KERNEL PARAMETER DEPENDENCE IN SPATIAL FACTOR ANALYSIS

Allan A. Nielsen

Technical University of Denmark, DTU Space – National Space Institute
DK-2800 Kgs. Lyngby, Denmark

1. INTRODUCTION

Principal component analysis (PCA) [1] is often used for general feature generation and linear orthogonalization or compression by dimensionality reduction of correlated multivariate data, see Jolliffe [2] for a comprehensive description of PCA and related techniques. Schölkopf et al. [3] introduce kernel PCA. Shawe-Taylor and Cristianini [4] is an excellent reference for kernel methods in general. Bishop [5] and Press et al. [6] describe kernel methods among many other subjects. The kernel version of PCA handles nonlinearities by implicitly transforming data into high (even infinite) dimensional feature space via the kernel function and then performing a linear analysis in that space.

In this paper we shall apply a kernel version of maximum autocorrelation factor (MAF) [7, 8] analysis to irregularly sampled stream sediment geochemistry data from South Greenland and illustrate the dependence of the kernel width. The 2,097 samples each covering on average 5 km^2 are analyzed chemically for the content of 41 elements.

2. DATA AND PREPROCESSING

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 identical to the ones used in [10] but different from the ones reported in [9].

2.1. 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)

AAN's e-mail is aa@space.dtu.dk, his homepage is http://www.imm.dtu.dk/∼aa.

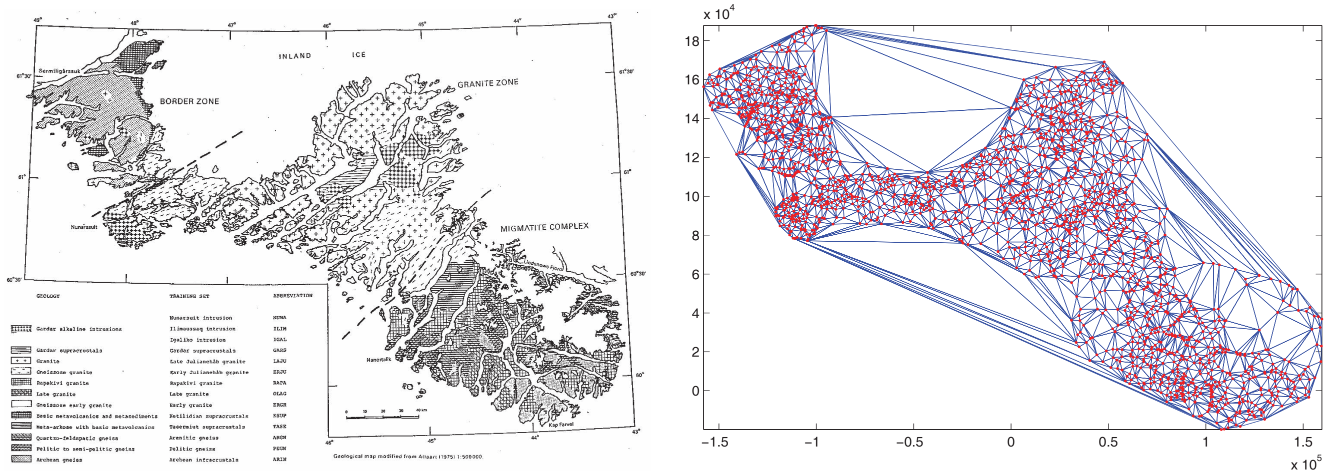

Fig. 1. Simplified geological map of South Greenland (left). All 2,097 sample sites and the Delaunay triangulation (right).

a southern migmatite complex of predominantly Palaeoproterozoic metasediments and metavolcanics intruded by post-tectonic rapakivi type granites, see Figure 1 (left) and [11]. The plate-tectonic setting of the orogen has been interpreted in [12]. 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, [13].

3. KERNEL PCA AND MAF

A kernel formulation of principal component analysis (PCA) [1] may be obtained from Q-mode or dual formulation of the problem combined with the so-called kernel trick [3]. In a similar fashion maximum autocorrelation factor (MAF) analysis [7, 8, 14] which may be considered as a form of spatial factor analysis may be kernelized. In this context a popular kernel is the Gaussian $\kappa(x_i, x_j) = \exp(-\frac{1}{2}(\|x_i - x_j\|/\sigma)^2)$ where the kernel width is given by the scale parameter σ, and x_i and x_j are (here) 41-dimensional vectors of concentrations. Below we give results of the kernel MAF analysis with different choices of σ.

4. RESULTS AND DISCUSSION

Figure 1 (right) shows the 2,097 sample sites in Southern Greenland in red. The study area is approximately 320 km east-west and 210 km north-south. The Delaunay triangulation is shown in blue. The analyses shown below are based on concentrations standardized to unit variance, see also [9, 10, 15].

For σ equal to the mean distance between observations in 41-dimensional feature space kMAFs 1, 2 and 3 in Figure 2 left focus on extreme observations associated with the intrusions marked with dense plus signs "+" in the Granite zone (Figure 1 left). Also they neatly adapt to an even strongly varying multivariate background. Although other samples have high scores, this is true also for kMAFs with σ equal to ten times the mean, Figure 2 right. In spite of a tendency to highlight more samples in the so-called Gardar intrusion, the same overall impression is true for for kMAFs with σ equal to a hundred times the mean,

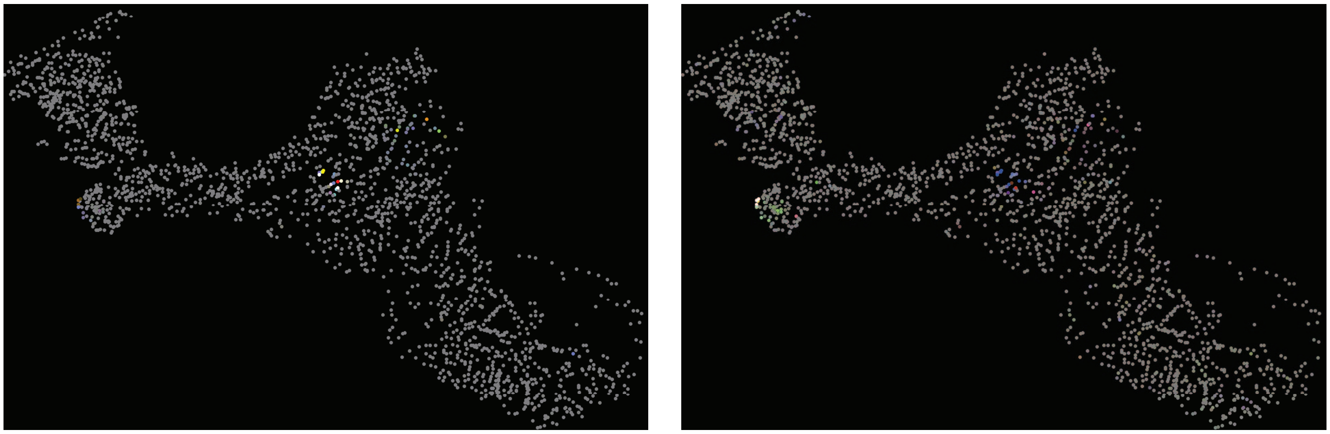

Fig. 2. Kernel MAFs 1, 2 and 3 as RGB, kernel width σ is mean of distances in feature space (left), kernel MAFs 1, 2 3 as RGB, kernel width σ is 10 times mean of distances in feature space (right).

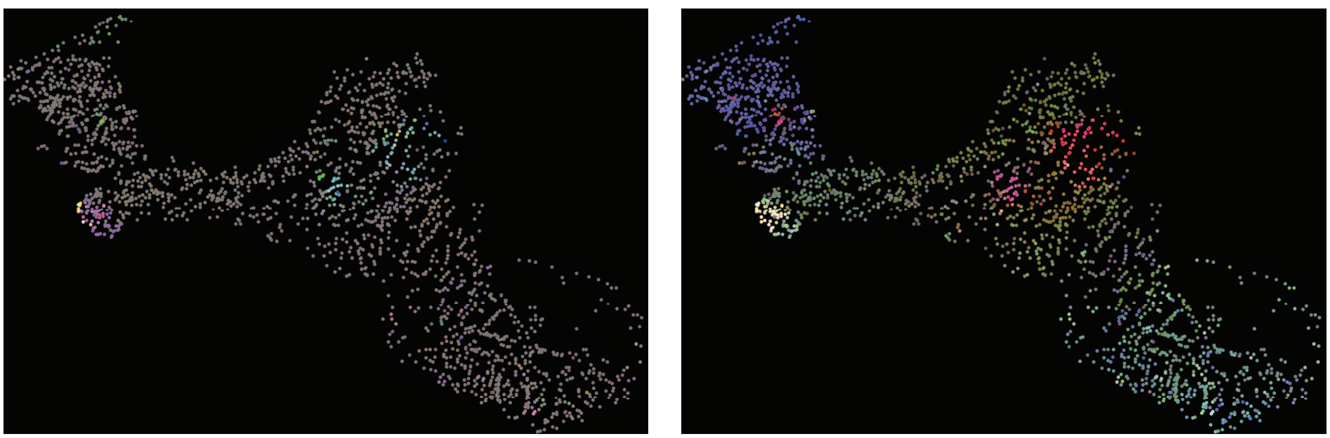

Fig. 3. Kernel MAFs 1, 2 and 3 as RGB, kernel width σ is 100 times mean of distances in feature space (left), kernel MAFs 1, 2 3 as RGB, kernel width σ is 1,000 times mean of distances in feature space (right).

Figure 3 left. For kMAFs with σ equal to a thousand times the mean (Figure 3 right) we see a depiction of the three major geological units named "Border Zone", "Granite Zone" and "Migmatite Complex" in the geological map, Figure 1 left.

In conclusion we see that by varying the kernel width σ we may analyse the phenomenon under study at different scales which highlight different relevant geological features.

5. REFERENCES

[1] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *Journal of Educational Psychology*, vol. 24, pp. 417–441 and 498–520, 1933.

[2] I. T. Jolliffe, *Principal Component Analysis*, second edition, Springer, 2002.

[3] B. Schölkopf, A. Smola, and K.-R. Müller, “Nonlinear component analysis as a kernel eigenvalue problem,” *Neural Computation*, vol. 10, no. 5, pp. 1299–1319, 1998.

[4] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press, 2004.

[5] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[6] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, third edition, Cambridge University Press, 2007.

[7] P. Switzer and A. A. Green, “Min/max autocorrelation factors for multivariate spatial imagery,” Tech. Rep. 6, Stanford University, 1984.

[8] P. Switzer and S. E. Ingebritsen, “Ordering of time-difference data from multispectral imagery,” *Remote Sensing of Environment*, vol. 20, pp. 85–94, 1986.

[9] A. A. Nielsen, K. Conradsen, J. L. Pedersen and A. Steenfelt, “Spatial Factor Analysis of Stream Sediment Geochemistry Data from South Greenland,” In V. Pawlowsky-Glahn (ed.) Proceedings of the Third Annual Conference of the International Association for Mathematical Geology, IAMG’97, pp. 955–960, Barcelona, Spain, 22-27 September 1997, Internet http://www.imm.dtu.dk/pubdb/p.php?5686.

[10] A. A. Nielsen, K. Conradsen, J. L. Pedersen and A. Steenfelt, “Maximum Autocorrelation Factorial Kriging”, In W. J. Kleingeld and D. G. Krige (eds.) Proceedings of the 6th International Geostatistics Congress, Geostats 2000, pp. 538–547, Cape Town, South Africa, 10-14 April 2000, Internet http://www.imm.dtu.dk/pubdb/p.php?3639.

[11] J. H. Allaart, “Ketilidian mobile belt in South Greenland.” In A. Escher and W. S. Watt (Eds.) *Geology of Greenland*. The Geological Survey of Greenland, Copenhagen, pp. 120–151, 1976.

[12] B. Chadwick and A. A. Garde, “Palaeoproterozoic oblique plate convergence in South Greenland: a reappraisal of the Ketilidian Oroge”. In T. S. Brewer (Ed.) *Precambrian Crustal Evolution in the North Atlantic Region*. Geological Society Special Publication No. 112, pp. 179–196, 1996.

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

[14] A. A. Nielsen, “Kernel maximum autocorrelation factor and minimum noise fraction transformations,” Submitted, 2009.

[15] A. A. Nielsen, “A kernel version of spatial factor analysis,” 57th Session of the International Statistical Institute, ISI, Durban, South Africa, 16-22 August 2009. Internet http://www.imm.dtu.dk/pubdb/p.php?5742.