Kernel parameter dependence in spatial factor analysis

Nielsen, Allan Aasbjerg

Published in:
IGARSS

Link to article, DOI:
10.1109/IGARSS.2010.5653545

Publication date:
2010

Document Version
Publisher's PDF, also known as Version of record

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 GEOLOGY

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

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

2.1. Geological Setting

The study area is underlain by a Palaeoproterozoic orogen, the Ketilidian orogen, which consists of three major tectonostratigraphic units: (1) a northern Border zone of tectonically reworked Archaean gneissic basement overlain by Palaeoproterozoic metasediments and metavolcanics in the northeast, (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 type granites, see Figure 1 (top) 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].

Let us consider a data set with n observations and p variables organized as a matrix X with n rows and p columns; each column contains measurements over all observations from one variable and each row consists of a vector of measurements x_i^T from p variables for a particular observation

$$X = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{bmatrix}. \quad (1)$$

The superscript ^{T} denotes the transpose. X is sometimes called the data matrix or the design matrix. Without loss of generality we assume that the variables in the columns of X have mean value zero.

3.1. R-mode PCA

In ordinary (primal also known as R-mode) PCA we analyze the variance-covariance matrix $S = X^T X/(n-1) = 1/(n-$

$1) \sum_{i=1}^{n} x_i x_i^T$ which is p by p. If $X^T X$ is full rank $r = \min(n, p)$ this will lead to r non-zero eigenvalues λ_i and r orthogonal or mutually conjugate unit length eigenvectors u_i ($u_i^T u_i = 1$) from the eigenvalue problem

$$\frac{1}{n-1} X^T X u_i = \lambda_i u_i. \quad (2)$$

We see that the sign of u_i is arbitrary. To find the principal component scores for an observation x we project x onto the eigenvectors, $x^T u_i$. The variance of these scores is $u_i^T S u_i = \lambda_i u_i^T u_i = \lambda_i$ which is maximized by solving the eigenvalue problem.

3.2. Q-mode PCA

In the dual formulation (also known as Q-mode analysis) we analyze $XX^T/(n-1)$ which is n by n and which may be very large. XX^T is called the Gram[1] matrix and its elements are the inner products $x_i^T x_j$ between the rows of the data matrix X. Multiply both sides of Equation 2 from the left with X

$$\frac{1}{n-1} XX^T (Xu_i) = \lambda_i (Xu_i) \quad (3)$$

or

$$\frac{1}{n-1} XX^T v_i = \lambda_i v_i \quad (4)$$

with v_i proportional to Xu_i, $v_i \propto Xu_i$, which is normally not unit length if u_i is. Now multiply both sides of Equation 4 from the left with X^T

$$\frac{1}{n-1} X^T X (X^T v_i) = \lambda_i (X^T v_i) \quad (5)$$

to show that $u_i \propto X^T v_i$ is an eigenvector of S with eigenvalue λ_i. We scale these eigenvectors to unit length assuming that v_i are unit vectors ($1 = v_i^T v_i \propto u_i^T X^T X u_i = (n-1)\lambda_i u_i^T u_i = 1$)

$$u_i = \frac{1}{\sqrt{(n-1)\lambda_i}} X^T v_i. \quad (6)$$

We see that if $X^T X$ is full rank $r = \min(n, p)$, $X^T X/(n-1)$ and $XX^T/(n-1)$ have the same r non-zero eigenvalues λ_i and that their eigenvectors are related by $u_i = X^T v_i/\sqrt{(n-1)\lambda_i}$ and $v_i = Xu_i/\sqrt{(n-1)\lambda_i}$.

3.3. Kernel Formulation of PCA

We now replace x by $\phi(x)$ which maps x nonlinearly into a typically higher dimensional feature space. As an example consider a two-dimensional vector $[z_1 \ z_2]^T$ being mapped into $[z_1 \ z_2 \ z_1^2 \ z_2^2 \ z_1 z_2]^T$. This maps the original two-dimensional vector into a five-dimensional feature space so that for example a linear decision rule becomes general enough to differentiate between all linear and quadratic forms including ellipsoids.

The mapping by ϕ takes X into Φ which is an n by q ($q \geq p$) matrix

$$\Phi = \begin{bmatrix} \phi(x_1)^T \\ \phi(x_2)^T \\ \vdots \\ \phi(x_n)^T \end{bmatrix}. \quad (7)$$

For the moment we assume that the mappings in the columns of Φ have zero mean. In this higher dimensional feature space $C = \Phi^T \Phi/(n-1) = 1/(n-1) \sum_{i=1}^{n} \phi(x_i)\phi(x_i)^T$ is the variance-covariance matrix and for PCA we get the primal formulation

$$\frac{1}{n-1} \Phi^T \Phi u_i = \lambda_i u_i \quad (8)$$

where we have re-used the symbols λ_i and u_i from above.

For the corresponding dual formulation we get

$$\frac{1}{n-1} \Phi \Phi^T v_i = \lambda_i v_i \quad (9)$$

where we have re-used the symbol v_i from above. As above the non-zero eigenvalues for the primal and the dual formulations are the same and the eigenvectors are related by

$$u_i = \frac{1}{\sqrt{(n-1)\lambda_i}} \Phi^T v_i \quad (10)$$

and $v_i = \Phi u_i/\sqrt{(n-1)\lambda_i}$.

3.3.1. Kernel Substitution

Applying kernel substitution also known as the kernel trick we replace the inner products $\phi(x_i)^T \phi(x_j)$ in $\Phi\Phi^T$ with a kernel function $\kappa(x_i, x_j) = \kappa_{ij}$ which could have come from some unspecified mapping ϕ. In this way we avoid the explicit mapping ϕ of the original variables. We obtain

$$Kv_i = (n-1)\lambda_i v_i \quad (11)$$

where $K = \Phi\Phi^T$ is an n by n matrix with elements $\kappa(x_i, x_j)$. K is symmetric and must be positive semi-definite, i.e., its eigenvalues are non-negative; we say that κ is a Mercer kernel. Normally we let the eigenvalues subsume the factor $n-1$

$$Kv_i = \lambda_i v_i. \quad (12)$$

In this case $u_i = \Phi^T v_i/\sqrt{\lambda_i}$ and $v_i = \Phi u_i/\sqrt{\lambda_i}$.

It is easy to show that both centering to zero mean of the mappings in the columns of Φ as well as the projection of observations x onto the primal eigenvectors u_i may be expressed by means of the kernel function $\kappa(x_i, x_j)$ without explicit use of the nonlinear mapping.

[1] named after Danish mathematician Jørgen Pedersen Gram (1850-1916)

3.4. Kernel MAF

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, for details see [15]. 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 (here) are 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 (bottom) 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 [10, 9, 16].

For σ equal to the mean distance between observations in 41-dimensional feature space kMAFs 1, 2 and 3 in Figure 2 top focus on extreme observations associated with the intrusions marked with dense plus signs "+" in the Granite zone (Figure 1 top). 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 bottom. In spite of a tendency to highlight more samples in the so-called Gardar intrusion, the same overall impression is true for kMAFs with σ equal to a hundred times the mean, Figure 3 top. For kMAFs with σ equal to a thousand times the mean (Figure 3 bottom) we see a depiction of the three major geological units named "Border Zone", "Granite Zone" and "Migmatite Complex" in the geological map, Figure 1 top.

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.

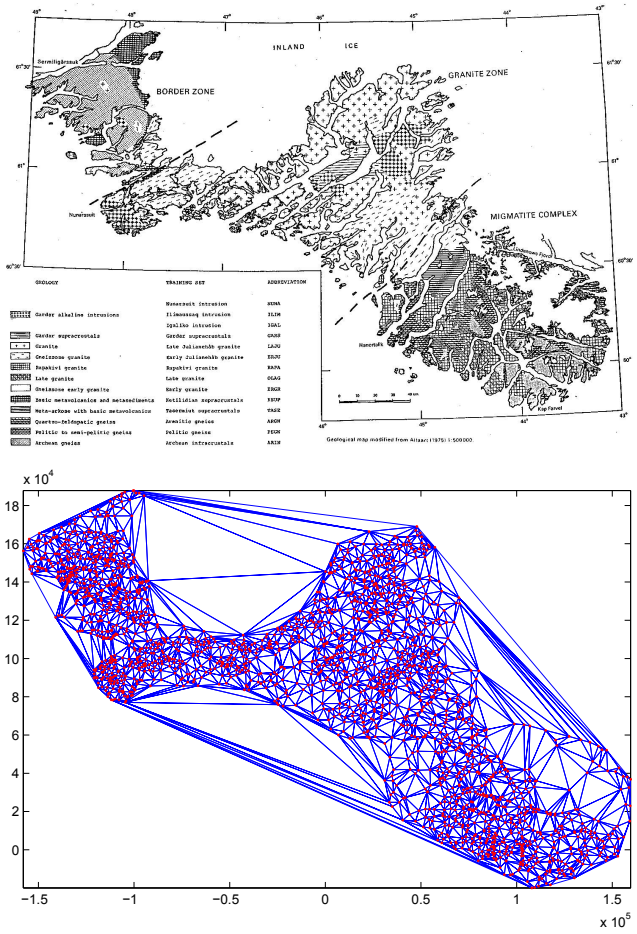

Fig. 1. Simplified geological map of South Greenland (top). All 2,097 sample sites and the Delaunay triangulation (bottom).

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, "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.

[10] A. A. Nielsen, K. Conradsen, J. L. Pedersen and A. Steenfelt, "Spatial Factor Analysis of Stream Sediment Geochemistry

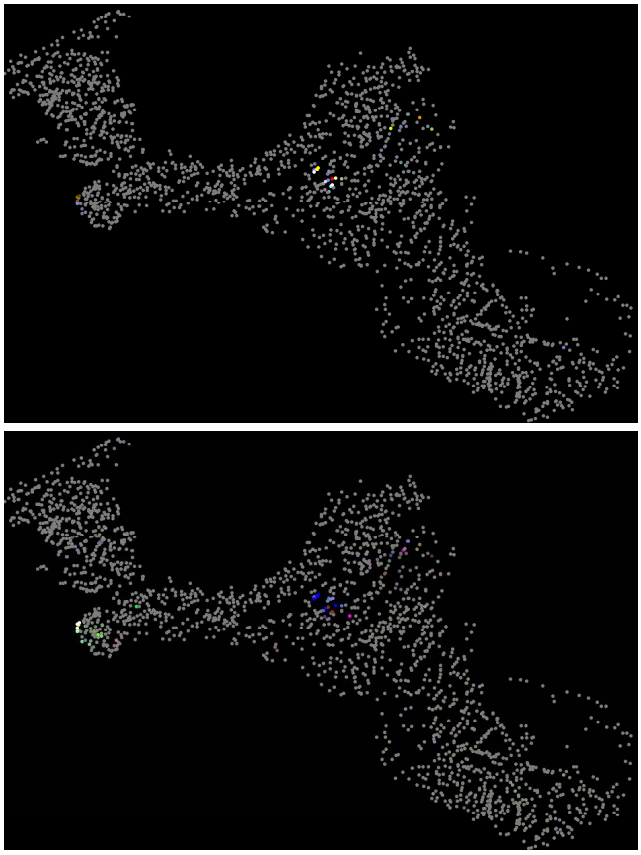

Fig. 2. Kernel MAFs 1, 2 and 3 as RGB, kernel width σ is mean of distances in feature space (top), kernel MAFs 1, 2 3 as RGB, kernel width σ is 10 times mean of distances in feature space (bottom).

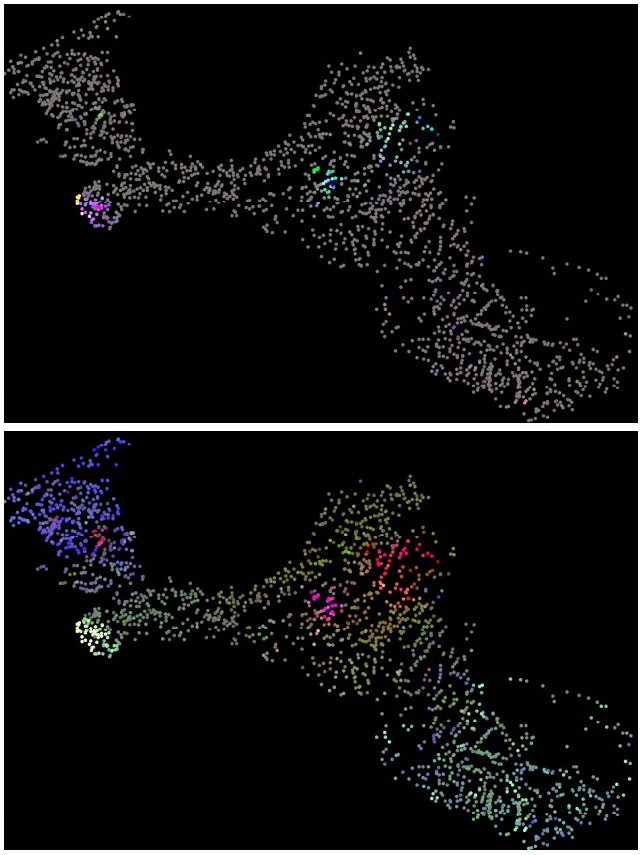

Fig. 3. Kernel MAFs 1, 2 and 3 as RGB, kernel width σ is 100 times mean of distances in feature space (top), kernel MAFs 1, 2 3 as RGB, kernel width σ is 1,000 times mean of distances in feature space (bottom).

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.

[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 Orogen". 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: the Gardar province". 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, A. Hecheltjen, F. Thonfeld, and M. J. Canty, "Automatic change detection in RapidEye data using the combined MAD and kernel MAF methods," IGARSS, 2010.

[15] A. A. Nielsen, "Kernel maximum autocorrelation factor and minimum noise fraction transformations," submitted, 2010.

[16] 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.