Application of factorial kriging analysis to the FOREGS European topsoil geochemistry database

Claire E. Imrie*, Anna Korre, Gabriela Munoz-Melendez, Iain Thornton, Sevket Durucan

Department of Earth Science and Engineering, Imperial College London, London, SW7 2AZ, UK

ARTICLE INFO

ABSTRACT

Concern about increasing levels of trace elements in the environment has led to the development and implementation of a global programme to determine the current baseline levels of these chemicals in the Earth's surface. The FORum of European Geological Surveys (FOREGS) has recently published a geochemical database for Europe, while progress on similar databases is continuing in other major regions of the world. The FOREGS database comprises multimedia samples collected at a resolution of approximately 72×72 km from 26 European countries. This enables the investigation of the factors governing geochemical variation on a continental scale, potentially allowing contributions of natural processes to be appreciated prior to setting environmental quality standards. This paper investigates the variation in European topsoil geochemistry using factorial kriging analysis, which performs principal components analysis at different spatial scales. The results are interpreted with the aid of a GIS database. Four spatial scales were identified: a nugget component representing variation over a range less than the sampling density; a 'short' scale component with a range of 296 km; an 'intermediate' scale component (875 km); and a 'long' scale component (1750 km). The first three principal components (PCs) of the nugget covariance matrix explained 22.2% of the overall variance, representing local variation in geology, land use, weathering and organic matter content. The first two PCs of the short range structure explained 12.6% of the variance, representing variation according to the major structural divisions of Europe, and to carbonate content. The first PC of the intermediate structure explained 7.2% of the variance and was found to relate to glacial history and Quaternary deposition. Finally, the first three PCs of the long range structure explained 29.6% of the variance and represented variation due to mineralisation, soil texture, climate and possibly anthropogenic contamination.

1. Introduction

Throughout the world concern continues to grow over the threat posed to humans, animals and vegetation by chemicals in the environment. As pressures on the Earth's resources increase with economic and population growth, land degradation and pollution threaten to have serious implications for the sustainability of soil and water ecosystems and the essential resources they provide. As such, it is important to gauge changes to the levels of chemical elements in the Earth's surface, and this can only be achieved with a knowledge of the current baseline.

At present, a large number of local and national geochemical databases exist. However, due to the lack of an internationally agreed set of sampling and analysis standards, these databases have been collected according to different surveying and analytical protocols and as such are unsuitable for the straightforward establishment of geochemical baselines on a global scale. This has prompted an international initiative to document the levels of elements in the Earth's

surface in the form of a Global Reference Network (GRN) and using specified methods (Darnley et al., 1995).

To proceed with the collection of the GRN samples, the International Union of Geological Sciences (IUGS) formed a working group on Global Geochemical Baselines in 1997. A number of large-scale projects have been initiated, with varying degrees of progress (IUGS/IAGC, 2005). A sampling campaign has been commenced in India, workshops have been held in Africa and pilot studies have been completed in Australia and North America (Smith et al., 2005). In Europe, however, the FOREGS Geochemical Atlas has been completed.

Implementation of the FOREGS Geochemical Baseline Mapping Programme was motivated by an urgent need for data on geochemical baselines in Europe (Salminen et al., 2005). While this benefits the environmental authorities in most European countries who are specifying limits for contaminants in soils according to prospective land use, the European Union is also in the process of constructing the Soil Protection Directive. State authorities tend to define standards without consideration of natural variability in soil geochemistry due to geology and soil-forming processes.

During the FOREGS campaign, samples of stream waters, stream sediments, soils and humus were collected using a fixed sampling protocol from random locations within a regular grid. The samples were then analysed at dedicated laboratories for an extensive range of elements. The survey was not designed to show the lowest natural background concentrations in Europe, but to demonstrate the current geochemistry of the surface environment (Salminen et al., 1998).

If it is assumed that the local ecosystems existing in and on soils are adapted to the prevailing background geochemistry, then added anthropogenic contamination may induce adverse effects at certain levels. Therefore, the setting of maximum allowable concentrations for metals in soils should take into account the local natural background levels. However, the processes contributing towards soil geochemistry are complex and manifold, and as such it is very difficult to know with certainty what the natural background concentration should be. This paper describes an attempt to elucidate the processes acting on the topsoil geochemistry of Europe using factorial kriging analysis, with a view to the future quantification of natural background concentrations of Co, Cr, Cu, Ni, Pb and Zn.

Many researchers have endeavoured to calculate ranges of background values for many soil constituents. Several methods (summarised in Kabata-Pendias, 2001) require a distribution of values for a given element measured within a geochemically and pedologically homogeneous area. Many of these methods assume that a geochemical background is already known, or that samples from completely pristine areas are available. These techniques are not applicable on a continental scale, where a wide range of parent materials and climatic factors will govern the topsoil geochemistry.

A commonly used approach for identifying anthropogenic input is based on enrichment factors. To account for some natural variation due to parent materials, these use ratios of the element of interest with a reference element which is not generally associated with human redistribution, such as Al, Zr, Sc and Ti. The ratios are compared with the ratio of the two elements' concentrations in the crust as defined by Clarke (1924). Reimann and de Caritat (2000, 2005), however, warn

against the use of enrichment factors for the following reasons: misinterpretations may arise due to differences in measurement techniques; depositions of dust may be naturally enriched in heavy metals; there are significant differences amongst researchers as to what level of enrichment constitutes proof of contamination; and the method does not account for enrichment due to biogenic processes.

Rawlins et al. (2002) attempted to determine levels of diffuse pollution by correlating metal concentrations with population density. The authors reported that diffuse input had given rise to increases of 11 to 40% in trace element concentrations in populous areas. Rawlins et al. (2003) applied multivariate geostatistical techniques to demonstrate the dependence of soil geochemistry on its parent materials. The authors suggested that parent materials could be used to imply mean regional background values where soil chemistry data are sparse.

Various authors (e.g., Zhang et al., 1999; Orescanin et al., 2004) have employed principal components analysis and cluster techniques to identify the processes governing the distribution of heavy metals in soils and other media. Korre (1999a,b) applied factor analysis and subsequent canonical correlation, GIS and geostatistical techniques to assess trace element contamination and its spatial distribution in soils.

Principal components analysis (PCA) and factor analysis (FA) are popular methods for reducing the dimensionality of multivariate datasets in order to create a smaller number of 'factors'. An attempt can then be made to interpret these factors in terms of various underlying processes. Factorial kriging analysis (FKA, Matheron, 1982) is a multivariate geostatistical technique that takes this one step further by allowing the causes of variability in a dataset to be first distinguished according to the spatial scales at which they operate. As well as filtering out variance due to error, this enables the resulting factors to be interpreted according to local, short range and regional processes.

FKA has been successfully applied in various fields including remote sensing (e.g., Rodgers and Oliver, 2007; Oliver et al., 2000; van Meirvenne and Goovaerts, 2002), hydrogeology (e.g., Lin et al., 2004, 2006) and ecology (e.g., Castriganò et al., 2000). Within the field of soil science, it has been used to assess variation in soil properties for crop management (e.g., Dobermann et al., 1995; Bocchi et al., 2000), geochemical exploration (e.g., Jimenez-Espinoza and Chica-Olmo, 1999; Batista et al., 2001; Reis et al., 2004), pollution (e.g., Einax and Soldt, 1998) and analysis of underlying geochemical processes (e.g., Bourennane et al., 2003; Xu and Tao, 2004).

The FOREGS Geochemical Atlas (Part 2) includes an exploratory application of FA to the soil, stream and floodplain sediment data (Batista et al., 2006). In brief, the first seven factors were interpreted as follows: 1) rare earth and lithophile elements; 2) mafic and ultramafic rocks; 3) grain size and chalcophile elements; 4) felsic rocks; 5) carbonate versus siliciclastic rocks; 6) elements typical for pollution and association with organic matter; and 7) siliciclastic rocks. However, it was acknowledged that the study could have been improved by removing the redundant variables and avoiding the effect of closure of compositional variables.

The current paper describes the application of factorial kriging analysis to the FOREGS topsoil data. The elements