Journal of Geochemical Exploration 162 (2016) 16–28

Contents lists available at ScienceDirect

Journal of Geochemical Exploration

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

The single component geochemical map: Fact or fiction?

Jennifer M. McKinley [a,*], Karel Hron [b], Eric C. Grunsky [c], Clemens Reimann [d], Patrice de Caritat [e,f], Peter Filzmoser [g], Karl Gerald van den Boogaart [h], Raimon Tolosana-Delgado [h]

[a] School of Geography, Archaeology and Palaeoecology, Queen's University Belfast, BT7 1NN, UK
[b] Dept. of Math. Anal. & Appl. of Math., Palacky University Olomouc, 17. listopadu 12, CZ-771 46 Olomouc, Czech Republic
[c] Department of Earth and Environmental Sciences, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada
[d] Geological Survey of Norway, PO Box 6315, Sluppen, N-7491 Trondheim, Norway
[e] Geoscience Australia, GPO Box 378, Canberra, ACT 2601, Australia
[f] Research School of Earth Sciences, Australian National University, Canberra, ACT 2601, Australia
[g] Institute of Statistics & Mathematical Methods in Economics, Vienna University of Technology, Wiedner Hauptstr. 8-10, 1040 Vienna, Austria
[h] Helmholtz Center Dresden Rossendorf, Helmholtz Institute Freiberg for Resource Technology, Freiberg, Germany

ARTICLE INFO

Article history:
Received 28 May 2015
Revised 1 December 2015
Accepted 13 December 2015
Available online 15 December 2015

Keywords:
Soil geochemistry
Compositional data analysis
Log-ratios
Mapping

ABSTRACT

Single component geochemical maps are the most basic representation of spatial elemental distributions and commonly used in environmental and exploration geochemistry. However, the compositional nature of geochemical data imposes several limitations on how the data should be presented. The problems relate to the constant sum problem (closure), and the inherently multivariate relative information conveyed by compositional data. Well known is, for instance, the tendency of all heavy metals to show lower values in soils with significant contributions of diluting elements (e.g., the quartz dilution effect); or the contrary effect, apparent enrichment in many elements due to removal of potassium during weathering. The validity of classical single component maps is thus investigated, and reasonable alternatives that honour the compositional character of geochemical concentrations are presented. The first recommended such method relies on knowledge-driven log-ratios, chosen to highlight certain geochemical relations or to filter known artefacts (e.g. dilution with SiO_2 or volatiles). This is similar to the classical normalisation approach to a single element. The second approach uses the (so called) log-contrasts, that employ suitable statistical methods (such as classification techniques, regression analysis, principal component analysis, and clustering of variables) to extract potentially interesting geochemical summaries. The caution from this work is that if a compositional approach is not used, it becomes difficult to guarantee that any identified pattern, trend or anomaly is not an artefact of the constant sum constraint. In summary the authors recommend a chain of enquiry that involves searching for the appropriate statistical method that can answer the required geological or geochemical question whilst maintaining the integrity of the compositional nature of the data. The required log-ratio transformations should be applied followed by the chosen statistical method. Interpreting the results may require a closer working relationship between statisticians, data analysts and geochemists.

1. Introduction

Over the last decade the focus of digital soil and sediment databases at a regional, national, transnational and continent scale has increasingly become to provide information for a range of purposes including geological and soil mapping, baseline quality documentation, mineral prospecting, land and soil resource assessment, risk evaluation, environmental and educational purposes, and prediction of soil provenance for forensic purposes (Smith et al., 1997; Reimann et al., 1998, 2003; Morris et al., 2003; Lagacherie et al., 2007; McKinley, 2013). These digital soil databases are generally accompanied by geochemical atlases (e.g. de Caritat and Cooper, 2011a; Young and Donald, 2013; Reimann et al., 2014a, 2014b) showing a collection of distribution maps for individual geochemical elements. This has been matched by a corresponding increase in published studies utilising these soil geochemical surveys (e.g. Chiprés et al., 2008; Grunsky et al., 2009; Carranza, 2010; Ohta et al., 2011; de Caritat and Grunsky, 2013; Cheng et al., 2014; Lancianese and Dinelli, 2014; Birke et al., 2015). The resolution of the ground-based sampling scheme used for the generation of these databases is the best compromise between the extent of the region covered, and time and resources available. Over a local to regional scale, soil sampling can be managed on a 2 km^2 grid as applied by the Geological

* Corresponding author.
E-mail addresses: j.mckinley@qub.ac.uk (J.M. McKinley), hronk@seznam.cz (K. Hron), egrunsky@gmail.com (E.C. Grunsky), Clemens.Reimann@NGU.NO (C. Reimann), Patrice.deCaritat@ga.gov.au (P. de Caritat), P.Filzmoser@tuwien.ac.at (P. Filzmoser), boogaart@hzdr.de (K.G. van den Boogaart), r.tolosana@hzdr.de (R. Tolosana-Delgado).

http://dx.doi.org/10.1016/j.gexplo.2015.12.005
0375-6742/

Survey of Northern Ireland (GSNI)'s Tellus project. A regional scale dataset such as this is close to exhaustive sampling in terms of geological survey mapping, reducing the need for interpolation. It is worth noting however that this is not exhaustive sampling for mineral exploration. Generally surveys are carried out at local (1:10,000), regional (1:250,000) or continental (1:1,000,000) scales. A number of authors discuss the scale of geochemical mapping (e.g. Bolviken et al., 1992; Darnley et al., 1995; Reimann et al., 2010).

The sampling strategy generally follows standard protocols (e.g. UNESCO's IGCP Global Geochemical Database — Darnley et al., 1995; G-BASE standard established by BGS — Johnson, 2005; FOREGS field handbook — Salminen et al., 1998; GEMAS field handbook — EGS, 2008; North American Soil Geochemical Landscapes Project — Smith et al., 2011; National Geochemical Survey of Australia field manual — Lech et al., 2007; Canadian component of the North American Soil Geochemical Landscapes Project — Friske et al., 2013; China Geochemical Baseline Project — Wang and the CGB Sampling Team, 2015).

Geochemical survey data are typically represented and interpreted using single element geochemical maps. The interpretability and validity of these single components have repeatedly been challenged because they are prone to several artefacts: spurious negative bias on correlations (Chayes, 1960), dependence of interpretation on other (potentially non-reported) components (Aitchison, 1986), dependence on units (e.g. mass, molar), and dependence on processes acting on some components (e.g. weathering, dilution) but influencing all of them (van den Boogaart and Tolosana-Delgado, 2013). All the issues mentioned are due to the fact that geochemical data constitute amounts of components with relative portions of a total even if this total is unknown. The components may be reported in different physical units (ppm, mg/kg or as percentages) and all the components may not be reported or measured. However, each component has an amount which represents its importance as part of the whole composition. The constraints of constant sum or the closed nature of the relative amounts of components have implications for the analysis of geochemical data. In statistics and mathematical geosciences, powerful solutions to deal with these issues have been developed in a field known as Compositional Data Analysis (CoDA) (Aitchison, 1986; Grunsky, 2010; Pawlowsky-Glahn and Buccianti, 2011; van den Boogaart and Tolosana-Delgado, 2013; Templ et al., 2011). This paper investigates the question of what the compositional nature of geochemical data means in regional geochemical mapping, specifically with regard to single element (univariate) distribution maps.

The collaborative research presented stems from the first GeoMap Workshop (held in Olomouc, Czech Republic, 17–20 June 2014) that discussed the consequences of these challenges and the usefulness of CoDA for regional geochemistry. Present were representatives from regional geochemical surveys (Tellus Survey covering Northern Ireland and Tellus Border Survey covering the border counties of the Republic of Ireland, Young and Donald, 2013), the GEMAS project covering a large part of Europe (Reimann et al., 2014a, 2014b), the soil geochemical survey of the conterminous USA (Smith et al., 2011; Drew et al., 2010) and the continental scale National Geochemical Survey of Australia (de Caritat and Cooper, 2011a, 2011b). This paper, whilst acknowledging the historically important role of single component maps, aims to: (1) discuss their appropriateness, (2) provide some examples to highlight the problems raised above, and (3) offer some solutions to present interpretable maps free of the issues arising from the compositional nature of geochemical data.

2. Problems related to single component maps

Geochemical surveys generate datasets with several tens of components (between 50 and up to 70 elements are commonly reported), obtained from different sample materials (soil horizons, size fractions, vegetal tissues, sediments, water, etc.) and with different analytical techniques (total analyses, partial or selective digests). For regional geochemistry, the key applications of the data are generally either to produce and use elemental concentration maps (i.e. one-component regional distribution maps) or to explore associations between elements affected by geological/geochemical processes, which can also be mapped (e.g. principal components). For the first of these tasks, standard practice has included producing a single component map thought to represent the raw or "absolute" input data in the form of dot (or point) maps, but also as interpolated maps of these raw concentrations. Reimann (2005) defines the purpose of such geochemical maps as *"to display different processes in a map form and to detect local deviations from the dominant process in any one sub-area"*. The problems discussed in this section result from the closure property of geochemical compositions, i.e. the unavoidable fact that samples are *aliquots* of the geological bodies we want to investigate and therefore do not really convey information about the element mass distributions (Aitchison, 1986). This has implications for baseline quality mapping, mineral prospecting, land and soil resource assessment or risk evaluation, though these issues are beyond the scope of this contribution. Nevertheless, in the following we show that the traditional meaning of closure effect (linked to closed data with unit, or any other fixed sum constraint of components, and the resulting distortion of the correlation structure) should be considered also in a broader sense. Namely, compositional data are primarily observations that contain quantitatively expressed relative contributions of parts on a whole. From this perspective, the unit sum constraint is just a representation, obtained without altering the source information, conveyed by ratios between the components. Therefore, even with a variable sum of geochemical concentrations, resulting, e.g., from designed omitting of some components, one should be aware that the relative nature of data is still present and needs to be taken into account by proper statistical processing, as exemplified below by CoDA (Pawlowsky-Glahn et al., 2015).

2.1. Point maps — the data "as is"

It is often thought that raw one-component maps report "what is there", that they report a sort of "objective ground truth". However, Reimann (2005) and others (Reimann and Filzmoser, 2000; Reimann and Garrett, 2005; Reimann et al., 2005; Reimann et al., 2008) highlight that to give sense to that set of spatially dispersed values, space-dependent geochemical processes must be interpreted, and that these are highlighted with a proper representation. Reimann (2005) discusses the advantages and disadvantages of such different representations. Reimann (2005) concludes that, actually, the most important issue becomes the scaling chosen to define the points (colour, size and symbol). Splitting the data into groups (classes) on the basis of order statistics in exploratory data analysis (such as the quartiles of boxplots, or other percentiles), Reimann suggests, may shed light on the spatial structure that reflects at least a number of these processes in a map. Fig. 1 shows how impactful this choice of symbol/colour scaling can be with a real dataset. Obviously, equidistant colour scales do not necessarily yield the most easily interpretable maps, neither in raw (or "absolute") nor in logarithmic values. The proportion of the entire study area actually measured by the dataset is largely exaggerated by the size of the dots. The conclusion is that it may be more appropriate to use quantile-based intervals (following the guidelines on scaling as discussed in Reimann, 2005) to present geochemical data to give an initial assessment of the distribution of elemental concentrations. Readers are referred to Reimann (2005) for further details on methods used to establish intervals in data scaling. It should be pointed out that this choice of scaling does not address the problems resulting from the relative character of elemental concentrations, an aspect that will be explored later.

2.2. Geospatial continuity — the interpolated map

It is becoming increasingly common for geochemical atlases (online and published printed versions; Reimann et al., 2003, 2014a, 2014b;