MAXIMUM AUTOCORRELATION FACTORIAL KRIGING

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

[1]IMM, Department of Mathematical Modelling, Technical University of Denmark, B-321
DK-2800 Lyngby, Denmark, aa@imm.dtu.dk
[2]JLP Minerals, Traneholmen 17, DK-3460 Birkerød, Denmark, johnlp.geo@post.tele.dk
[3]GEUS, The Geological Survey of Denmark and Greenland, Thoravej 8
DK-2400 København NV, Denmark, ast@geus.dk

ABSTRACT

This paper describes maximum autocorrelation factor (MAF) analysis, maximum autocorrelation factorial kriging, and its application to irregularly sampled stream sediment geochemical data from South Greenland. Kriged MAF images are compared with kriged images of varimax rotated factors from an ordinary non-spatial factor analysis, and they are interpreted in a geological context. It is demonstrated that MAF analysis contrary to ordinary non-spatial factor analysis gives an objective discrimination of large-scale as well as small-scale geological features which are related to crustal structure and the occurrence of specific rock types.

INTRODUCTION

A spatial extension to principal components and factor analysis termed maximum autocorrelation factor (MAF) analysis is described in the literature for multivariate data sampled on a regular grid, [16, 5, 10]. Other references deal with spatial factor analysis based on parameterisations of observed correlations of irregularly sampled data, [6, 7]. In this paper the MAF analysis is extended to irregularly sampled data, see also [11, 12, 14], and maximum autocorrelation factorial kriging is introduced. The technique is applied to stream sediment geochemical data from South Greenland, and results from maximum autocorrelation factorial kriging are compared with kriged varimax rotated principal factors.

DATA

In 1979-80 the GGU, the Geological Survey of Greenland (now GEUS, the Geological Survey of Denmark and Greenland), collected stream sediment samples from a 10,000 km^2 area in South Greenland. Sample sites were small active streams with catchment areas of 1-10 km^2. Samples were sieved at 100 mesh and the undersize was analysed. The present study is based on a dataset with 41 variables and 2,097 samples. Two analytical techniques have been used. The concentrations of 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 and the concentrations of Au, Ag, As, Ba, 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. These analyses of the samples are not identical to the ones used in the case reported in [11, 12]. Statistical analysis is done on natural logarithms of the element concentrations with values below the detection limit simulated from a triangular distribution from zero to the detection limit.

GEOLOGICAL SETTING

The study area is underlain by a Palaeoproterozoic orogen, the Ketilidian orogen, which consists of three major tectono-stratigraphic units: (1) a northern Border zone of tectonically reworked Archaean gneissic basement overlain by Palaeoproterozoic metasediments and metavolcanics in the north-east, (2) a central zone occupied by a calc-alkaline granitic batholith, and (3) a southern migmatite complex of predominantly Palaeoproterozoic metasediments and metavolcanics intruded by post-tectonic rapakivi

WJ Kleingeld and DG Krige (eds), Geostats 2000 Cape Town, Copyright 2000
Document Transformation Technologies. Printed in South Africa.

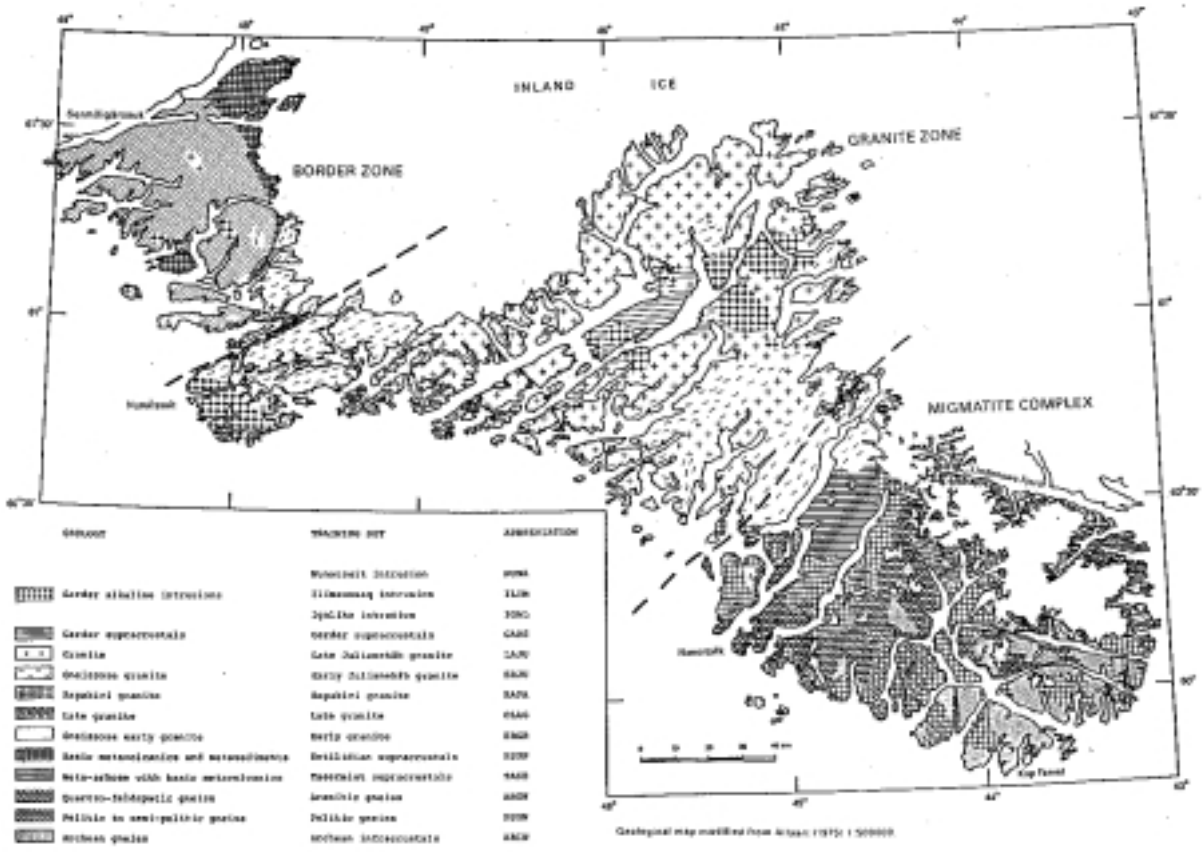

Figure 1: Simplified geological map of South Greenland

type granites, see Figure 1 and [1]. The plate-tectonic setting of the orogen has recently been interpreted in [3]. In Mesoproterozoic times the boundary region between the border and the granite zones was subjected to rifting and intrusions of numerous dykes of basaltic to trachytic compositions as well as of felsic alkaline complexes including carbonatites. The region affected by the alkaline magmas is termed the Gardar province, [18].

STATISTICAL TECHNIQUES

Orthogonal transformations

The popular principal component (PC) analysis transforms a multivariate variable into new variables that are mutually orthogonal. The first PC, PC 1, is the linear combination of the (zero mean) 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 forms is principal factor analysis. Mathematically, principal factors can be thought of as scaled PCs. The factors can be rotated for instance 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 or 1. Most good textbooks on multivariate statistics give descriptions of PC and factor analysis, see for instance [2].

As opposed to PC and factor analysis the maximum autocorrelation factor (MAF) transformation allows for the spatial nature of image data. The first MAF, MAF 1, is the linear combination of the original variables that contains maximum autocorrelation between neighbouring observations. A higher order MAF is the linear combination of the original variables that contains maximum autocorrelation subject to the constraint that it is orthogonal to lower order MAFs. MAF analysis thus constitutes a (conceptually)

more satisfactory way of orthogonalising spatial data than PC and factor 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 for example to unit variance—as in factor analysis—before the analysis is performed.

Let us consider the multivariate random variable $\boldsymbol{Z}^T = [Z_1(\boldsymbol{x}), \ldots, Z_m(\boldsymbol{x})]$ with expectation $\mathrm{E}\{\boldsymbol{Z}(\boldsymbol{x})\} = \boldsymbol{0}$ and dispersion or covariance matrix $\mathrm{D}\{\boldsymbol{Z}(\boldsymbol{x})\} = \boldsymbol{\Sigma}$. We denote a spatial shift by $\boldsymbol{\Delta}^T = [\Delta_x, \Delta_y]$. The spatial covariance function is defined by

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

We are interested in the correlations between projections of the variables and the shifted variables. Therefore we find for the autocovariance

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

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 get

$$\frac{1}{2}(\boldsymbol{\Gamma}(\boldsymbol{\Delta}) + \boldsymbol{\Gamma}(\boldsymbol{\Delta})^T) = \boldsymbol{\Sigma} - \frac{1}{2}\boldsymbol{\Sigma}_\Delta$$

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 maximise the Rayleigh coefficient

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

we minimise the autocorrelation. Let $\lambda_1 \leq \cdots \leq \lambda_m$ 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 ith MAF.

Maximum autocorrelation factor analysis was suggested by Switzer and Green in [16]. Other references to MAF, related transformations and applications are [4, 5, 10, 11, 13].

Maximum autocorrelation factorial kriging

For regularly spaced data the differencing to obtain $\boldsymbol{\Sigma}_\Delta$ can be performed by combining horizontal and vertical shifts. For irregularly spaced data the differencing can be done by using the nearest neighbour only irrespective of distance and direction. This is done below. More elaborate noise models such as residuals from local mean or median filters, or residuals from fits to local surfaces can be obtained by means of the Voronoi tessellation and its dual concept the Delaunay triangulation, [15]. To each point in the plane we associate a Voronoi polygon which is the part of the plane that is nearer to that point than to any other point. From the Voronoi tessellation we can construct the Delaunay triangulation by joining points with common Voronoi polygon edges. Figure 2 shows a Delaunay triangulation and the dual Voronoi tessellation of the same point set. The definition of MAFs for irregularly sampled data can be modified to allow for other neighbourhoods for instance confined by distance and/or direction constraints.

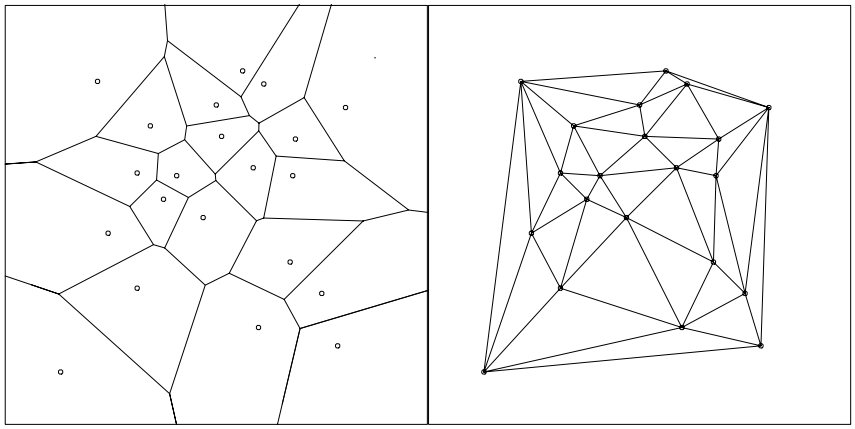

Figure 2: Voronoi tessellation (left), Delaunay triangulation (right)

Varimax rotated factors (VRFs) and MAFs are linear combinations of the (zero mean) original variables, in this case natural logarithms of element concentrations. To facilitate interpretation Figure 3 shows correlations between the original variables and the first three VRFs (all 41 factors retained) with amounts of variance explained and the first three MAFs with associated autocorrelations.

The signal MAFs by design have high autocorrelation. In general the semivariograms of the MAFs exhibit decreasing range of influence and increasing nugget effect with increasing order. This is clearly seen in Figures 4 to 6 which show 21×21 5 km lag 2-D semivariograms of the (standardised) original variables, the VRFs (all 41 factors retained) and the MAFs, all ordered row-wise, and 1 km lag 1-D isotropic semivariograms of the (first 40) MAFs. The elements in Figure 4 top-right are ordered row-wise as in Figure 3: Au, Ag, As, Ba, Br, Ca, Co, Cr, Cs, Cu, Fe, Ga, Hf, K, Mn, Mo, Na, Nb, Ni, Pb, Rb, Sb, Sc, Se, Sr, Ta, Th, Ti, U, W, Y, Zn, Zr, La, Ce, Nd, Sm, Eu, Tb, Yb and Lu. Figure 4 top-left shows autocorrelations associated with the MAFs. The ordering of the 2-D semivariograms shows that the VRFs do not exhibit the desired semivariogram characteristics of the MAFs.

Because of their 1-D and 2-D semivariogram characteristics the signal MAFs are well suited for interpolation. This ability inspires us to perform maximum autocorrelation factorial kriging which is kriging (minimum estimation variance prediction, see for example [9, 8, 11]) of the (hopefully few) signal MAFs. To obtain kriged versions of all the original variables the inverse MAF transformation applies. Figures 7 and 8 show RGB images of kriged VRFs 1-3 and MAFs 1-3.

GEOLOGICAL INTERPRETATION

In stream sediment geochemical mapping it is generally observed that element distribution patterns are very stable despite the fact that the chemical compositions of individual samples are much influenced by local conditions. It is also observed that the chemistry of a stream sediment reflects that of the surrounding rocks. For South Greenland this is demonstrated in single element maps in [17], where certain element distributions are seen to display known lithological units well. Thus the Gardar province is mapped by high Nb, the Archaean by high Cu, Mg and Ni, the batholith by low Sc, the migmatite complex by high As, and the rapakivi granites by high Hf and Zr.

While single element maps depict local lithological units, the multivariate maps shown in Figure 7 and especially in Figure 8 enable the recognition of large-scale geochemical provinces. This is shown particularly well by the low order MAF images which are less influenced by short range variations than the corresponding VRF images. VRF images, on the other hand, may be used to highlight local multielement anomalies which may be of interest in localising mineral occurrences.

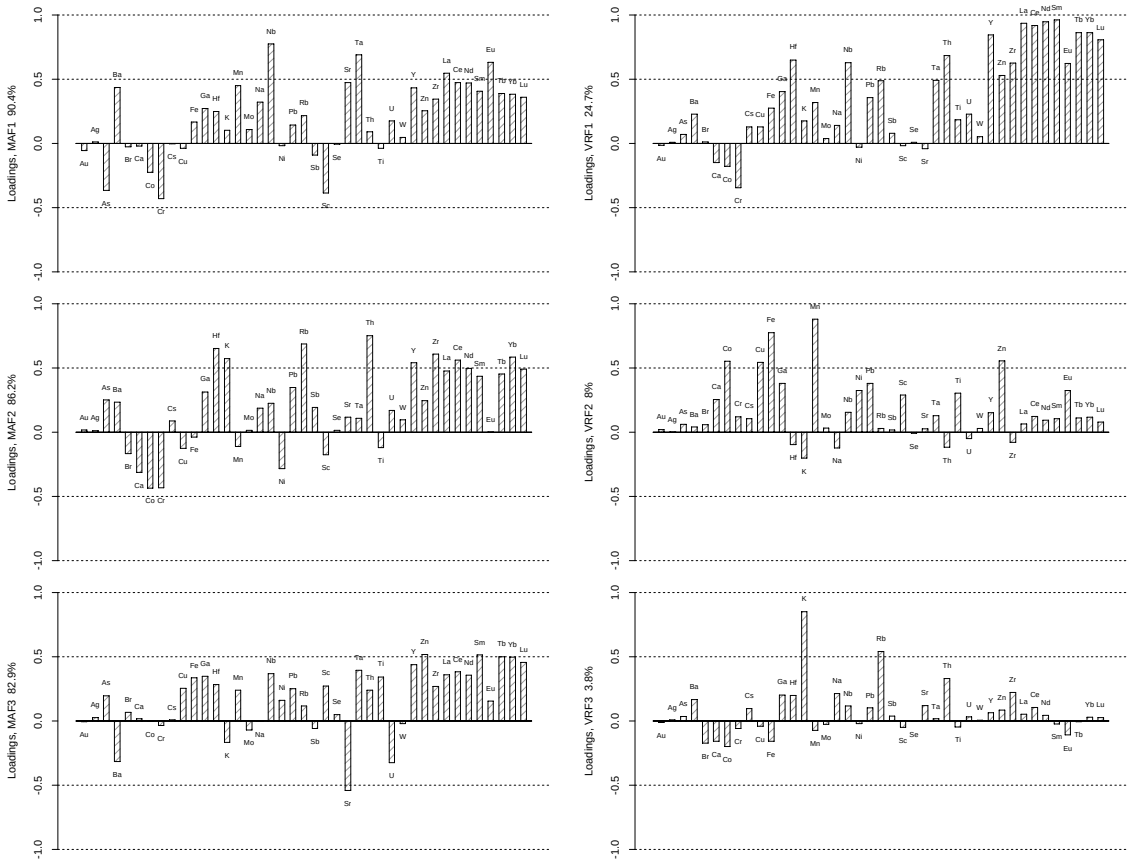

Figure 3: Correlations between original variables and the first three MAFs (left) and VRFs (right)

Table 1: Scores of MAFs 1-3 in main lithotectonic units

	Archean	Batholith	Migmatite	Gardar
MAF 1	Medium	Medium	Low	High
MAF 2	Low	Medium	High	Medium
MAF 3	High	Low	Medium	High

The main features of the MAF images in relation to the main lithotectonic units are listed in Table 1.

The most prominent geochemical feature of South Greenland is the Gardar province, and this is well displayed by white areas in the map of MAF 1. Figure 3 shows that the province is geochemically anomalous in a long range of elements and most notably in Nb and Ta. MAF 2, governed by high Hf, K, Rb, Th, Y and Zr, highlights the province of rapakivi granites in the migmatite complex and emphasises the difference between these granites and those of the batholith. MAF 3 is governed by much the same element association as MAF 1, though high Cu, Fe, Ga and Zn are more dominant in MAF 3 with the result that the mafic rocks of the Archaean and Palaeoproterozoic parts of the Border zone and of the migmatite complex show up in light shades, and individual alkaline complexes within the Gardar province are enhanced.

When maps of MAFs 1, 2 and 3 are combined in a colour ternary image (not shown) the major provinces of South Greenland are particularly well discriminated. This demonstrates the ability of the MAF analysis to identify the most important geochemical boundaries in a region. Such boundaries are often difficult or impossible to discern in single element maps because the distribution patterns for different elements have slightly different boundaries.

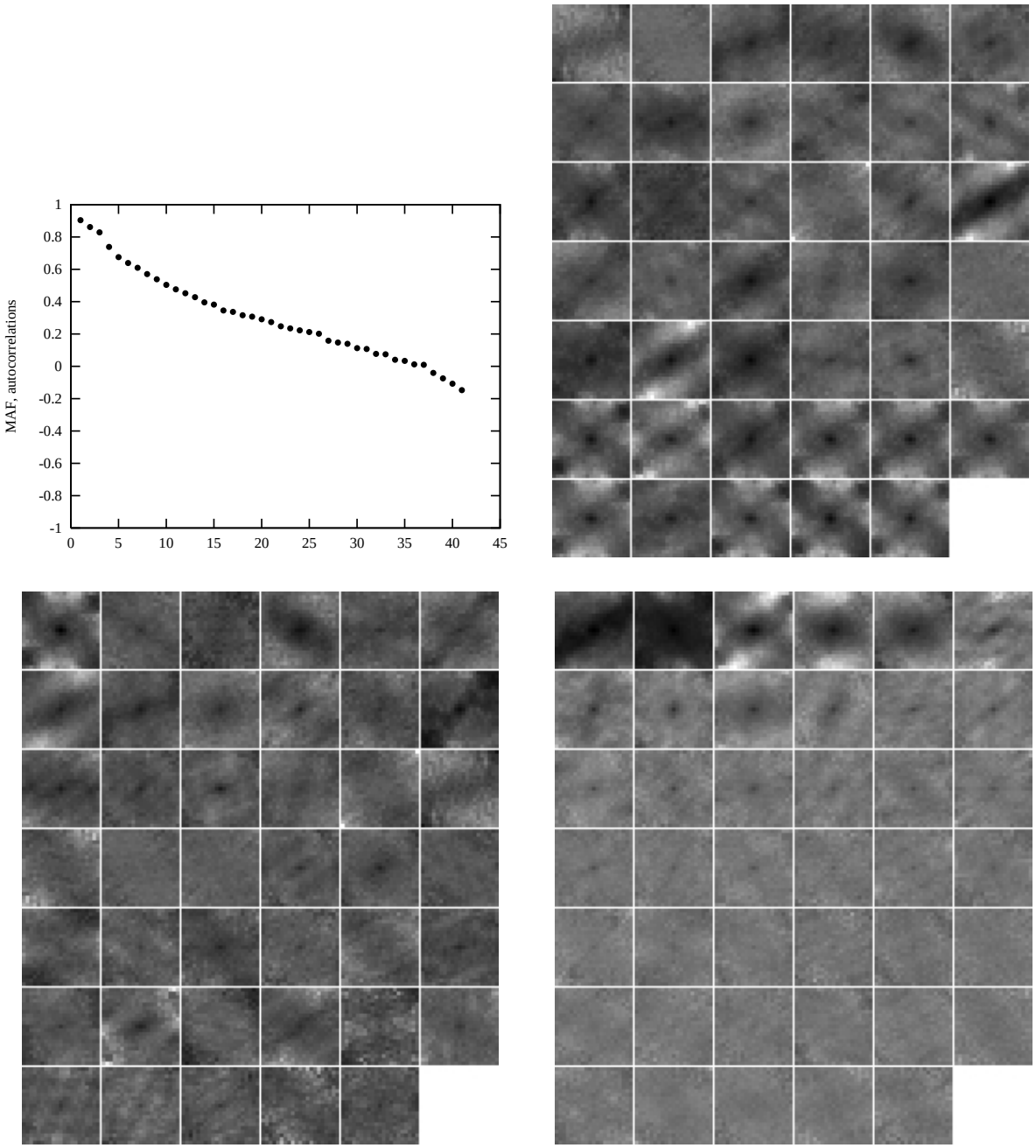

Figure 4: Autocorrelations for MAFs (top-left), 2-D semivariograms of standardised original data (top-right), varimax rotated factors (bottom-left), and maximum autocorrelation factors (bottom-right)

Many maps of higher order MAFs display short range features some of which may be related to certain lithological units, while others are difficult to relate to the known lithology. However, there are cases where high order MAFs reveal geochemical features which may be interpreted to reflect other geological phenomena like hydrothermal mineralisation or alteration.

The MAF analysis offers a range of possibilities which have only been tested to a limited degree. For instance, analysis can be performed on any element or spatial subset of the data. In the South Greenland case analyses could be carried out, for example, on rock forming elements and mineralisation related elements separately.

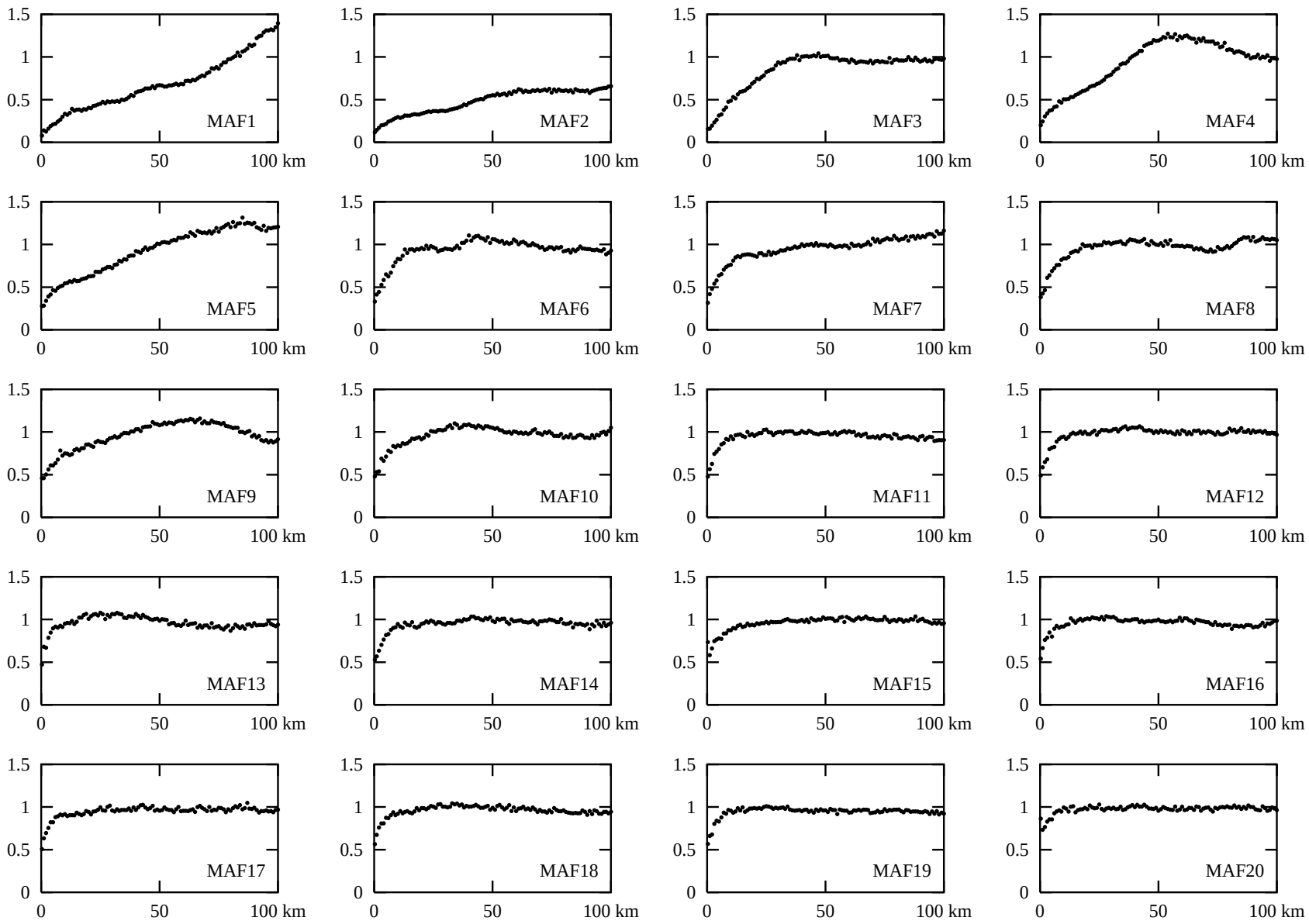

Figure 5: Isotropical semivariograms of maximum autocorrelation factors

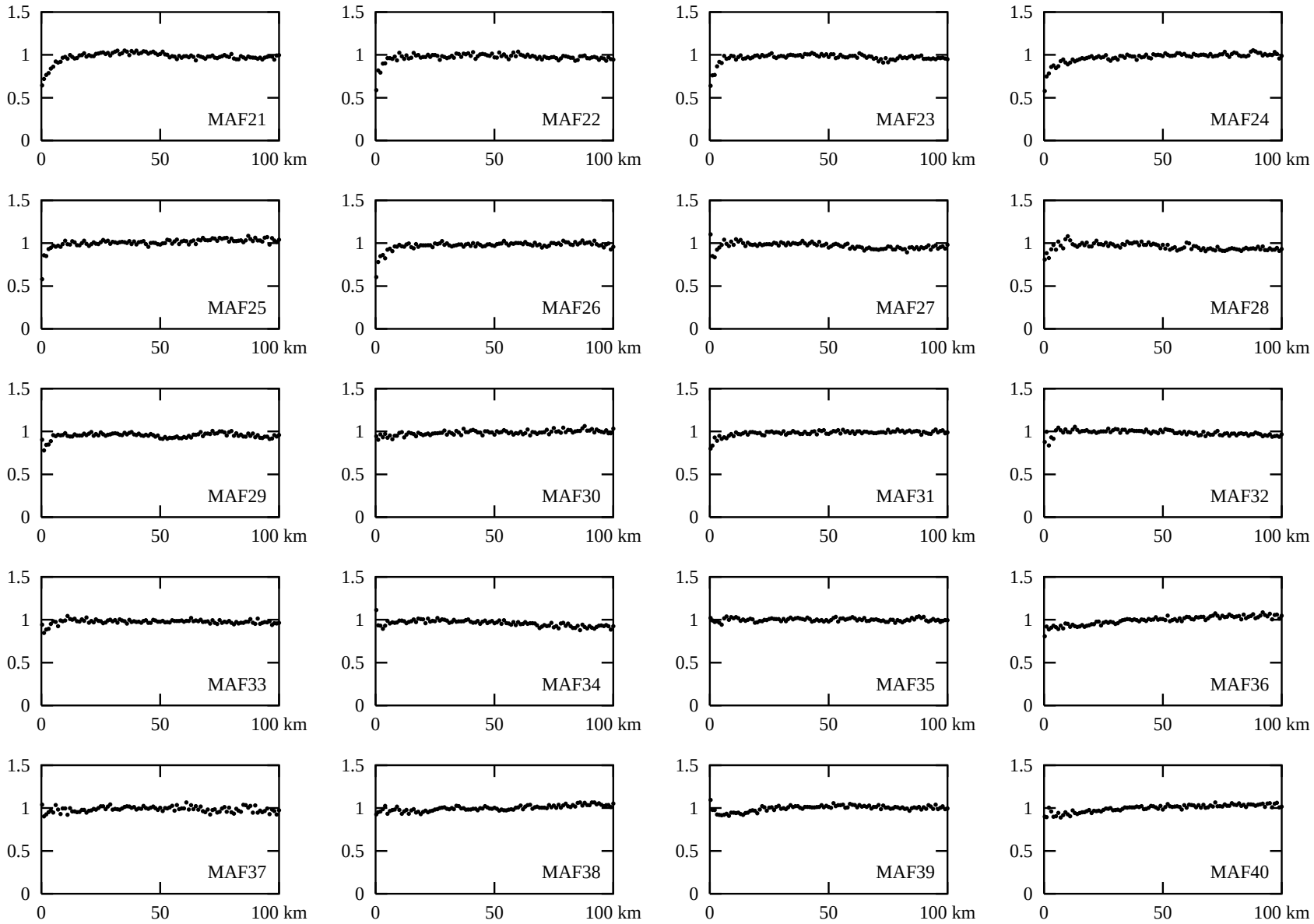

Figure 6: Isotropical semivariograms of maximum autocorrelation factors

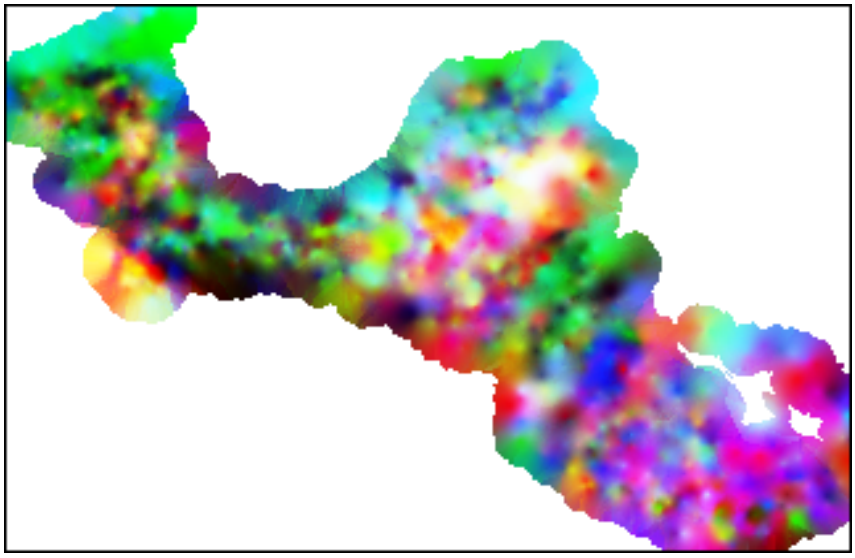

Figure 7: Kriged varimax rotated principal factors 1, 2 and 3 as RGB

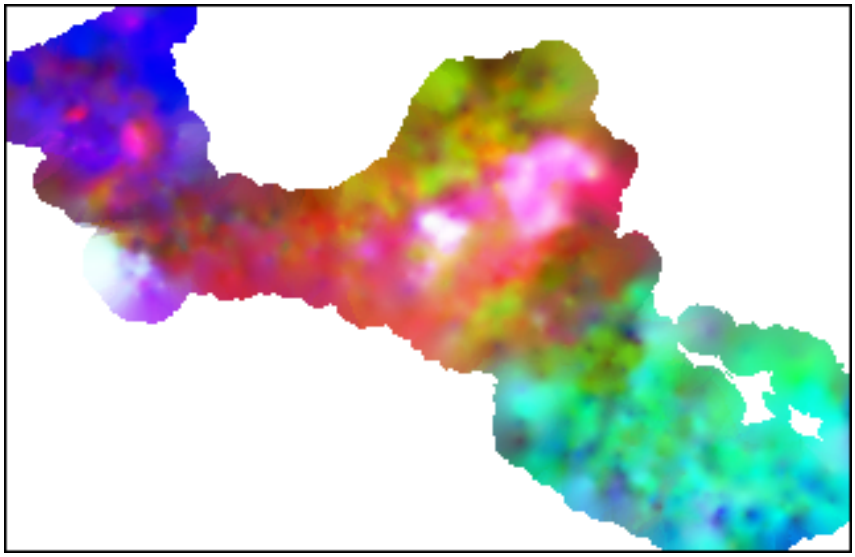

Figure 8: Kriged maximum autocorrelation factors 1, 2 and 3 as RGB

CONCLUSIONS

Maximum autocorrelation factor (MAF) analysis performs a decomposition of the original variables into new orthogonal variables which by design exhibit characteristics ranging from large-scale, low-frequency features to small-scale, high-frequency features with increasing order of the MAFs. In the South Greenland case with analysis of stream sediment data we obtain an efficient and un-biased segmentation into geochemical provinces and identification of significant boundaries within the region. Such information is valuable to the understanding of the crustal stucture of an area and may be particularly useful in plate-tectonic reconstructions. MAF analysis is superior to ordinary non-spatial factor analysis in displaying large-scale features. The multivariate approach reveals data relations which cannot be recognised by studying single element maps.

Producing maps of kriged MAFs is a good way of condensing large amounts of geochemical data for presentation. The combination of such maps as ternary images has proved to be particularly useful in presenting the essential geological features of the region.

ACKNOWLEDGMENTS

Dr. Rasmus Larsen, Dr. Karsten Hartelius, Dr. Kristian Windfeld and Henrik J. Hansen wrote some of the computer programs used. AAN was partly funded by the European Union under contract numbers MA2M-CT90-0010 and BRE2-CT92-0201.

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, 675 pp., 1984.

[3] 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.

[4] Conradsen, K., Nielsen, B.K. and Thyrsted, T. "A comparison of min/max autocorrelation factor analysis and ordinary factor analysis". In *Proceedings from Symposium in Applied Statistics*, Technical University of Denmark, Lyngby, Denmark, 47-56, 1985.

[5] Green, A.A., Berman, M., Switzer, P. and Craig, M.D. "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] Isaaks, E.H. and Srivastava, R.M. "*An Introduction to Applied Geostatistics*". Oxford University Press, 561 pp., 1989.

[9] Journel, A.G. and Huijbregts, C.J. "*Mining Geostatistics*". Academic Press, 600 pp., 1978.

[10] Lee, J.B., Woodyatt, A.S. and Berman, M. "Enhancement of high spectral resolution remote-sensing data by a noise-adjusted principal components transform". *IEEE Transactions on Geoscience and Remote Sensing*, **28**(3), 295-304, 1990.

[11] Nielsen, A.A. "*Analysis of Regularly and Irregularly Sampled Spatial, Multivariate, and Multitemporal Data*". Ph.D. thesis, Department of Mathematical Modelling, Technical University of Denmark, 213 pp., 1994. http://www.imm.dtu.dk/~aa/phd/.

[12] Nielsen, A.A., Conradsen, K., Pedersen, J.L. and Steenfelt, A. "Spatial factor analysis of stream sediment geochemistry from South Greenland". In Pawlowsky-Glahn, V. (Ed.) *Proceedings of the International Association of Mathemetical Geology, IAMG'97*, Barcelona, Spain, 955-960, 1997.

[13] Nielsen, A.A., Conradsen, K. and Simpson, J.J. "Multivariate alteration detection (MAD) and MAF post-processing in multispectral, bi-temporal image data: new approaches to change detection studies". *Remote Sensing of Environment*, **19**, 1-19, 1998.

[14] Nielsen, A.A., Ersbøll, B.K., Pälchen, W. and Rank, G. "Spatial analysis of multivariate, (ir)regularly sampled data: geochemistry from the Eastern Erzgebirge". In Baafi, E.Y. and Schofield, N.A. (Eds.) *Geostatistics Wollongong '96: Quantitative Geology and Geostatistics*, Wollongong, New South Wales, Australia, 1173-1184, Kluwer Academic Publishers, 1996.

[15] Ripley, B.D. "*Spatial Statistics*". John Wiley & Sons, 252 pp., 1981.

[16] Switzer, P. and Green, A.A. "Min/max autocorrelation factors for multivariate spatial imagery". Tech. rep. 6, Department of Statistics, Stanford University, 1984.

[17] Thorning, L., Tukiainen, T. and Steenfelt, A. (Eds.). "*Regional compilations of geoscience data from the Kap Farvel-Ivittuut area, South Greenland*". The Geological Survey of Greenland, Thematic Map Series, 1994.

[18] Upton, B.G.J. and Emeleus, C.H. "Mid-Proterozoic alkaline magmatism in southern Greenland: Gardar province". In Fitton, F.G. and Upton, B.G.J. (Eds.) *Alkaline Igneous Rocks*. Geological Society Special Publication No. 30, 449-471, 1987.

Allan Aasbjerg Nielsen is an Associate Research Professor with the Section for Image Analysis, Department of Mathematical Modelling, Technical University of Denmark. M.Sc. in 1978 from the Department of Electrophysics, Technical University of Denmark. Ph.D. in 1994 from the Department of Mathematical Modelling, Technical University of Denmark. With the Danish Defense Research Establishment 1977-1978. Work on energy conservation in housing with the Thermal Insulation Laboratory, Technical University of Denmark, from 1978 to 1985. Since 1985 with the Section for Image Analysis, Department of Mathematical Modelling, Technical University of Denmark. Since then work on several national and international projects on application of statistical methods and remote sensing in mineral exploration, mapping, geology, environment and agriculture funded by industry, the European Union, Danida (the Danish International Development Agency), and the Danish Research Councils. Very interested and experienced in development and implementation of mathematical and statistical methods concerning spatial, multi-/hyperspectral and multi-temporal data, and applications hereof. National point of contact (NPOC) for concerted actions on advanced methods in remote sensing funded by the European Union. Member of host, program and scientific committees for international conferences. Referee for international journals and conferences. Homepage http://www.imm.dtu.dk/~aa/.

Knut Conradsen is a Full Professor in statistical image analysis at the Department of Mathematical Modelling, Technical University of Denmark and the Deputy Rector of the Technical University of Denmark. Main research interest is the application of statistics and statistical models to real-life problems. Work on air pollution monitoring (research which produced standards that are presently used by government agencies in their approval of the location of polluting industries) and remote sensing in mineral exploration (Greenland, Spain, Germany), in agricultural monitoring (Kenya), and in environmental monitoring (Spain, Øresund) developing new techniques in classification, in texture analysis, and in orthogonal transformations of multispectral data. Project manager for several projects funded by the Danish Environmental Protection Agency (on air quality monitoring), by the Danish National Research Councils (on storm surge forecasting, on image analysis and computer vision), by the Commission of European Communities and the European Union (on the application of statistics, remote sensing, and image processing in the raw materials sector, in environmental monitoring, and in industry), and by Danida, the Danish International Development Agency (on agricultural monitoring in Kenya).