Geoderma 323 (2018) 83–106

Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

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

R.M. Lark[a,*], M. Patton[b], E.L. Ander[a], D.M. Reay[b]

[a] British Geological Survey, Keyworth, Nottinghamshire NG12 5GG, UK
[b] Geological Survey of Northern Ireland, Dundonald House, Upper Newtownards Road, Ballymiscaw, Belfast BT4 3SB, UK

ARTICLE INFO

Keywords:
Geochemistry
Anomalies
Singularity
Fractal
Mixture model

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.

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

* Corresponding author at: University of Nottingham, Sutton Bonington Campus, Sutton Bonington, LE12 5RD, UK.
E-mail address: murray.lark@nottingham.ac.uk (R.M. Lark).

https://doi.org/10.1016/j.geoderma.2018.02.032
Received 2 June 2017; Received in revised form 6 February 2018; Accepted 21 February 2018

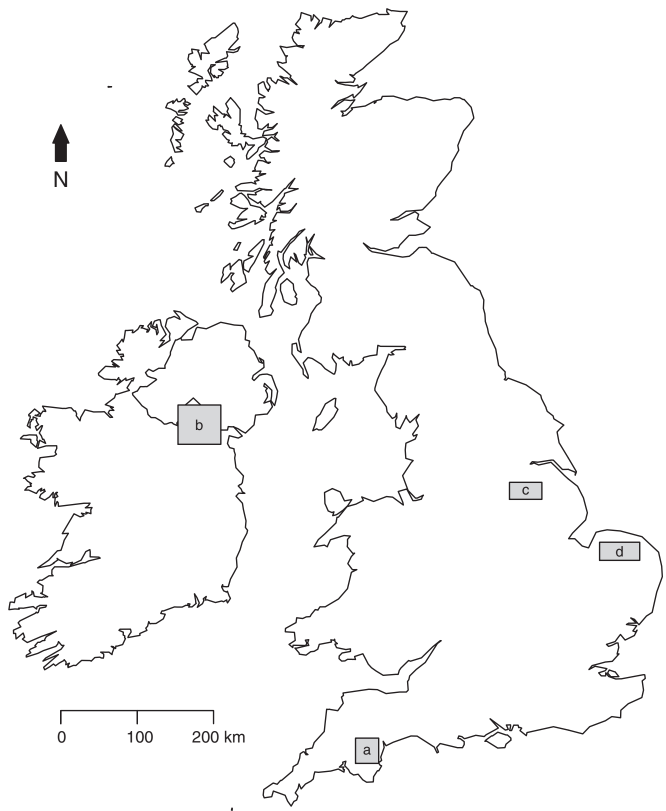

Fig. 1. Map of the United Kingdom and Ireland showing the locations of the rectangular regions within which the singularity index was evaluated for some element in the topsoil. a). East of Dartmoor in the south-west of England (Zn); b). Longford-Down terrane in Counties Monaghan and Armagh in the north 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).

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 $\mathscr{B}_{\mathbf{x}}(\varepsilon)$. The amount of some material within the support, $\mu(\mathscr{B}_{\mathbf{x}}(\varepsilon))$ depends on the local background concentration $c(\mathbf{x})$ scaled according to a local singularity index, $\alpha(\mathbf{x})$:

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

The equivalent expression for the mean concentration over the support is

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

Allègre 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,

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

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 Fig. 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. Lin 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 centred 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