IDENTIFICATION OF MINERALIZATION IN GEOCHEMISTRY BASED ON THE SPATIAL CURVATURE OF LOG-RATIOS

D. Mikšová, C. Rieser, P. Filzmoser
TU Wien
Vienna, AUSTRIA
e-mail: dominika.miksova@tuwien.ac.at

Abstract

Detecting subcropping mineralizations but also deeply buried mineralizations is one important goal in geochemical exploration. The identification of useful indicators for mineralization is a difficult task as mineralization might be influenced by many factors, such as location, investigated media, depth, etc. We propose a statistical method which indicates chemical elements related to mineralization. The identification is based on GAM models for the element concentrations across the spatial coordinate(s). The log-ratios of the GAM fits are taken to compute the curvature, where high curvature is supposed to indicate mineralization. By defining a measure for the quantification of high curvature, the log-ratios can be ranked, and elements can be identified that are indicative of the anomaly patterns.

Keywords: data science, spatial curvature, mineralization

1 Introduction

Identifying geochemical processes as mineralization is defined as the presence of higher concentrations of particular chemical elements compared to the background concentration. In other words, one would expect a rapid spatial change in the concentration on top of the mineralization, depending on the type and extent of the mineralization. Data coming from geochemistry are naturally compositional data, which are strictly non-negative values, forming parts of a whole. Therefore, using log-ratios seems to be an appropriate approach for constructing meaningful features that indicate mineralization. The important information is reflected in the ratios between the variables rather than in the absolute values. Relative information might lead to a proper understanding of the data.

The problem of identifying mineralization is of major interest in the project UpDeep [2], where TU Wien is the project partner responsible for developing statistical methods. The exploration techniques might then even lead to suggestions considering ore discoveries. In this project several data sets are available, and for the development and illustration of the method we use a geochemical data set originating from Finland. The data have been sampled along a linear transect, and concentrations of various chemical elements have been measured in different sample media.

The presented method is based upon the behavior of the curvature of log-ratios. A big (absolute) value of the curvature corresponds to a rapid change of the log-ratio in the area of interest, and this may indicate mineralization. In order to suppress

the effect of measurement uncertainties, we use as a first step GAM models (see, e.g., [4], [5]) to smooth the absolute concentrations ensuring sufficient smoothness. Based on the curvature of the log-ratio of the smoothed concentrations, we then employ an unsupervised learning method leading to a hitlist of log-ratios most suitable for finding mineralization.

The proposed method has been tested on the mentioned real data set, where the mineralizations are even known, and the results seem to be reliable and promising.

2 Methodology

Smoothing splines, as developed by [3], are nowadays an indispensable tool in the modern days statistician's toolbox. They have been used with great success in a variety of areas and continue to this day to be a very active field of research.

Usually as a starting point one considers, for given data $(x_1, y_1), \ldots, (x_n, y_n)$, the following non-parametric model with Gaussian i.i.d. errors,

$$y_i(x_i) = f(x_i) + \epsilon_i, \tag{1}$$

$\epsilon_i \sim \mathcal{N}(0, \sigma^2)$, with $i = 1, \ldots, n$. The presumably smooth linear predictor f is estimated by solving the problem given by

$$\max_f \sum_{i=1}^{n} \omega_i l(y_i|x_i; f) - \lambda \int (f''(x))^2 dx, \tag{2}$$

where $f(x) = \mathbf{h}(x)^t \boldsymbol{\beta}$ for $\mathbf{h}$ being a B-spline basis, and l stands for log-likelihood of the Gamma distribution, which is appropriate in this context since we deal with positive concentrations of elements. Note, ω_i represent predefined weights enforcing the higher concentrations. The part $\lambda \int (f''(x))^2 dx$ in (2) is important for controlling the smoothness via the number of basis functions. The parameter λ is the smoothing parameter controlling the trade-off between fitting the data closely and having a smooth model, and finally f'' is the second derivative of f.

Using log-ratios of the obtained fitted values from the GAM model for a pair of variables, we calculate the absolute curvature, denoted as $\kappa(x)$. A big value of the curvature indicates clear peaks in the log-ratio.

A further step is to define a measure to rank the log-ratios according to their curvature. For this purpose, we define a threshold as mean plus standard deviation of the curvature, marked as $\tau := \mu + \sigma$. This allows to determine the number J of separated regions, obtained as half of number of points N crossing a threshold of its curvature, as well as the length I_J of the interval of each region. A measure for the quantification of high curvature of particular log-ratio can be then defined as

$$c = \frac{1}{J} \sum_{j=1}^{J} \max_{x \in I_J} (\kappa(x) - \mathcal{T})_+^2, \tag{3}$$

where $\max_{x \in I_J}(\kappa(x) - \mathcal{T})_+^2$ is the highest distance of curvature subtracted from threshold for Jth interval. The log-ratios of all pairs of variables can now be ranked according to this value c, and log-ratios on top of the hitlist indicate the locations of potential mineralization.

In the presentation we will show results from our data set, and also results from other geochemical data sets. All these results indicate that the method indeed is able to identify pathfinder elements for mineralization. Note that our proposed approach in unsupervised – thus it is not necessary to know the locations of the potential mineralization.

References

[1] Friedman J., Hastie T., Tibshirani R. (2001). *The Elements of Statistical Learning*. Springer Series in Statistics, New York.

[2] KAVA Reference: 16329, UpDeep, Upscaling deep buried geochemical exploration techniques into European business (2017–2020).

[3] Reinsch C.H. (1967). Smoothing by spline functions. *Numerische Mathematik*. Vol.**10**, pp. 177-183.

[4] Wood S.N. (2017). *Generalized Additive Models: An Introduction With R*. Chapman and Hall/CRC, Boca Raton, USA.

[5] Yee T.W. (2015). *Vector Generalized Linear and Additive Models: With an Implementation in R*. Springer, New York.