C. R. Geoscience 346 (2014) 64–74

Contents lists available at ScienceDirect

Comptes Rendus Geoscience

www.sciencedirect.com

Applied geophysics

Multivariate geostatistical simulation by minimising spatial cross-correlation

Babak Sohrabian *, Abdullah Erhan Tercan

Hacettepe University, Department of Mining Engineering, 06800 Beytepe, Ankara, Turkey

ARTICLE INFO

Article history:
Received 9 January 2014
Accepted after revision 16 January 2014
Available online 18 April 2014

Keywords:
Joint simulation
MAF
Multivariate geostatistics
Orthogonalization
Spatial correlation

ABSTRACT

Joint simulation of attributes in multivariate geostatistics can be achieved by transforming spatially correlated variables into independent factors. In this study, a new approach for this transformation, Minimum Spatial Cross-correlation (MSC) method, is suggested. The method is based on minimising the sum of squares of cross-variograms at different distances. In the approach, the problem in higher space ($N \times N$) is reduced to $N \times (N-1)/2$ problems in the two-dimensional space and the reduced problem is solved iteratively using Gradient Descent Algorithm. The method is applied to the joint simulation of a set of multivariate data in a marble quarry and the results are compared with Minimum/Maximum Autocorrelation Factors (MAF) method.

1. Introduction

Linear transformation of spatially correlated variables into uncorrelated factors has been one of the most challenging issues in mining engineering and earth sciences. In this direction, several methods have been introduced by researchers. Xie et al. (1995) and Tercan (1999) used simultaneous diagonalisation in finding approximately uncorrelated factors at several lag distances by simultaneously diagonalising a set of variogram matrices. Switzer and Green (1984) developed the method of Minimum/Maximum Autocorrelation Factors (MAF) for the objective of separating signals from noise in multivariate imagery observations. The method is first introduced to geostatistical community by Desbarats and Dimitrakopoulos (2000) in the context of multivariate geostatistical simulation of pore-size distributions.

Fonseca and Dimitrakopoulos (2003) used MAF method for assessing risks in grade-tonnage curves in a complex copper deposit. Boucher and Dimitrakopoulos (2009) presented a method for the conditional block simulation of a non-Gaussian vector random field. They, first, orthogonalised a vector random function with MAF method and then used LU simulation to generate possible realizations. Rondon (2011) discussed the joint simulation of spatially cross-correlated variables using MAF factors in detail and gave some examples.

MAF is basically a two-stage principal component analysis applied to variance–covariance matrices at short and long lag distances. Goovaerts (1993) proves that in the presence of a two-structure linear model of co-regionalization (2SLMC), the factors are spatially uncorrelated. But the assumption of a 2SLMC is not reasonable for most of real data sets and also an extension of MAF to more than two distinct structure matrices is not possible (Vargas-Guzmán and Dimitrakopoulos, 2003). In the case where fitting a 2SLMC is not possible, a data-driven version of MAF method can be used (Desbarats and Dimitrakopoulos, 2000; Tercan, 1999; Sohrabian and Ozcelik, 2012a). In this approach, the variance–covariance matrices are calculated directly from the data set.

There are also some studies that use the independency property of the generated factors. For example, Sohrabian

* Corresponding author.
E-mail address: babak_sohrabian@hacettepe.edu.tr (B. Sohrabian).

http://dx.doi.org/10.1016/j.crte.2014.01.002
1631-0713/

and Ozcelik (2012b) introduce Independent Component Analysis (ICA) to transform spatially correlated attributes of an andesite quarry into independent factors. Then they estimated each factor independently and back-transformed the results into the real data space to determine exploitable blocks. Tercan and Sohrabian (2013) used independent component analysis in joint simulation of some quality attributes of a lignite deposit.

Goovaerts (1993) proves that in the general case, it is impossible to find factors that are exactly uncorrelated at all lag distances. When spatially uncorrelated factors cannot be produced, one looks for algorithms that produce approximately uncorrelated factors. For that, Mueller and Ferreira (2012) used Uniformly Weighted Exhaustive Diagonalization with Gauss iterations (U-WEDGE), introduced by Tichavsky and Yeredor (2009), for joint simulation of a multivariate data set from an iron deposit. In their case study, Mueller and Ferreira (2012) show that the U-WEDGE algorithm performs better than MAF.

In the present study, Minimum Spatial Cross-Correlation (MSC) method is introduced for generating approximately uncorrelated factors. It aims to minimise cross-variance matrices at different lag distances using the gradient descent algorithm. In this method the decorrelation problem is reduced to the solution of a sequence of 2 by 2 problems. Against other blind source separation algorithms such as U-WEDGE and ICA which generally work in high-dimensional spaces and try to solve the problem by choosing $N \times N$ initial matrices the MSC method is more convenient. In addition, ICA and U-WEDGE algorithms are sensitive to the choice of the initial matrices so that several applications of these algorithms do not converge to the same result (Hyvarinen et al., 2001). But, the MSC method approximately converges to the same result and this can be considered as an advantage over methods that directly solve $N \times N$ optimization problems.

The outline of the paper is as follows: the second section describes multivariate random field model. The third section explains the theory of MSC. The method is presented in 2D and then is generalized into an N-dimensional space. In the fourth section the method is applied to joint simulation of multivariate data obtained from an andesite quarry and the efficiency of the method in generating spatially orthogonalised factors is measured and compared to that of MAF method. Then MSC and MAF factors are used to simulate some attributes of a marble quarry. The last section includes the conclusions.

2. The multivariate random field

Let $\boldsymbol{Z}(\boldsymbol{u}) = [Z_1(\boldsymbol{u}), Z_2(\boldsymbol{u}), \ldots, Z_N(\boldsymbol{u})]$ be an N-dimensional stationary random field with zero mean and unit variance. In this multivariate case, the variogram matrix is given by

$$\boldsymbol{\Gamma}_Z(\boldsymbol{h}) = \frac{1}{2} E\left[(\boldsymbol{Z}(\boldsymbol{u}+\boldsymbol{h}) - \boldsymbol{Z}(\boldsymbol{u}))(\boldsymbol{Z}(\boldsymbol{u}+\boldsymbol{h}) - Z(\boldsymbol{u}))^{\mathrm{T}}\right] \quad (1)$$

The variogram matrix depends only on the lag distance $\boldsymbol{h}$, assuming that correlations vanish as $\boldsymbol{h} \to \infty$. In the presence of spatial cross-correlation, each variable should be simulated by considering the cross-correlations. In such cases, the traditional method is co-simulation, but it is impractical and time consuming due to difficulties arising from the fitting of a valid model of coregionalisation and the solving of large cokriging systems (Goovaerts, 1993). To ease multivariate simulation, $\boldsymbol{Z(u)}$ can be transformed into spatially uncorrelated factors $\boldsymbol{F(u)}$ in such way that

$$\boldsymbol{F}(\boldsymbol{u}) = \boldsymbol{Z}(\boldsymbol{u})\boldsymbol{W} \quad (2)$$

where $\boldsymbol{W}$ is an orthogonal transformation matrix. Then each factor $F_1, F_2, \ldots, F_N$, can be simulated separately and simulated factors can be back-transformed into the original space. This is a linear transformation that can remove linear correlations of variables and in the presence of non-linear correlations among variables it cannot be helpful. The transformation process results in factors which should be simulated by using one of the geostatistical methods for which stationary assumption holds. We assume that before running MSC the multivariate data are whitened with principal component analysis. By using principal component analysis, we guarantee the orthogonality of the produced factors. Orthogonal factors can be parameterised by half the parameters which are needed in any arbitrary matrix. Whitening also restricts the possible results to a unit circle (Hyvarinen et al., 2001).

3. MSC Method

Researchers have proposed various methods for generating spatially orthogonal factors. Some criteria have also been introduced to measure how well these methods orthogonalise the variogram matrices at different lag distances. For example, Tercan (1999) proposed the following measure:

$$\tau(\boldsymbol{h}) = \frac{\varphi(\boldsymbol{h})}{\xi(\boldsymbol{h})}, |\boldsymbol{h}| > 0 \quad (3)$$

where $\varphi(\boldsymbol{h}) = \sum_{k=1}^{N}\sum_{k<j}^{N}\left|\gamma_{\boldsymbol{F}}(\boldsymbol{h}; k, j)\right|$ and $\xi(\boldsymbol{h}) = \sum_{k=1}^{N}\gamma_{\boldsymbol{F}}(\boldsymbol{h}; k, k)$. This measure compares the sum of off-diagonal elements of the factor variogram matrix $\Gamma_{\boldsymbol{Z}}(\boldsymbol{h})$ to the sum of its diagonal elements for each lag distance $\boldsymbol{h}$. It is used by Rondon (2011) and Mueller and Ferreira (2012) to compare various factorization algorithms. Efficient factorization algorithms would produce $\tau(\boldsymbol{h})$ as close as possible to zero at each lag distance. This measure can also be considered as an optimization criterion in producing the desired factors.

While developing our method in deriving spatially uncorrelated factors, we will consider Eq. 3, proving that $\xi(\boldsymbol{h})$ is constant (see Appendix A for a proof). Therefore it suffices to minimize the sum of $\varphi(\boldsymbol{h})$ values at various lag distances. In the following section, a simple method based on the gradient descent algorithm is presented for iteratively minimizing the φ value.

$$\varphi = \sum_{i=1}^{l}\sum_{k=1}^{N}\sum_{k<j}^{N}\left|\gamma_{\boldsymbol{F}}(\boldsymbol{h}_i; k, j)\right|, \ |\boldsymbol{h}| > 0 \quad (4)$$

In Eq. 4, l denotes the number of lags that are considered in the calculations. The number of lags depends on the