



The concept of compositional data analysis in practice – Total major element concentrations in agricultural and grazing land soils of Europe

Clemens Reimann ^{a,*}, Peter Filzmoser ^b, Karl Fabian ^a, Karel Hron ^c, Manfred Birke ^d, Alecos Demetriades ^e, Enrico Dinelli ^f, Anna Ladenberger ^g and The GEMAS Project Team ¹

^a Geological Survey of Norway, PO Box 6315 Sluppen, N-7491 Trondheim, Norway

^b Institute for Statistics and Probability Theory, Vienna University of Technology, Wiedner Hauptstrasse 8–10, A-1040 Wien, Austria

^c Department of Mathematical Analysis and Applications of Mathematics, Palacký University, Faculty of Science, Listopadu 12, CZ-77146 Olomouc, Czech Republic

^d Federal Institute for Geosciences and Natural Resources (BGR), Branch office Berlin, Wilhelmstr. 25–30, D-13593 Berlin, Germany

^e Institute of Geology and Mineral Exploration, Entrance C, Olympic Village, Acharnae, Athens, GR-13677, Greece

^f University of Bologna, Department of Earth Science, Piazza di Porta San Donato 1, I-40126 Bologna, Italy

^g Geological Survey of Sweden (SGU), Box 670, S-751 28 Uppsala, Sweden

ARTICLE INFO

Article history:

Received 20 December 2011

Received in revised form 14 February 2012

Accepted 14 February 2012

Available online 12 April 2012

Keywords:

Agricultural soil

XRF

Major elements

Europe

Geochemistry

Compositional data

ABSTRACT

Applied geochemistry and environmental sciences invariably deal with compositional data. Classically, the original or log-transformed absolute element concentrations are studied. However, compositional data do not vary independently, and a concentration based approach to data analysis can lead to faulty conclusions. For this reason a better statistical approach was introduced in the 1980s, exclusively based on relative information. Because the difference between the two methods should be most pronounced in large-scale, and therefore highly variable, datasets, here a new dataset of agricultural soils, covering all of Europe (5.6 million km²) at an average sampling density of 1 site/2500 km², is used to demonstrate and compare both approaches. Absolute element concentrations are certainly of interest in a variety of applications and can be provided in tabulations or concentration maps. Maps for the opened data (ratios to other elements) provide more specific additional information. For compositional data XY plots for raw or log-transformed data should only be used with care in an exploratory data analysis (EDA) sense, to detect unusual data behaviour, candidate subgroups of samples, or to compare pre-defined groups of samples. Correlation analysis and the Euclidean distance are not mathematically meaningful concepts for this data type. Element relationships have to be investigated via a stability measure of the (log-)ratios of elements. Logratios are also the key ingredient for an appropriate multivariate analysis of compositional data.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Geochemistry aims to quantitatively determine the chemical composition of the Earth and its parts and to discover the factors that control the distribution of individual elements (Goldschmidt, 1937,

1954). Geochemical studies need to be carried out from the atomic to the continental and finally global (some may argue cosmic) scale (for discussions of scale see: Darnley et al., 1995; Reimann et al., 2009, 2010) to meet these aims.

Geochemical data are usually reported as concentrations in units of mg/kg or weight percent (wt.%) and are thus a classical example of compositional (closed) data (CoDa – Aitchison, 1986). If all chemical elements in a sample are analysed, the analytical results sum up to a constant (1,000,000 mg/kg or 100 wt.%). Thus no single variable is free to vary separately from the rest of the total composition. Even if not all chemical elements are analysed, the total element concentrations still depend on each other. The relevant information for each single variable in a geochemical dataset thus lies in the ratios between all variables and not in the measured element concentrations as such. An interpretation and statistical evaluation of the observed concentration values is only meaningful if the relationship to the values of the remaining variables is taken into account (Aitchison, 1986; Filzmoser et al., 2010). It could hence be argued that a multi-element geochemical

* Corresponding author. Tel.: +47 73 904 307.

E-mail address: Clemens.Reimann@ngu.no (C. Reimann).

¹ S. Albanese, M. Andersson, A. Arnoldussen, R. Baritz, M.J. Batista, A. Bel-lan, D. Cicchella, B. De Vivo, W. De Vos, M. Duris, A. Dusza-Dobek, O.A. Eggen, M. Eklund, V. Ernstsen, T.E. Finne, D. Flight, S. Forrester, M. Fuchs, U. Fugedi, A. Gilucis, M. Gosar, V. Gregorauskiene, A. Gulan, J. Halamić, E. Haslinger, P. Hayoz, G. Hobiger, R. Hoffmann, J. Hoogewerff, H. Hrvatovic, S. Husnjak, L. Janik, C.C. Johnson, G. Jordan, J. Kirby, J. Kivisilla, V. Klos, F. Krone, P. Kwecko, L. Kuti, A. Lima, J. Locutura, P. Lucivjansky, D. Mackovych, B.I. Malyuk, R. Maquil, M.J. McLaughlin, R.G. Meuli, N. Miosic, G. Mol, P. Négrel, P. O'Connor, K. Oorts, R. T. Ottesen, A. Pasieczna, V. Petersell, S. Pfeleiderer, M. Poňavič, C. Prazeres, U. Rauch, . Salpeteur, A. Schedl, A. Scheib, I. Schoeters, P. Sefcik, E. Sellersjö, F. Skopljak, I. Slaninka, A. Šorša, R. Srvkota, T. Stafilov, T. Tarvainen, V. Trendavilov, P. Valera, V. Verougstraete, D. Vidojević, A.M. Zissimos, Z. Zomeni.

dataset should only be analysed in multivariate space, without even considering the univariate case. However, a careful univariate data analysis has always been the starting point of statistical analyses of regional geochemical datasets (Reimann et al., 2008). This is a reasonable approach because it helps to better understand the behaviour of the data before more sophisticated multivariate techniques are applied. For example, the very aim of a regional geochemical mapping project is to study and predict the distribution (concentration) of a chemical element in two-dimensional space. Such maps have been successfully used to aid geological mapping, for mineral exploration, for documenting contamination, and for detecting a multitude of additional processes that determine the distribution of chemical elements at the Earth's surface. It will be hard to convince a regional geochemist that all these maps are “wrong” and that dimensionless ratio maps (which ratio?) are the only correct maps. Thus, while many solutions to the closure problem exist for multivariate data analysis (e.g., Aitchison and Greenacre, 2002; Buccianti and Pawlowsky-Glahn, 2005; Buccianti et al., 2006; Egozcue and Pawlowsky-Glahn, 2011; Filzmoser and Hron, 2008; Filzmoser et al., 2009b; Hron et al., 2010; Otero et al., 2005; Pawlowsky-Glahn and Buccianti, 2002, 2011; Tolosana-Delgado and van den Boogaart, 2011; von Eynatten et al., 2003), a sensible approach to univariate and bivariate data analysis of compositional data, satisfying the statistician as well as the geochemist, is still under development (Filzmoser et al., 2009a, 2010).

The administration of the new European REACH (Registration, Evaluation and Authorisation of Chemical – EC, 2006) regulation, which came into force on the 1st of June, 2007, requires knowledge about “soil quality” at the European scale. The GEMAS (Geochemical mapping of agricultural and grazing land soils) project, a cooperation project between EuroGeoSurveys and Eurometaux, aims at providing such data for Europe. Samples of agricultural soil and of soil under permanent grass cover were taken during 2008 at an average density of 1 site/2500 km², covering the member states of the European Union (exception Malta and Romania) and several neighbouring countries (e.g., Norway, Serbia, Ukraine). In total, over 4000 samples were collected, prepared and analysed (see also Reimann et al., 2012a). The total concentrations of the major elements (Al₂O₃, CaO, Fe₂O₃, K₂O, MgO, MnO, Na₂O, P₂O₅, SiO₂, and TiO₂, plus Loss on Ignition (LOI)) in the soil samples, reported for the GEMAS project analysed by X-ray fluorescence spectrometry (XRF – LOI gravimetric) are a “classical” example of a “closed” dataset. This dataset is used here:

- (1) To report the concentration of major elements in European agricultural soils.
- (2) To study the regional distribution of the major elements in order to better understand the processes governing the distribution of chemical elements in European agricultural soils and their relative importance at the continental scale.
- (3) To investigate effects of data closure and to understand which evaluation procedures may be applied to such data, and which should be avoided.
- (4) To compare alternative data analysis techniques to the classical way of treating geochemical data.
- (5) To further develop recommendations for the uni-, bi- and multivariate investigation of compositional datasets.

1.1. The survey area

Maps covering topography and land use for Europe can be found in almost any atlas. A number of maps covering different themes at about the scale of the GEMAS project (topography, geology, tectonics, fault and fracture zones, distribution of different rock types, distribution of the main sedimentary basins, precipitation and population density) are collected in Reimann and Birke (2010). Fig. 1 shows a simplified geological map including the main geological structures discussed in this paper. For Europe an excellent source of land use information is

the CORINE land cover map of Europe (GLC2000 database, 2003). A detailed geological map of Europe is provided by Asch (2003), and concise descriptions of the geology of Europe can be found in Ziegler (1990), Blundell et al. (1992) and McCann (2008). The soil atlas of Europe provides a wealth of information on European soils, but also contains maps of average precipitation, temperature, land use, population density, extent of the last glaciation, and soil texture (Jones et al., 2005).

2. Material and methods

2.1. Project background and sampling

GEMAS is a cooperation project between the Geochemistry Expert Group of EuroGeoSurveys (EGS) and Eurometaux. The GEMAS project aims to produce consistent soil geochemistry data at the continental scale in accordance to REACH (EC, 2006) requirements. REACH specifies that industry must prove that it can produce and handle its substances safely. Risks due to the exposure to a substance during production and use at the local, regional and European scale all need to be assessed. Industries handling metals needed harmonised data on the natural distribution of chemical elements, and of soil properties governing metal availability in soils at the continental scale. REACH requires that risk assessment is performed according to land use. The GEMAS project focused on agricultural soils from arable and grazing land, both linked to the human food chain. According to REACH the sample depth should be 0–20 cm for agricultural soils (arable land, Ap-horizon) and 0–10 cm for grazing land soils (land under permanent grass cover) and the <2 mm grain size is the fraction to be analysed. With the exception of the sample density, the sampling requirements were thus rigidly fixed by external requirements.

With regard to sample density it was decided to follow the example of an earlier project, covering Northern Europe (the Baltic Soil Survey: Reimann et al., 2003) and to sample one site per 2500 km² (50 × 50 km grid). The grid cells were centrally provided, but the sample teams were free to decide where in a grid cell the two samples of agricultural and grazing land soil were taken. Sample materials and especially the bags used for storing the samples were centrally provided to all field teams.

Samples were taken as composites from 5 sites spread over a ca. 100 m² area in a large agricultural field (Ap-sample) and on land under permanent grass cover (Gr sample). The average weight of a sample was 3.5 kg. It was attempted to find sample sites for the Ap and Gr samples in as close proximity as possible. The average distance between the two sites is 500 m, but, depending on land use, single sample pairs, where the sites are more than 50 km apart do occur. All sites and the soil profile at any one site were documented in a series of photographs. Field procedures are detailed in a field handbook which is freely available on the internet (EGS, 2008). For quality control purposes, a field duplicate was taken at every 20th sample site with an offset distance of ca. 10–20 m from the original sample site.

2.2. Sample preparation

All samples were prepared in a central laboratory (Geological Survey of the Slovak Republic). The samples were air dried and sieved to pass a 2 mm nylon screen. All samples were then randomised and analytical duplicates and project standards were introduced at a rate of 1 in 20. All samples were then split into ten aliquots using a Jones Riffle splitter. Four splits of ~200 g each are stored for future reference, and 6 splits of 50–100 g each were sent to the different contract laboratories for the immediate analytical work. For analysis by XRF the samples needed to be milled prior to further sample preparation. One of the small sample splits was milled to less than 63 μm in an agate disc mill at BGR's laboratory in Germany. Loss on Ignition (LOI) was then determined on all samples via slowly heating to 1030 °C, keeping them at this temperature for 15 min in a muffle furnace, letting them cool to room

Download English Version:

<https://daneshyari.com/en/article/4429472>

Download Persian Version:

<https://daneshyari.com/article/4429472>

[Daneshyari.com](https://daneshyari.com)