Implementation of the Min/Max Autocorrelation Factors and Application to a Real Data Example

S. Elogne and O. Leuangthong

The implementation in a GSLIB version of the minimum maximum autocorrelation factor (MAF) technique is considered in this note. This note reviews in a practical way the technique by describing the main lines of the approach. An academic application of the method has been conducted on a real data example.

Introduction

This note implements and describes the minimum maximum autocorrelation factor which is based on the principal component analysis that spatially brings correlated variables into non-correlated factors at any lag h. Principal Component Analysis (PCA) is a well-known technique involving a linear transformation of one vector into another, by an orthogonalization procedure.

This method which was initially proposed by Switzer and Green (1984) in image processing and remote sensing has been smoothly introduced in geostatistics (see references). Like the stepwise conditional transform (Leuangthong and Deutsch, 2003), the MAF is a powerful tool for analyzing and simulating coregionalized variables. It is practically convenient to avoid manipulating large systems in cokriging and the modeling of cross-covariance and use of the linear model of coregionalization.

Algorithm

Consider a multivariate data matrix of dimension p x n

$$\mathbf{Z} = \begin{bmatrix} z_1(\mathbf{u}_1) & z_1(\mathbf{u}_2) & \ldots\ldots\ldots & z_1(\mathbf{u}_n) \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ z_p(\mathbf{u}_1) & z_p(\mathbf{u}_2) & \ldots\ldots\ldots & z_p(\mathbf{u}_n) \end{bmatrix} \tag{1}$$

with the columns being the collocated data and the row being the variables. The main idea of the minimum/maximum autocorrelation factor (MAF) is to transform the correlated multivariate variables $Z_1, \ldots, Z_p$ into uncorrelated factors $\tilde{Z}_1, \ldots, \tilde{Z}_p$. Switzer and Green (1984) proposed the following algorithm which mainly uses the properties of the eigenvectors and eigenvalues of symmetric matrices.

Part one: principal components analysis at h=0.

- Compute the empirical covariance matrix as

$$\hat{\Sigma}_0 = \frac{1}{n}\left[\mathbf{Z}\,\mathbf{Z}^T\right] \tag{2}$$

- Consider the spectral decomposition (eigenvalue decomposition) of the above matrix as

$$\hat{\Sigma}_0 = Q_1\, D_1\, Q_1^T \tag{3}$$

where D_1 is the matrix of the eigenvalues and Q_1 the matrix of the orthonormal eigenvectors.

- Ensure that the entries of D_1 are in a decreasing order and that the corresponding eigenvectors are in rows

- Compute the (standardized) first PCA transformations $V_1, \dots, V_p$ by the following

$$V(\mathbf{u}) = D_1^{-1/2} \; Q_1 \, \mathbf{Z}(\mathbf{u}) \tag{4}$$

Part two: principal components analysis at a nonzero lag h using the previous factors.

- Select a nonzero lag distance h.
- Compute the experimental omni-directional symmetric cross-variance matrix $\hat{\Gamma}_V(\mathbf{h})$ for matrix V.
- Then consider the spectral decomposition of $\hat{\Gamma}_V(\mathbf{h})$ as

$$\hat{\Gamma}_V(\mathbf{h}) = Q_2 \, D_2 \, Q_2^T \tag{5}$$

- Finally, the principal components factors give the MAF factors as

$$\mathbf{T}(\mathbf{u}) = Q_2 \, V(\mathbf{u}) \tag{6}$$

Forward transform of the MAF

The forward transform of the MAF technique is obtained through the matrix A given by

$$A = Q_2 \; D_1^{-1/2} \; Q_1 \tag{7}$$

Indeed by plugging Equation (4) into Equation (6), we derive that

$$\mathbf{T}(\mathbf{u}) = Q_2 \, V(\mathbf{u}) = Q_2 \; D_1^{-1/2} \; Q_1 \, \mathbf{Z}(\mathbf{u}) = A \, \mathbf{Z}(\mathbf{u}) \tag{8}$$

Back transform of the MAF

The back transform operation of the MAF technique is utterly computed through the inverse of the matrix A given in Equation (7).

Data application: Jura data set

We apply the MAF technique by considering three correlated variables from the Jura data set (Goovaerts, 1997). The correlated variables used in this note are Cd, Co and Cr of sample size n=259. We consider the shift lag h=0.187 with a tolerance number taken to be half of this lag. We transform each variable into their normal score space (Desbarats and Dimitrakopoulos, 2000). The sample correlation matrix is as

$$\hat{\Sigma}_0 = \begin{bmatrix} 1.000 & 0.338 & 0.669 \\ 0.338 & 1.000 & 0.447 \\ 0.669 & 0.447 & 1.000 \end{bmatrix}. \tag{12}$$

The spectral eigen decomposition $\hat{\Sigma}_0 = Q_1 \, D_1 \, Q_1^T$ gives

$$Q_1 = \begin{bmatrix} 0.598 & 0.493 & 0.631 \\ -0.472 & 0.853 & -0.218 \\ -0.646 & -0.167 & 0.744 \end{bmatrix} \tag{13}$$

and

$$D_1 = \begin{bmatrix} 1.984 & 0 & 0 \\ 0 & 0.697 & 0 \\ 0 & 0 & 0.317 \end{bmatrix} \quad (14)$$

Thus we obtain the first factors V(u) using Equation (4) as

$$V(\mathbf{u}) = D_1^{-1/2} \; Q_1 \, \mathbf{Z}(\mathbf{u})$$

Taking the shift lag h=0.187, the sample semivariogram matrix of the standardized factor V is as

$$\hat{\Gamma}_V(0.187) = \begin{bmatrix} 0.578 & -0.117 & 0.095 \\ -0.117 & 0.668 & 0.112 \\ 0.095 & 0.112 & 0.734 \end{bmatrix}. \quad (15)$$

An eigen decomposition as $\hat{\Gamma}_V(\mathbf{h}) = Q_2 \, D_2 \, Q_2^T$ produces

$$Q_2 = \begin{bmatrix} 0.038 & 0.579 & 0.813 \\ -0.699 & 0.597 & -0.392 \\ 0.713 & 0.554 & -0.428 \end{bmatrix} \quad (16)$$

and

$$D_2 = \begin{bmatrix} 0.819 & 0 & 0 \\ 0 & 0.731 & 0 \\ 0 & 0 & 0.429 \end{bmatrix} \quad (17)$$

Thus the forward transform MAF matrix A follows by

$$A = \begin{bmatrix} -1.244 & 0.3637 & 0.9402 \\ -0.1845 & 0.4816 & -0.9870 \\ 0.4806 & 0.9425 & -0.3908 \end{bmatrix} \quad (18)$$

and the back transform process is conducted through the following inverse

$$A^{-1} = \begin{bmatrix} -0.4937 & -0.6841 & 0.5401 \\ 0.3636 & -0.0229 & 0.9324 \\ 0.2697 & -0.8964 & 0.3539 \end{bmatrix} \quad (19)$$

Figure (1) gives the cross plots of the MAF factors as the correlations between the factors are removed.

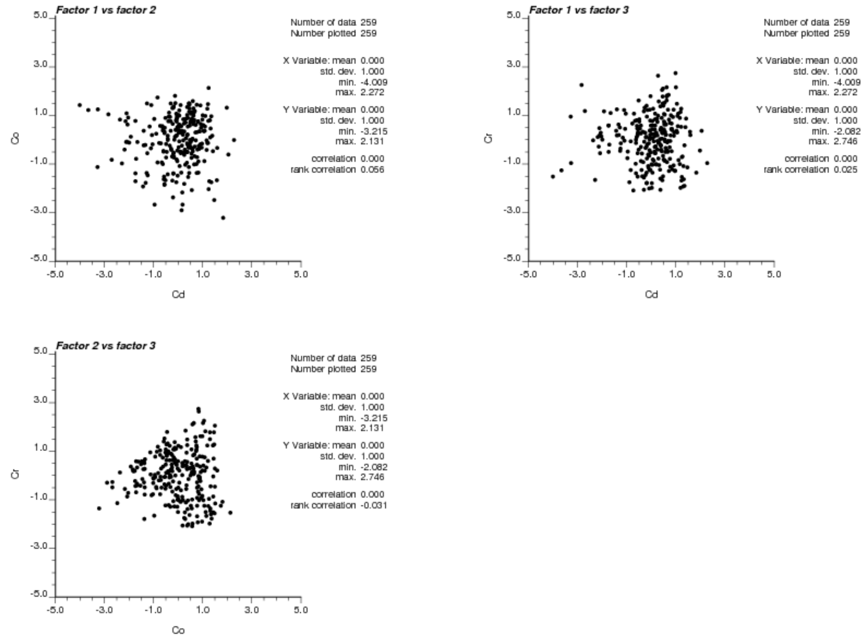

Figure 1: Cross plots of the three MAF factors using the lag h=0.187.

Program: maf.exe

```
START OF PARAMETERS:
NS_transf.out                         - File with data (after normal transformation)
3                                     - Number of variables
5   7   9                             - Columns for variables
1   2   0                             - Columns for locations
-20000.5      1.0e21                  - Trimming limits
0.30     0.15   0.15                  - Hlag, Htol1, Htol2 (lag and tolerance)
first-Testing.out                     - Eigenvector first PCA
first-Trans.out                       -first PCA transformation factors
Second-Testing.out                    -file for eigen values second PCA at nonzero lag
output.out                            -file for final MAF factors
```

Appendix: The min/max autocorrelation formalism

Consider p multivariate data $Z_1(\mathbf{s}), \ldots, Z_p(\mathbf{s})$ and let

$$\mathbf{Z}(\mathbf{s}) = \Big(Z_1(\mathbf{s}), \ldots, Z_p(\mathbf{s})\Big)^T.$$

Consider p orthogonal linear combinations $T_1(\mathbf{s}), \ldots, T_p(\mathbf{s})$ such that

$$T_j(\mathbf{s}) = \sum_{k=1}^{p} a_{j,k} Z_k(\mathbf{s})$$

for j=1,…, p. The idea of the maximum/minimum autocorrelation factor is similar to the principal component analysis. Basically each transform T_j is determined so as to exhibit greater spatial correlation than T_{j-1} that is the previously determined factor while being orthogonal to the other transformed factors

(Desbarats and Dimitrakopoulos, 2000). The orthogonality requirement being that for any j different of j', one checks that

$$\mathbb{C}ov\Big(T_j(\mathbf{s}), T_{j'}(\mathbf{s}')\Big) \equiv 0.$$

The covariance of the factor T_j is given as

$$\begin{aligned}
C_{T_j}(\mathbf{s}, \mathbf{s}+\mathbf{h}) &= \mathbb{C}ov\Big(T_j(\mathbf{s}), T_j(\mathbf{s}+\mathbf{h})\Big) \\
&= \sum_{k_1=1}^{p} \sum_{k_2=1}^{p} a_{j,k_1} a_{j,k_2} \mathbb{C}ov\Big(Z_{k_1}(\mathbf{s}), Z_{k_2}(\mathbf{s}+\mathbf{h})\Big) \\
&= \sum_{k_1=1}^{p} \sum_{k_2=1}^{p} a_{j,k_1} a_{j,k_2} C_{k_1,k_2}(\mathbf{h}) \\
&= \sum_{k_1=1}^{p} \sum_{k_2=1}^{p} a_{j,k_1} a_{j,k_2} C_{k_2,k_1}(-\mathbf{h}) \\
&= \sum_{k_1=1}^{p} \sum_{k_2=1}^{p} a_{j,k_1} a_{j,k_2} \left(\frac{C_{k_1,k_2}(\mathbf{h}) + C_{k_2,k_1}(-\mathbf{h})}{2}\right).
\end{aligned}$$

Consider the variance-covariance matrix

$$\Sigma_{\mathbf{h}} = \mathbb{C}ov\Big(\mathbf{Z}(\mathbf{s}), \mathbf{Z}(\mathbf{s}+\mathbf{h})\Big).$$

Thus

$$C_{T_j}(\mathbf{s}, \mathbf{s}+\mathbf{h}) = \frac{1}{2}\mathbf{a}_j\Big(\Sigma_{\mathbf{h}} + \Sigma_{-\mathbf{h}}\Big)\mathbf{a}_j^T.$$

Now under the second order stationary assumption, consider the variance matrix

$$\Gamma_{\mathbf{h}} = \mathbb{V}ar\Big(\mathbf{Z}(\mathbf{s}) - \mathbf{Z}(\mathbf{s}+\mathbf{h})\Big) = 2\Sigma_0 - \Sigma_{\mathbf{h}} - \Sigma_{-\mathbf{h}}.$$

It follows that

$$C_{T_j}(\mathbf{s}, \mathbf{s}+\mathbf{h}) = \mathbf{a}_j\Big(\Sigma_0 - \frac{\Gamma_{\mathbf{h}}}{2}\Big)\mathbf{a}_j^T$$

where the correlation coefficient is derived by

$$Corr(\mathbf{s}, \mathbf{s}+\mathbf{h}) = 1 - \frac{1}{2}\frac{\mathbf{a}_j\,\Gamma_{\mathbf{h}}\,\mathbf{a}_j^T}{\mathbf{a}_j\,\Sigma_0\,\mathbf{a}_j^T}$$

Minimization of the above correlation is equivalent to maximize the second term of the right hand side. It is shown (Switzer and Green, 1984) that the coefficients $\mathbf{a}_j$ yielding the MAF factors are obtained as the left hand eigen vectors of the matrix

$$\Delta(\mathbf{h}) = \Gamma_{\mathbf{h}}\,\Sigma_0^{-1}$$

The technique for deriving the eigen vectors $\mathbf{a}_j$ is the above MAF method with two principal components analysis.

References

Davis, M.W., Greenes, K.A., 1983. Estimation using spatially distributed multivariate data: an example with coal quality. *Mathematical Geology 15 (2), 287–300.*

Desbarats, A.J., Dimitrakopoulos, R., 2000. Geostatistical simulation of regionalized pore-size distributions using min/max autocorrelation factors. Mathematical Geology 32 (8), p. 919–941.

Dimitrakopoulos, R., Fonseca, M.B., 2003. Assessing risk in grade–tonnage curves in a complex copper deposit, northern Brazil, based on an efficient joint simulation of multiple correlated variables. In: Proceedings of the Application of Computers and Operations Research in the Minerals Industries, Cape Town, South Africa, p. 373-382.

Dimitrakopoulos, R., Mackie, S., 2003. Stochastic simulation for the quantification of mine spoil variability and rehabilitation decision making. In: Proceedings of MODSIM 2003, International Congress on Modelling and Simulation, Townsville, Australia, pp. 1703–1708.

Green Andrew A. Green, Berman Mark, Switzer Paul, and Craig Maurice D. , A transformation for ordering multispectral data in terms of image quality with implications for noise removal, IEEE Transactions on Geoscience and Remote Sensing, vol. 26, no. 1, p. 65-74, 1988.

Goovaerts, P., 1993. Spatial orthogonality of the principal components computed from coregionalized variables. Mathematical Geology 25 (3), 281–302.

Jolliffe, I.T., 1986. Principal Component Analysis. Springer, New York, 271pp.

Leuangthong, O., Deutsch, C.V., 2003. Stepwise conditional transformation for simulation of multiple variables. Mathematical Geology 35 (2), 155–173.

Switzer, P., and Green, A. A., 1984, Min/Max autocorrelation factors for multivariate spatial imaging Technical Report No. 6, Department of Statistics, Stanford University.

Switzer Paul, Min/max autocorrelation factors for multivariate spatial imagery, in Computer Science and Statistics, L. Billard, Ed. 1985, pp. 13–16, Elsevier, Science Publishers B.V. (North Holland)

Vargas-Guzman, J.A., Dimitrakopoulos R., 2003. Computational properties of min/max autocorrelation factors. Computers & Geosciences 29 (6), 715–723.

Wackernagel, H., 1998. Principal component analysis for autocorrelated data: a geostatistical perspective. Centre de Geostatistique-Ecole des Mines de Paris Technical Report N-22/98/G, 47pp.