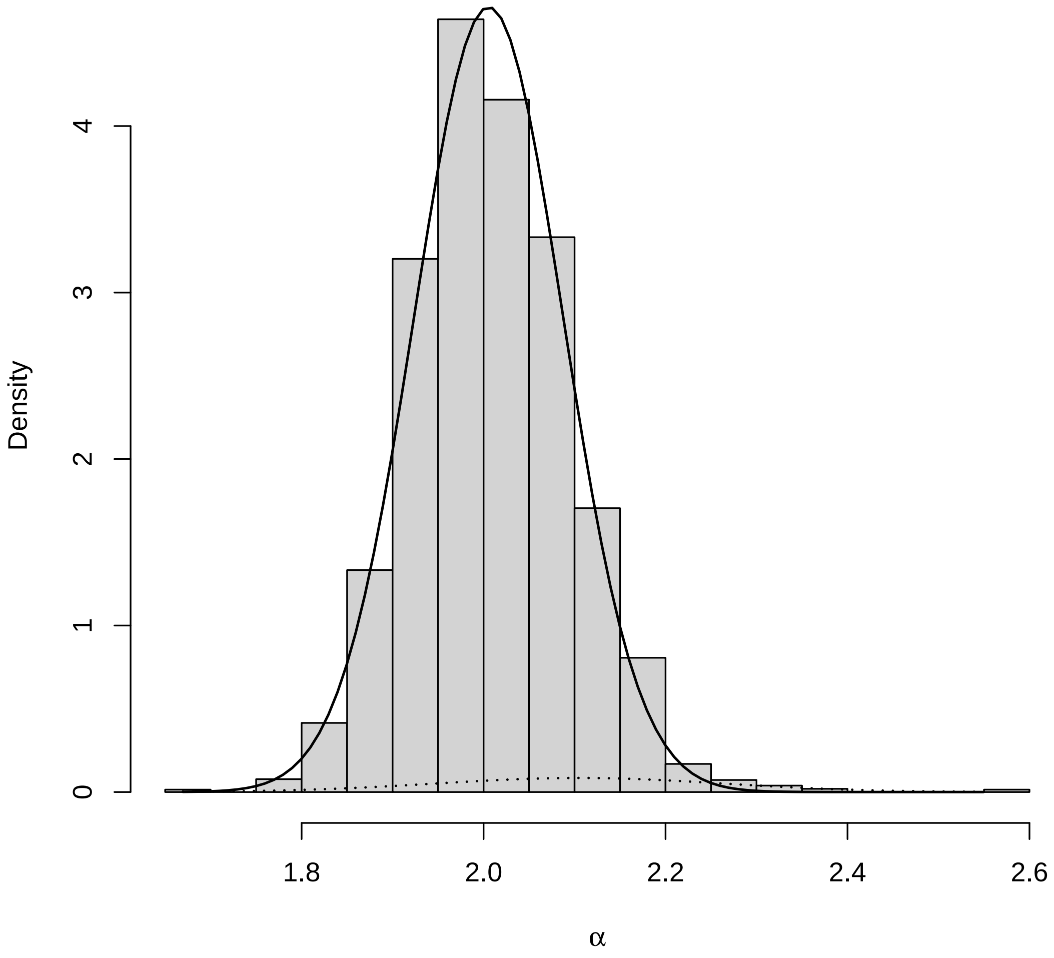
21: Fig 21


a).
b).
c).
d).
Variance
0.00
0.05
0.10
0.15
0.20
0
5000
10000
20000
Distance /m

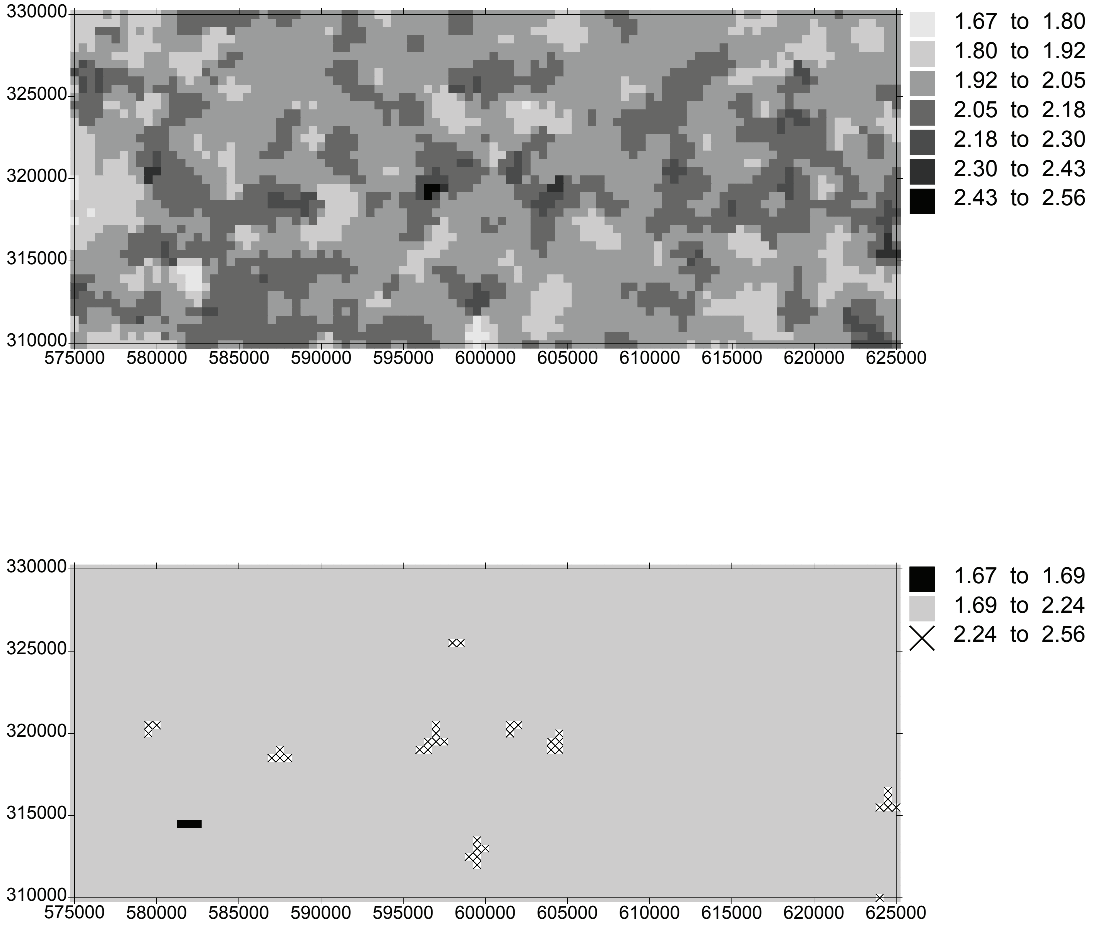
22: Fig 22


a).
b).
c).
d).
Frequency
0
50
100
150
200
250
300
−1.5
−0.5
0.0
0.5
1.0
1.5
Prediction error

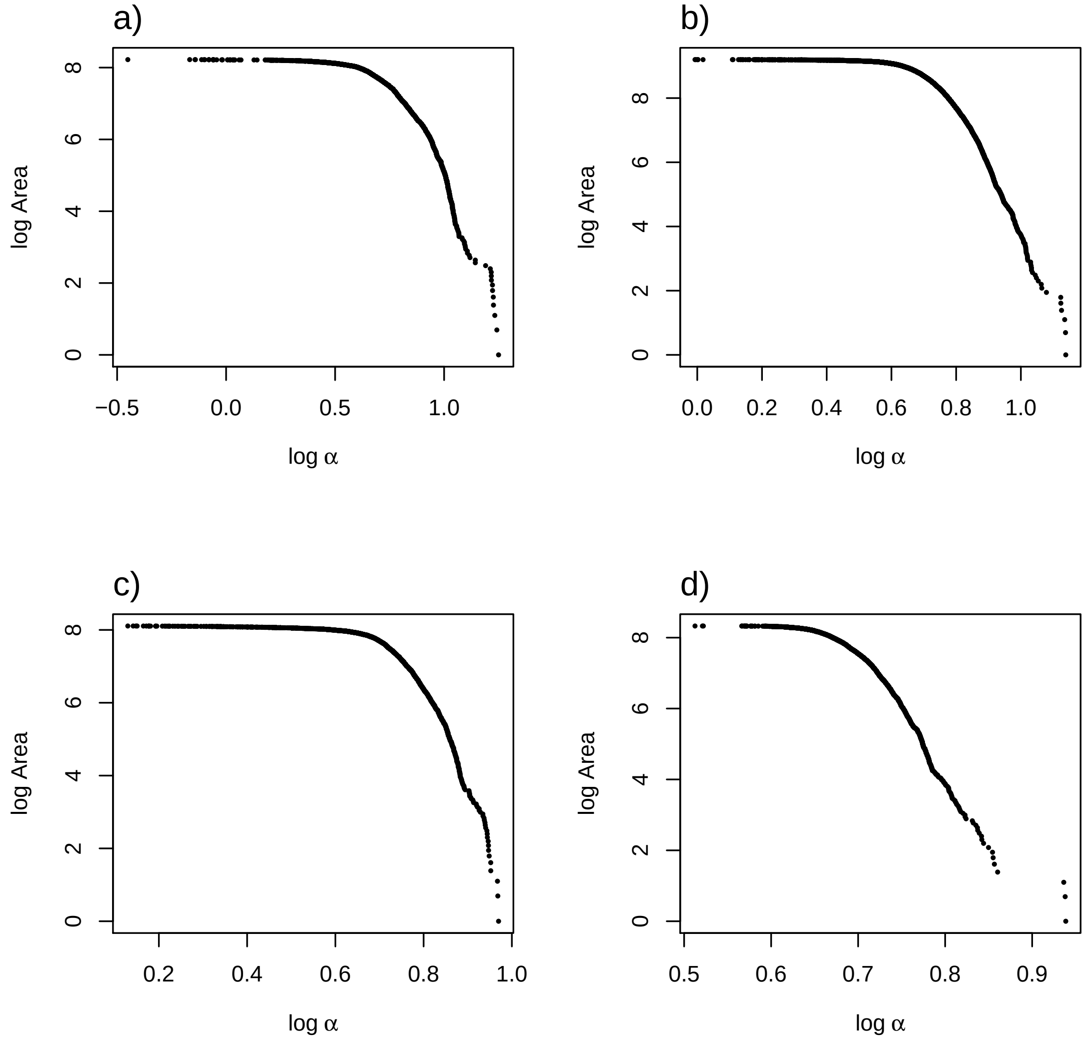
23: Fig 23



24: Fig 24

330000
325000
320000
315000
310000
575000
580000
585000
590000
595000
600000
605000
610000
615000
620000
625000
1.67 to 1.80
1.80 to 1.92
1.92 to 2.05
2.05 to 2.18
2.18 to 2.30
2.30 to 2.43
2.43 to 2.56
330000
325000
320000
315000
310000
575000
580000
585000
590000
595000
600000
605000
610000
615000
620000
625000
1.67 to 1.69
1.69 to 2.24
2.24 to 2.56

25: Fig 25


a)
b)
c)
d)
log Area
log α
8
6
4
2
0
−0.5
0.0
0.5
1.0
0.0
0.2
0.4
0.6
0.8
1.0
0.2
0.4
0.6
0.8
1.0
0.5
0.6
0.7
0.8
0.9

26: Fig 26