# SPATIAL FACTOR ANALYSIS OF STREAM SEDIMENT GEOCHEMISTRY DATA FROM SOUTH GREENLAND

Allan A. Nielsen[1], Knut Conradsen[1], John L. Pedersen[2] and Agnete Steenfelt[3]

[1]Department of Mathematical Modelling, Technical University of Denmark, B-321, DK-2800 Lyngby, Denmark. aa,kc@imm.dtu.dk
[2]Nunaoil A/S, Pilestræde 52, DK-1112 København K, Denmark. jlp@nunaoil.dk
[3]GEUS, The Geological Survey of Denmark and Greenland, Thoravej 8, DK-2400 København NV, Denmark. ast@geus.dk

## SUMMARY

This paper describes the application of min/max autocorrelation factor (MAF) analysis to irregularly sampled stream geochemical data from South Greenland. Kriged MAF images are compared with kriged images based on varimax rotated factors (VRF) from an ordinary non-spatial factor analysis, and the image patterns are discussed in a geological context. It is demonstrated that MAF analysis has a potential for objectively dividing multi-element data into spatial segments coinciding with large lithotectonic units.

## 1. INTRODUCTION

A spatial extension to principal components (PC) and factor analysis termed minimum/maximum autocorrelation factor (MAF) analysis is described in the literature for multivariate data sampled on a regular grid, [10], [5]. Other references deal with spatial factor analysis based on parameterisations of observed correlations of irregularly sampled data, [6], [7]. In this contribution the MAF analysis is extended to irregularly sampled data, see also [9]. The technique is applied to stream geochemical data from South Greenland, and results are compared with varimax rotated factors (VRFs).

## 2. DATA BASE

The data are derived from chemical analyses of 2,100 stream sediment samples collected in local drainage basins over approximately 10,000 km$^2$ in South Greenland by the Geological Survey of Greenland [3], [11]. Two analytical techniques giving total concentrations have been used. The elements Ca, Cu, Fe, Ga, K, Mn, Nb, Ni, Pb, Rb, Sr, Ti, Y, Zn and Zr have been determined by energy-dispersive, isotope excited X-ray fluorescence, while Au, Ag, As, B, Br, Co, Cr, Cs, Hf, Mo, Na, Sb, Sc, Se, Ta, Th, U, W, La, Ce, Nd, Sm, Eu, Tb, Yb and Lu have been determined by instrumental neutron activation analysis. Thus there are 41 variables in the dataset.

## 3. GEOLOGICAL SETTING

The study area (see [1]) comprises a Palaeoproterozoic orogen, which consists of three major tectono-stratigraphic units: a northern border zone of tectonically reworked Archaean gneissic basement, a central zone occupied by a Proterozoic granite batholith

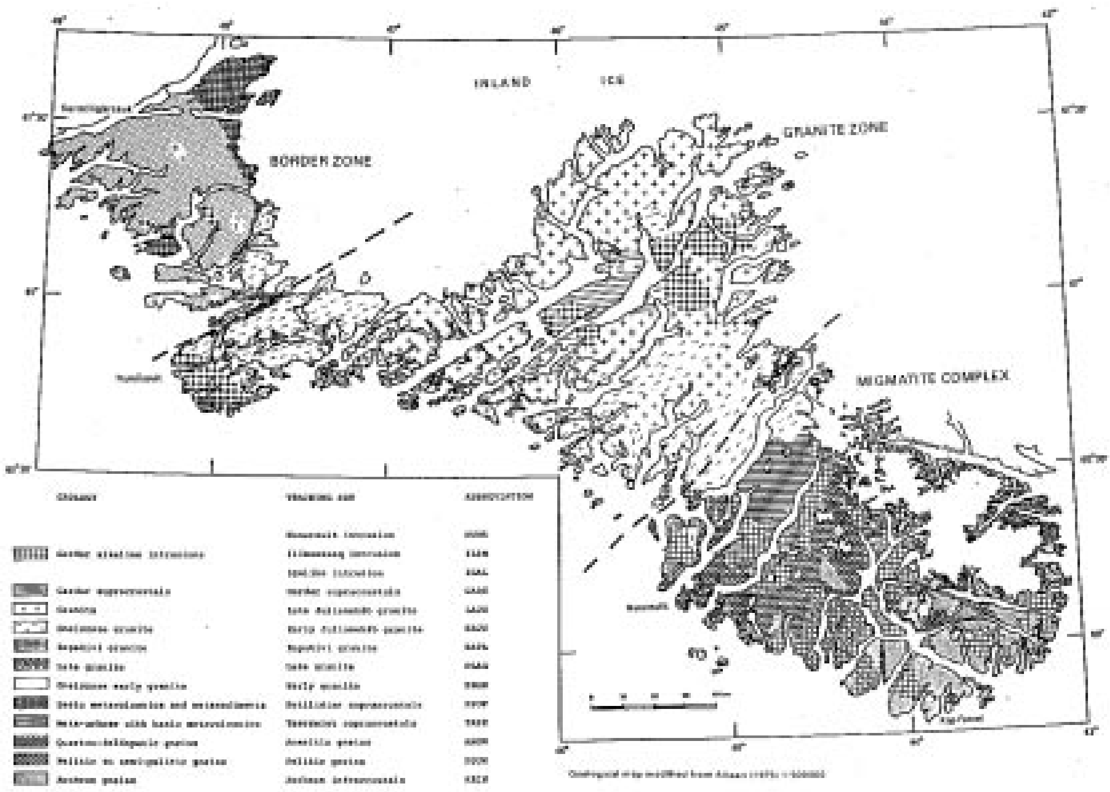


Figure 1: Geological map of South Greenland

complex and a southern migmatite complex of predominantly Proterozoic metasediments and metavolcanics intruded by rapakivi type granites (Figure 1). In Mesoproterozoic times the boundary region between the Archaean and batholith was subjected to rifting and intrusions of numerous dykes of basaltic to trachytic compositions as well as of felsic alkaline complexes including carbonatites.

## 4. STATISTICAL TECHNIQUES

The popular principal components (PC) analysis transforms a multivariate variable into new variables that are mutually orthogonal. The first PC, PC1, is the linear combination of the original variables that explains maximal variance in all the original variables. Higher order PCs explain maximal variance subject to the orthogonality. Factor analysis is a common name for a family of multivariate techniques. One of the simpler is principal factor analysis. Mathematically, principal factors can be thought of as scaled PCs. A result of this scaling is that the factors can be rotated, e.g. to obtain easy interpretability. The so-called varimax rotation criterion aims at obtaining correlations between original variables and factors that are close to –1, 0 and 1. Most good textbooks on multivariate statistics give descriptions of PC and factor analysis, see e.g. [2].

As opposed to PC and factor analysis the min/max autocorrelation factor (MAF) transformation allows for the spatial nature of the image data. The MAF transform minimises the autocorrelation rather than maximising the data variance. The first MAF, MAF1, is the linear combination of the original bands that contains minimum autocorrelation between neighbouring pixels. A higher order MAF is the linear combination of the original bands that contains minimum autocorrelation subject to the constraint that it is orthogonal to lower order MAFs. MAF analysis thus constitutes a (conceptually) more satisfactory way of orthogonalising image data than PC analysis. An important property of the MAF procedure is its invariance to linear transformations, a property not shared by ordinary PC analysis. This means that it doesn't matter whether the data have been scaled e.g. to unit variance before the analysis is performed. Min/max autocorrelation factor analysis was suggested by Switzer and Green in [10]. A good and easily obtained reference is [5].

Let us consider the random variable $\boldsymbol{Z} = [Z_1(\boldsymbol{x}), \ldots, Z_m(\boldsymbol{x})]^T$ and without loss of generality we assume that $\mathrm{E}\{\boldsymbol{Z}(\boldsymbol{x})\} = \boldsymbol{0}$ and $\mathrm{D}\{\boldsymbol{Z}(\boldsymbol{x})\} = \boldsymbol{\Sigma}$. We denote a spatial shift by $\boldsymbol{\Delta} = [\Delta_x, \Delta_y]^T$. The spatial covariance function is defined by

$$\mathrm{Cov}\{\boldsymbol{Z}(\boldsymbol{x}), \boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})\} = \boldsymbol{\Gamma}(\boldsymbol{\Delta}).$$

$\boldsymbol{\Gamma}$ has the following properties: i) $\boldsymbol{\Gamma}(\boldsymbol{0}) = \boldsymbol{\Sigma}$, and ii) $\boldsymbol{\Gamma}(\boldsymbol{\Delta})^T = \boldsymbol{\Gamma}(-\boldsymbol{\Delta})$. We are interested in the correlations between projections of the variables and the shifted variables. Therefore we find

$$\begin{aligned}\mathrm{Cov}\{\boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}), \boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})\} &= \boldsymbol{a}^T\boldsymbol{\Gamma}(\boldsymbol{\Delta})\boldsymbol{a} = \boldsymbol{a}^T\boldsymbol{\Gamma}(\boldsymbol{\Delta})^T\boldsymbol{a} \\ &= \boldsymbol{a}^T\boldsymbol{\Gamma}(-\boldsymbol{\Delta})\boldsymbol{a} = \frac{1}{2}\boldsymbol{a}^T\left(\boldsymbol{\Gamma}(\boldsymbol{\Delta}) + \boldsymbol{\Gamma}(-\boldsymbol{\Delta})\right)\boldsymbol{a}.\end{aligned}$$

Introducing

$$\boldsymbol{\Sigma}_\Delta = \mathrm{D}\{\boldsymbol{Z}(\boldsymbol{x}) - \boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})\} = \mathrm{E}\{[\boldsymbol{Z}(\boldsymbol{x}) - \boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})][\boldsymbol{Z}(\boldsymbol{x}) - \boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})]^T\},$$

which considered as a function of $\boldsymbol{\Delta}$ is a multivariate variogram, we have

$$\boldsymbol{\Gamma}(\boldsymbol{\Delta}) + \boldsymbol{\Gamma}(-\boldsymbol{\Delta}) = 2\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_\Delta$$

and thus for the autocorrelation

$$\mathrm{Corr}\{\boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}), \boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})\} = 1 - \frac{1}{2}\frac{\boldsymbol{a}^T\boldsymbol{\Sigma}_\Delta\boldsymbol{a}}{\boldsymbol{a}^T\boldsymbol{\Sigma}\boldsymbol{a}}.$$

If we want to minimise that correlation we must maximise the Rayleigh coefficient

$$R(\boldsymbol{a}) = \frac{\boldsymbol{a}^T\boldsymbol{\Sigma}_\Delta\boldsymbol{a}}{\boldsymbol{a}^T\boldsymbol{\Sigma}\boldsymbol{a}}.$$

Let $R_1 \geq \cdots \geq R_m$ be the eigenvalues and $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m$ the corresponding conjugate eigenvectors of $\boldsymbol{\Sigma}_\Delta$ with respect to $\boldsymbol{\Sigma}$. Then $\boldsymbol{Y}_i(\boldsymbol{x}) = \boldsymbol{a}_i^T\boldsymbol{Z}_i(\boldsymbol{x})$ is the i'th MAF. The reverse numbering of MAFs so that the signal MAF is referred to as MAF1 is used below.

For regularly spaced data the differencing to obtain $\boldsymbol{\Sigma}_\Delta$ can be done by combining horizontal and vertical shifts. For irregularly spaced data the differencing can be done by simply using the nearest neighbour only. For irregularly spaced data more elaborate noise models such as residuals from fits to local surfaces can be obtained by means of the Voronoi tessellation and its dual concept, the Delaunay triangulation, [9]. MAFs defined in this fashion can be altered to allow for other neighbourhoods for instance confined by distance and/or direction constraints.

Figure 2: Correlations between original variables and the first two VRFs (top) and MAFs (bottom)

## 5. RESULTS

Varimax rotated factors (VRF) and MAFs are linear combinations of the original variables in this case element concentrations. To facilitate interpretation Figure 2 shows correlations between the original variables and the first two VRFs and MAFs.

The signal MAFs by design have high autocorrelation. In general the semivariogram of the MAFs will exhibit decreasing range of influence and increasing nugget effect. Thus the signal MAFs are well suited for interpolation and this characteristic inspires a new form of kriging, namely maximum autocorrelation factorial kriging. To obtain kriged versions of the original data the inverse MAF transformation can be applied. For reproduction reasons Figure 3 shows greyscale images of kriged VRF 1-3 and MAF 1-3 instead of the more informative RGB plots of the same quantities.

## 6. GEOLOGICAL INTERPRETATION

It is a general observation in geochemical mapping using stream sediments that element distribution patterns are often very stable despite the fact that the chemical compositions of individual samples are much influenced by local conditions. This is particularly well displayed in images produced by kriging of single element or multivariate data although the choice of method greatly influences the resulting image as shown here. It is

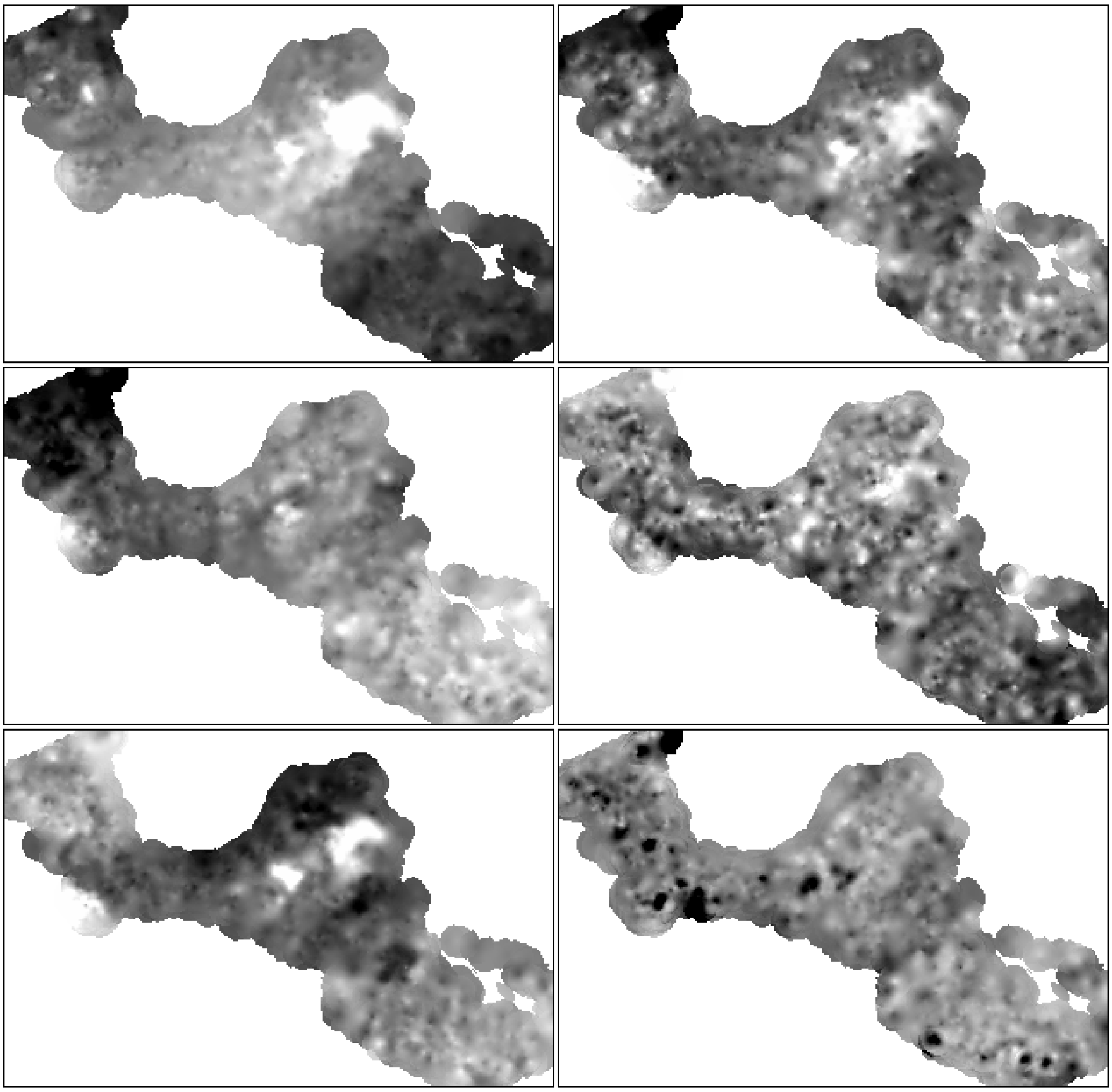

Figure 3: Kriged MAFs (left column) and VRFs (right column) 1, 2 and 3 of geochemical data

immediately observed (Figure 3) that the MAF images are less influenced by short range variations and appears to depict large scale geochemical features, whereas the VRF images reflect more localised features, many of which are difficult to relate to lithological features.

The most obvious feature depicted by MAF1, MAF3 and VRF1 as white areas reflects the Mesoproterozoic alkaline intrusive complexes. These complexes have a very distinct chemistry with high concentrations of lithophile and rare earth elements (compare Figure 2). There are minor chemical differences between the individual complexes which are seen when MAF1 is compared with MAF3.

The clear large scale spatial division of the data by the MAF images into three subprovinces was unexpected and very important for the future use of this method. The southeastern province is imaged as a fairly homogeneous unit coinciding with the mapped migmatite complex, see Figure 1. As this unit is lithologically diverse, it comprises rocks of both sedimentary, volcanic and intrusive origin, it was not expected to posses a geochemical uniformity (see MAF1) or to deviate much as a unit from the Border zone to the north-west (see MAF2).

REFERENCES


1. ALLAART, J.H. - "Ketilidian mobile belt in South Greenland". In Escher, A. and Watt, W.S. (Eds.) "*Geology of Greenland*". The Geological Survey of Greenland, Copenhagen. 120-151. 1976.

2. ANDERSON, T.W. - "*An Introduction to Statistical Analysis*" (second edition). John Wiley, New York. 675 pp. 1984.

3. ARMOUR-BROWN, A., STEENFELT, A. and KUNZENDORF, H. - "Uranium districts defined by reconnaissance geochemistry in South Greenland". *J. of Geochem. Explor.* **19**, 127-145. 1993.

4. CHADWICK, B. and GARDE, A.A. - "Palaeoproterozoic oblique plate convergence in South Greenland: a reappraisal of the Ketilidian Orogen". In Brewer, T.S. (Ed.) "*Precambrian Crustal Evolution in the North Atlantic Region*". Geological Society Special Publication No. 112, 179-196. 1996.

5. GREEN, A.A., BERMAN, M., SWITZER, P. and CRAIG, M. - "A transformation for ordering multispectral data in terms of image quality with implications for noise removal". *IEEE Transactions on Geoscience and Remote Sensing*, **26**(1), 65-74. 1988.

6. GRUNSKY, E.C. and AGTERBERG, F.P. - "Spatial and multivariate analysis of geochemical data from metavolcanic rocks in the Ben Nevis area, Ontario". *Mathematical Geology*, **20**(7), 825-861. 1988.

7. GRUNSKY, E.C. and AGTERBERG, F.P. - "SPFAC: a Fortran program for spatial factor analysis of multivariate data". *Computers & Geoscience*, **17**(1), 133-160. 1991.

8. JOURNEL, A.G. and HUIJBREGTS, C.J. - "*Mining Geostatistics*". Academic Press, London. 600 pp. 1978.

9. NIELSEN, A.A. - "*Analysis of Regularly and Irregularly Sampled Spatial, Multivariate, and Multi-temporal Data*". Department of Mathematical Modelling, Technical University of Denmark, Ph.D. thesis. 213 pp. 1994.

10. SWITZER, P. and GREEN, A.A. - "Min/max autocorrelation factors for multivariate spatial imagery". Tech. rep. 6, Department of Statistics, Stanford University. 1984.

11. THORNING, L., TUKIAINEN, T. and STEENFELT, A. - "Regional compilations of geoscience data from the Kap Farvel - Ivittuut area, South Greenland". Thematic Map Series. Geological Survey of Greenland, Copenhagen. 94/1, 27 pp. 1994.

12. WINDLEY, B.F. - "Early Proterozoic collision tectonics, and rapakivi granites as intrusions in an extensional thrust-thickened crust: the Ketilidian orogen, South Greenland". *Tectonophysics* **195**, 1-10. 1991.