



Calculation of the normal scores variogram used for truncated Gaussian lithofacies simulation: theory and FORTRAN code

Phaedon C. Kyriakidis^{a,*}, Clayton V. Deutsch^b, Marshall L. Grant^c

^a*Department of Geological and Environmental Sciences, Stanford University, Stanford, CA 94305-2115, USA*

^b*Department of Civil and Environmental Engineering, University of Alberta, Edmonton, Alberta, Canada, T6G 2G7*

^c*Department of Chemical Engineering, Yale University, New Haven, CT 06520-8286 USA*

Received 28 August 1997; revised 28 August 1998; accepted 28 August 1998

Abstract

Application of the truncated Gaussian method for categorical variable simulation requires the calculation of an appropriate normal scores variogram for generating the Gaussian random field. In the case of only two categories, the appropriate variogram can be determined by inverting the indicator variogram model from the bivariate Gaussian distribution. Even though no closed-form relation exists for such inversion, the proper normal scores variogram can be obtained through numerical integration via a series approximation. The procedure is illustrated with a small simulation example demonstrating the close reproduction of the variogram of the categorical data. © 1999 Elsevier Science Ltd. All rights reserved.

Code available at <http://www.iang.org/CGEditor/index.htm>

Keywords: Geostatistics; Stochastic simulation; Indicator variograms; Gaussian random fields

1. Introduction

Stochastic simulation is being increasingly utilized for modeling the spatial distribution of categorical and continuous attributes, such as lithofacies and soil types, or porosity and permeability (Srivastava, 1995; Damsleth and Omre, 1997). The procedure consists of generating alternative, high resolution 3D images/realizations, which mimic the heterogeneity expected in the real media and inferred from the available data. Such heterogeneity or spatial variability is usually quantified by two-point variogram or covariance models, which are at the heart of geostatistical simulation algorithms. The goal of stochastic simulation is

then the reproduction of such models of spatial dependence.

A routinely encountered situation is that of a binary 0/1 regionalized indicator random variable (RV) $J(\mathbf{u}; s_k)$ coding the presence or absence of a particular lithofacies s_k at a location \mathbf{u} within the study area A :

$$J(\mathbf{u}; s_k) \begin{cases} 1, & \text{if } S(\mathbf{u}) = s_k, \forall k = 1, \dots, K, \mathbf{u} \in A. \\ 0, & \text{if not} \end{cases}$$

where the categorical RV $S(\mathbf{u})$ can take K mutually exclusive and exhaustive outcomes $\{s_k, k = 1, \dots, K\}$.

Similarly, a regionalized indicator RV $J(\mathbf{u}; z)$ could code the event of a petrophysical attribute $Z(\mathbf{u})$ being greater than a particular threshold z_k :

$$J(\mathbf{u}; z_k) \begin{cases} 1, & \text{if } Z(\mathbf{u}) > z_k \forall k = 1, \dots, K, \mathbf{u} \in A. \\ 0, & \text{if not} \end{cases}$$

* Corresponding author. E-mail address: phaedon@pan-gea.stanford.edu.

where K threshold values $\{z_k, k = 1, \dots, K\}$ discretize the variability of the Z attribute.

Various algorithms exist for generating a number L of realizations $\{j^{(l)}(\mathbf{u}), \mathbf{u} \in A\}$, $l = 1, \dots, L$ of the indicator random function (RF) $\{J(\mathbf{u}), \mathbf{u} \in A\}$; the most common and straightforward being sequential indicator simulation (Journal and Alabert, 1988). Another alternative is to truncate a realization $\{y^{(l)}(\mathbf{u}), \mathbf{u} \in A\}$ of a standard normal RF $\{Y(\mathbf{u}), \mathbf{u} \in A\}$ to create an indicator realization $\{j^{(l)}(\mathbf{u}), \mathbf{u} \in A\}$ (Journal and Isaaks, 1984; Matheron et al., 1987):

$$j^{(l)}(\mathbf{u}) = \begin{cases} 1, & \text{if } y^{(l)}(\mathbf{u}) > y_0, \\ 0, & \text{if not} \end{cases}$$

with y_0 being some threshold value. The only data available are indicator data, yet the covariance model $C_Y(\mathbf{h})$ of the Gaussian RF $\{Y(\mathbf{u}), \mathbf{u} \in A\}$ to be truncated is needed for generating the realization $\{j^{(l)}(\mathbf{u}), \mathbf{u} \in A\}$.

One could use for $C_Y(\mathbf{h})$ the (standardized to unit sill) indicator covariance model $C_J(\mathbf{h})$ inferred from the categorical data. This practice leads to a mismatch between the target indicator variogram model $C_J(\mathbf{h})$ and the variogram of the resulting simulated indicator values $\{j^{(l)}(\mathbf{u}), \mathbf{u} \in A\}$. Such mismatch becomes larger for indicator variograms related to lithofacies with small proportions, that is when the threshold value y_0 deviates significantly from the zero median. In addition, Matheron (1989) showed that it is not consistent to use the same variogram type (e.g. spherical) for the Gaussian RF $\{Y(\mathbf{u}), \mathbf{u} \in A\}$, as the target indicator variogram $C_J(\mathbf{h})$ inferred from the sample indicator data.

Alternatively, the original categorical data are transformed to continuous pseudo normal scores data, as in Xu and Journal (1993), and $C_Y(\mathbf{h})$ is substituted by the covariance inferred from such data. The resulting pseudo normal score values depend critically on the procedure used to despik the original indicator data, i.e. break the ties between binary values. Random despiking leads to a too-high nugget effect and parameters for a despiking algorithm, which ranks the categorical data according to local areal proportions, are difficult to establish.

In the case of a single threshold y_0 separating only two lithofacies, the normal scores covariance $C_Y(\mathbf{h})$ is directly linked, through a one-to-one relation, to the lithofacies indicator covariance $C_J(\mathbf{h})$ (Journal and Isaaks, 1984). However, such relation exists only for the case of only two lithofacies. A standard Gaussian RF $Y(\mathbf{u})$ is fully specified by its sole covariance model $C_Y(\mathbf{h})$, hence it cannot be used to identify more than one indicator covariance model $C_J(\mathbf{h})$.

One way around this limitation would be to consider a series of Gaussian RFs $Y_k(\mathbf{u})$, $k = 1, \dots, K$, each

used to simulate (after truncation) a nested set of only two lithofacies; i.e. simulate each set nested into a previously simulated set. This approach can be used only to simulate lithofacies nested one into another.

An approximation consists of inverting the single covariance model $C_Y(\mathbf{h})$ from some average of the $(K-1)$ indicator variograms, in the case of K lithofacies. It is this single average indicator variogram that would be reproduced through simulation, not any particular lithofacies indicator variogram model.

Recent efforts to extend the truncated Gaussian method for simulation of multiple ($K > 2$), non-nested, lithofacies call for multiple Gaussian RFs. Contrary to the case $K = 2$, simulation of multiple, non-nested, lithofacies involves a highly non-unique inversion. Iterative procedures based on trial and error selection of the input normal scores covariances and cross-covariances between the multiple Gaussian RFs have been reported in the literature (see, for example, Loc'h and Galli, 1997).

This paper recalls the one-to-one relationship between $C_J(\mathbf{h})$ and $C_Y(\mathbf{h})$ in the case of only two lithofacies and proposes a power series development to approximate it.

2. Theoretical framework

Consider a stationary standard (zero mean and unit variance) multivariate normal RF $\{Y(\mathbf{u}), \mathbf{u} \in A\}$. The univariate and bivariate cumulative distribution functions (cdfs) are:

$$G(y) = \text{Prob}\{Y(\mathbf{u}) \leq y\} = p, \quad \forall \mathbf{u} \in A$$

where $y = G^{-1}(p)$ is the standard normal p -quantile and

$$G(\rho_Y(\mathbf{h}); y, y') = \text{Prob}\{Y(\mathbf{u}) \leq y, Y(\mathbf{u} + \mathbf{h}) \leq y'\}$$

where $\rho_Y(\mathbf{h}) = C_Y(\mathbf{h}) = E\{Y(\mathbf{u})Y(\mathbf{u} + \mathbf{h})\}$ is the covariance of the Gaussian RF $Y(\mathbf{u})$, fully characterizing its multivariate distribution.

Next, consider the indicator RF $J(\mathbf{u}; y_0)$ defined by truncating the Gaussian RF $Y(\mathbf{u})$ at the p_0 quantile:

$$J(\mathbf{u}; y_0) = \begin{cases} 1, & \text{if } Y(\mathbf{u}) > y_0 = G^{-1}(p_0), \\ 0, & \text{if not} \end{cases}$$

The resulting stationary indicator RF $J(\mathbf{u}; y_0)$ has the following moments:

- Mean m_J :

$$m_J = E\{J(\mathbf{u}; y_0)\} = \text{Prob}\{Y(\mathbf{u}) > y_0\} = 1 - p_0.$$

- Variance σ_J^2 :

$$\sigma_J^2 = \text{Var}\{J(\mathbf{u}; y_0)\} = m_J(1 - m_J) = p_0(1 - p_0).$$

- Non-centered indicator covariance $K_J(\mathbf{h}; y_0)$:

$$\begin{aligned} K_J(\mathbf{h}; y_0) &= E\{J(\mathbf{u}; y_0)J(\mathbf{u} + \mathbf{h}; y_0)\} \\ &= \text{Prob}\{Y(\mathbf{u}) > y_0, Y(\mathbf{u} + \mathbf{h}) > y_0\} \\ &= G(\rho_Y(\mathbf{h}); y_0, y_0) + 1 - 2p_0. \end{aligned} \quad (2)$$

The non-centered indicator covariance $K_J(\mathbf{h}; y_0)$ is actually the two-point distribution function of the Gaussian RF $\{Y(\mathbf{u}), \mathbf{u} \in A\}$. The centered indicator covariance $C_J(\mathbf{h}; y_0)$ is then:

$$\begin{aligned} C_J(\mathbf{h}; y_0) &= K_J(\mathbf{h}; y_0) - (1 - p_0)^2 \\ &= G(\rho_Y(\mathbf{h}); y_0, y_0) - p_0(1 - p_0). \end{aligned} \quad (3)$$

Inversion of Eq. (2) or Eq. (3) provides the normal scores covariance $\rho_Y(\mathbf{h})$ as a function of the non-centered indicator covariance $K_J(\mathbf{h}; y_0)$:

$$\rho_Y(\mathbf{h}) = G^{-1}(K_J(\mathbf{h}; y_0) - (1 - 2p_0); y_0) \quad (4)$$

or as a function of the centered indicator covariance $C_J(\mathbf{h}; y_0)$:

$$\rho_Y(\mathbf{h}) = G^{-1}(C_J(\mathbf{h}; y_0) + p_0(1 - p_0); y_0). \quad (5)$$

Conversely, the relation linking $K_J(\mathbf{h}; y_0)$ to $\rho_Y(\mathbf{h})$ is determined by integrating the bivariate Gaussian density over the area delineated by $Y_1 > y_0$ and $Y_2 > y_0$ (see hatched area in Fig. 1):

$$K_J(\mathbf{h}; y_0) = \int_{y_0}^{\infty} \int_{y_0}^{\infty} g_{Y_1 Y_2}(\mathbf{h}; y_1, y_2) dy_1 dy_2 \quad (6)$$

where $g_{Y_1 Y_2}(y_1, y_2)$ is the bivariate density function of two Gaussian RVs $Y_1 = Y(\mathbf{u})$ and $Y_2 = Y(\mathbf{u} + \mathbf{h})$ separated by vector \mathbf{h} (Anderson, 1958):

$$\begin{aligned} g_{Y_1 Y_2}(y_1, y_2) &= \frac{1}{2\pi\sqrt{1 - \rho_Y^2(\mathbf{h})}} \exp\left[-\frac{y_1^2 - 2\rho_Y(\mathbf{h})y_1y_2 + y_2^2}{2(1 - \rho_Y^2(\mathbf{h}))}\right]. \end{aligned} \quad (7)$$

Unfortunately, integrals of Gaussian densities, such as in Eq. (6), do not have closed-form analytical expressions. The task is then to approximate $K_J(\mathbf{h}; y_0)$ through a numerical integration procedure, and then inversion of Eq. (6) with respect to $\rho_Y(\mathbf{h})$ for a given value of \mathbf{h} . Note that $\rho_Y(\mathbf{h})$ is constrained by $-1 \leq \rho_Y(\mathbf{h}) \leq 1$ and that the integral in Eq. (6) is monotonic in $\rho_Y(\mathbf{h})$.

A power series expansion approach is proposed hereafter for numerically evaluating Eq. (6). First, the coefficients for the power series approximation of $C_J(\mathbf{h})$ in terms of $C_Y(\mathbf{h})$ are computed, and then the power series is reverted to obtain a development of $C_Y(\mathbf{h})$ in terms of $C_J(\mathbf{h})$.

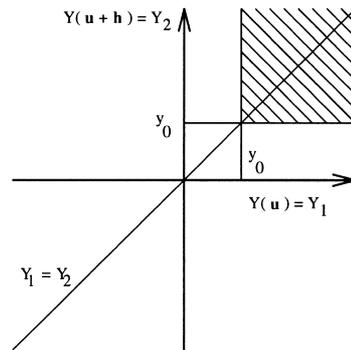


Fig. 1. Bivariate scatterplot of Y_1 versus Y_2 : non-centered indicator covariance $K_J(\mathbf{h}; y_0)$ is integral of $f_{Y_1 Y_2}(y_1, y_2)$ over hatched region.

3. Implementation

The first step towards simplifying Eq. (6) is to perform a 45° rotation and scaling of the original variables Y_1, Y_2 in order to eliminate the cross terms from the exponential integrand.

Consider the following change of variables:

$$u = \frac{y_1 + y_2}{2\sqrt{1 + \rho_Y(\mathbf{h})}} \quad \text{and} \quad v = \frac{y_2 - y_1}{2\sqrt{1 - \rho_Y(\mathbf{h})}}.$$

The original variables are then written as

$$\begin{aligned} y_1 &= u\sqrt{1 + \rho_Y(\mathbf{h})} - v\sqrt{1 - \rho_Y(\mathbf{h})}, \\ y_2 &= u\sqrt{1 + \rho_Y(\mathbf{h})} + v\sqrt{1 - \rho_Y(\mathbf{h})} \end{aligned}$$

with Jacobian:

$$J(u, v) = \begin{vmatrix} \sqrt{1 + \rho_Y(\mathbf{h})} & -\sqrt{1 - \rho_Y(\mathbf{h})} \\ \sqrt{1 + \rho_Y(\mathbf{h})} & \sqrt{1 - \rho_Y(\mathbf{h})} \end{vmatrix} = 2\sqrt{1 - \rho_Y^2(\mathbf{h})}.$$

The limits of integration are now modified as follows:

- For $y_1 = y_0$ and $y_2 = y_0$: $u = u_0 = y_0/(\sqrt{1 + \rho_Y(\mathbf{h})})$.
- For $y_1 = +\infty$ and $y_2 = +\infty$: $u = +\infty$.
- For $y_2 = y_0$ and $y_1 \in (-\infty, +\infty)$: $v = v_1 = -(u - u_0)\sqrt{(1 + \rho_Y(\mathbf{h}))}/(1 - \rho_Y(\mathbf{h}))$.
- For $y_1 = y_0$ and $y_2 \in (-\infty, +\infty)$: $v = v_2 = (u - u_0)\sqrt{(1 + \rho_Y(\mathbf{h}))}/(1 - \rho_Y(\mathbf{h}))$.

Substituting the expressions for y_1 and y_2 into the numerator of the bivariate density (Eq. (7)) yields

$$-(y_1^2 - 2\rho_Y(\mathbf{h})y_1y_2 + y_2^2) = -2(1 - \rho_Y^2(\mathbf{h}))(u^2 + v^2).$$

The integral in Eq. (6) giving the non-centered indicator covariance $K_J(\mathbf{h}; y_0)$ is then simplified to

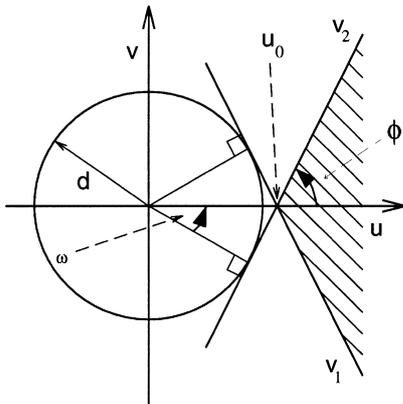


Fig. 2. Area of integration in terms of independent variables u and v defined after 45° rotation and subsequent scaling of original variables y_1 and y_2 .

$$K_J(\mathbf{h}; y_0) = \frac{1}{\pi} \int_{u_0}^{\infty} du \int_{v_1}^{v_2} \exp[-(u^2 + v^2)] dv$$

where integration is now performed over the hatched region in Fig. 2.

Note that the lines bounding the area of integration are tangent to the circle of radius $d = y_0/\sqrt{2}$ centered at the origin. From symmetry of the domain of integration, the previous expression reduces to

$$K_J(\mathbf{h}; y_0) = \frac{2}{\pi} \int_{u_0}^{\infty} du \int_0^{v_2} \exp[-(u^2 + v^2)] dv \tag{8}$$

with $u_0 = y_0/(\sqrt{1 + \rho_Y(\mathbf{h})})$

and $v_2 = (u - u_0)\sqrt{(1 + \rho_Y(\mathbf{h})) / (1 - \rho_Y(\mathbf{h}))}$.

The normal scores correlogram $\rho_Y(\mathbf{h})$ does not appear explicitly in the integrand, instead it is accounted for in the limits of integration u_0 and v_2 .

The second step is to convert to polar coordinates (see Fig. 3) defining

$$u = r \cos(\theta - \omega) \quad \text{and} \quad v = r \sin(\theta - \omega)$$

where r is the radial distance of $p = (u, v)$ and $a = \theta - \omega$ is the polar angle. Note that θ is the angle between the u axis and the line from p to the origin, and ω is the angle between the u axis and the radius perpendicular to the line bounding the area of integration (Fig. 3). The new area of integration can be then defined by considering that $a \in [0, \pi/2 - \omega]$, hence $\theta \in [\omega, \pi/2]$ and that $r \geq r_0 = d/\cos \theta$, hence $r \in [d \sec \theta, +\infty)$.

Substituting the expressions for u and v into the exponent of the integrand Eq. (8) yields

$$-(u^2 + v^2) = -[r^2 \cos^2(\theta - \omega) + r^2 \sin^2(\theta - \omega)] = -r^2.$$

Hence, Eq. (8) for the non-centered indicator covariance becomes

$$\begin{aligned} K_J(\mathbf{h}; y_0) &= \frac{2}{\pi} \int_{\omega}^{\pi/2} d\theta \int_{d \sec \theta}^{\infty} r e^{-r^2} dr \\ &= \frac{1}{\pi} \int_{\omega}^{\pi/2} e^{-d^2 \sec^2 \theta} d\theta \end{aligned} \tag{9}$$

where, in this situation, the calculation has been reduced to a single integration.

The third step involves another change of variables in order to use series expansion, by introducing the new variable $t = \sqrt{\sec^2 \theta - 1}$. From Fig. 3 and the definition of the limit of integration $v_2 = (u - u_0)\sqrt{(1 + \rho_Y(\mathbf{h})) / (1 - \rho_Y(\mathbf{h}))}$, one can see that

$$\tan \phi = \sqrt{\frac{1 + \rho_Y(\mathbf{h})}{1 - \rho_Y(\mathbf{h})}}$$

Therefore,

$$\tan \omega = \tan\left(\frac{\pi}{2} - \phi\right) = \sqrt{\frac{1 - \rho_Y(\mathbf{h})}{1 + \rho_Y(\mathbf{h})}}$$

The area of integration is now modified as follows:

- For $\theta = \omega$: $t = t_0 = \sqrt{(1 - \rho_Y(\mathbf{h})) / (1 + \rho_Y(\mathbf{h}))}$
- For $\theta = \pi/2$: $t = \tan \pi/2 \Rightarrow t = 0$.

The non-centered indicator covariance expression (Eq. (9)) then becomes

$$K_J(\mathbf{h}; y_0) = \frac{e^{-d^2}}{\pi} \int_{t_0}^0 \frac{e^{-d^2 t^2}}{t^2 + 1} dt \tag{10}$$

where dt is expressed in terms of $d\theta$ using the fact that $\sec^2 \theta - 1 = \tan^2 \theta$, hence

$$u = r \cos(\theta - \omega) \quad \text{and} \quad v = r \sin(\theta - \omega)$$

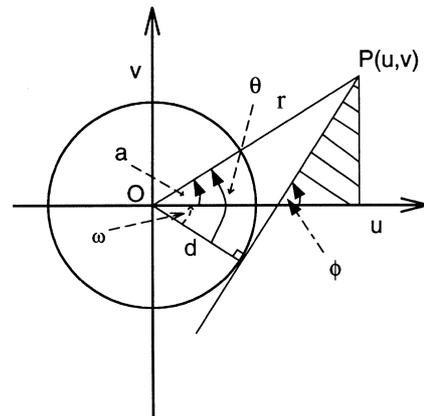


Fig. 3. Conversion to polar coordinates: r is radial distance of $p(u, v)$ and θ is polar angle.

$$\frac{\partial t}{\partial \theta} = \sec^2 \theta \implies \partial \theta = \frac{\partial t}{t^2 + 1}.$$

Application of power series expansion at $t = 0$ (see, for example, Apostol, 1967), leads to:

$$K_J(\mathbf{h}; y_0) = (1 - p) - \frac{e^{-d^2}}{\pi} \int_0^{t_0} \frac{e^{-d^2 t^2}}{t^2 + 1} dt \approx (1 - p) - \sum_{n=1}^N b_n t_0^n \tag{11}$$

where N is the number of terms retained in the series approximation, and coefficients b_n are evaluated as

$$b_n = \frac{1}{n \pi} \frac{\partial^{(n)}((e^{-d^2 t^2})/(t^2 + 1))}{\partial t}$$

or, equivalently, as

$$b_n = (-1)^{((n-1)/2)(e^{-d^2/n\pi})} \sum_{k=0}^{(n-1)/2} (d^{2k}/k!), \text{ if } n \text{ odd,}$$

$$b_n = 0, \text{ if } n \text{ even.}$$

Direct reversion of the series equation (Eq. (11)) leads to a power series for $t_0 = \sqrt{(1 - \rho_Y(\mathbf{h})) / (1 + \rho_Y(\mathbf{h}))}$ (see, for example, Abramovitz and Stegun, 1972):

$$t_0 = \sum_{n=1}^N a_n y^n \tag{12}$$

where $y = C_J(\mathbf{h})p_0(1 - p_0) + p_0(1 - p_0)$ and a_n are coefficients defined as:

$$a_n = -\frac{\sum_{i=1}^{n-1} a_i b(n, i)}{b(n, n)}$$

with $b(n, n)$ being the n th element of a matrix B containing the coefficients b_n .

The value of $\rho_Y(\mathbf{h})$ is then obtained by simple algebra as:

$$\rho_Y(\mathbf{h}) = \frac{1 - t_0^2}{1 + t_0^2}. \tag{13}$$

Note that, even if the series involves N terms the actual number of non-zero terms is $(N - 1)/2$.

Remark: The resulting normal scores correlogram values $\rho_Y(\mathbf{h})$ must constitute a positive definite covariance table for any lag \mathbf{h} . This could be checked a posteriori by verifying that the Fourier transform of the numerically derived correlogram $\rho_Y(\mathbf{h})$, i.e. the corresponding spectral density function, is non-negative for all possible lags \mathbf{h} (see, for example, Christakos, 1984). Note that for certain indicator covariance models $C_J(\mathbf{h}; y_0)$, such as a Gaussian model that behaves like a parabola near the origin, the resulting normal scores correlogram model $\rho_Y(\mathbf{h})$ may not be positive definite, since it behaves like a polynomial with an exponent

greater than 2. In general, for an arbitrary threshold y_0 , in the neighborhood of $\rho_Y(\mathbf{h}) = 1$, the variogram $\gamma_J(\mathbf{h})$ of the indicator RF behaves like the square root of the variogram $1 - \rho_Y(\mathbf{h})$ of the Gaussian RF (Matheron, 1989) and this poses problems when dealing with Gaussian indicator covariance models. However, Gaussian covariance models characterize phenomena with very high spatial continuity and should not be used for modeling a (by definition) discontinuous indicator RF. Hence, for realistic positive definite indicator covariance models $C_J(\mathbf{h})$ the normal scores correlogram values derived by **bigaus2** constitute a licit covariance table, since they are obtained by direct evaluation (integration and inversion) of Eq. (6).

Alternatively, one could ensure positive definiteness of the numerically derived normal scores correlogram $\rho_Y(\mathbf{h})$ by fitting a permissible parametric correlogram model $\rho_Y(\mathbf{h}; \theta)$ to $\rho_Y(\mathbf{h})$. Here, θ denotes the vector of parameters (sills, ranges) of each basic structure comprising the parametric correlogram model $\rho_Y(\mathbf{h}; \theta)$ adopted.

4. Program description

The GSLIB program **bigaus** (Deutsch and Journel, 1998) has been modified to include the reverse procedure, i.e. calculation of $\gamma_Y(\mathbf{h})$ from $\gamma_J(\mathbf{h})$. The new program **bigaus2** can handle the following cases:

- (1) Input: normal scores variogram model $\gamma_Y(\mathbf{h})$. Output: values of the indicator variogram $\gamma_J(\mathbf{h})$ defined at a specific threshold y_0 .
- (2) Input: indicator variogram model $\gamma_J(\mathbf{h})$ for a specified threshold y_0 . Output: values of the normal scores variogram $\gamma_Y(\mathbf{h})$ required for generating the Gaussian RF $\{Y(\mathbf{u}), \mathbf{u} \in A\}$.
- (3) Input: experimental normal scores variogram values $\gamma_Y^*(\mathbf{h})$. Output: ‘experimental’ values of the indicator variogram $\gamma_J(\mathbf{h})$ defined at a specific threshold y_0 .
- (4) Input: experimental indicator variogram values $\gamma_J^*(\mathbf{h})$ for a specified threshold y_0 . Output: ‘experimental’ values of the normal scores variogram $\gamma_Y^*(\mathbf{h})$ required for generating the Gaussian RF $\{Y(\mathbf{u}), \mathbf{u} \in A\}$.

Note that in case (4), a further modeling step is required in order to fit a licit normal scores variogram model to the values output from **bigaus2**. Test runs of **bigaus2** indicate that in case (2), the resulting normal scores variogram values could be used directly for generating the Gaussian RF $\{Y(\mathbf{u}), \mathbf{u} \in A\}$, provided that the threshold value y_0 is not too high, and that numerical integration errors are small.

```

Parameters for BIGAUS2
*****

START OF PARAMETERS:
1                               \input variogram: model (1) sample (2)
samplegam.var                 \file with input experimental variogram
0.25                          \threshold cdf value
2                               \ indicator (1) normal scores (2)
bigaus2.out                   \file for output of variograms
3                               \number of thresholds
0.25   0.50   0.75           \threshold cdf values
1   20                               \number of directions and lags
0.0   0.0   1.0             \azm(1), dip(1), lag(1)
1   0.2                       \nst, nugget effect
1   0.8   0.0   0.0   0.0     \it,cc,ang1,ang2,ang3
                               \a_hmax, a_hmin, a_vert
                               10.0   5.0   10.0

```

Fig. 4. Parameter file for **bigaus2**.

The **bigaus2** program follows GSLIB conventions. The parameters required for the program, shown also in Fig. 4, are listed as follows:

- **imd**: if **imd** is set to 1, then an input variogram model is required. If **imd** is set to 2, then an input file **infi** containing experimental semivariogram values is required.
 - **infi**: if **imd** is set to 2, the input file containing the experimental semivariogram values should be provided.
 - **pcut**: if **imd** is set to 2, a single threshold is required. This threshold is expressed in units of cumulative probability, e.g. the median is 0.50.
 - **icl**: the type of calculation is specified. If **icl** is set to 1, then the input consists of a standardized indicator semivariogram (either model or experimental). If **icl** is set to 2, the input consists of a normal scores semivariogram (either model or experimental).
 - **outfi**: the output file for the normal scores or indicator semivariograms depending on **icl**. The format is the same as that created by GSLIB; therefore, GSLIB program **vargplt** could be used to plot these indicator variograms.
 - **ncut**: the number of thresholds.
 - **zc(ncut)**: **ncut** threshold values are required. These thresholds are expressed in units of cumulative probability, for example, the lower quartile is 0.25, the median is 0.50. Note that the results are symmetric: the variogram for the 5th percentile (0.05) is the same as the variogram for the 95th percentile (0.95).
 - **ndir** and **nlag**: the number of directions and the number of lags to be considered.
 - **azm**, **dip** and **lag**: for each of the **ndir** directions, an azimuth and a dip must be specified (by **azm** and **dip**, respectively) along with a unit lag offset (**lag**).
 - **nst** and **c0**: the number of nested structures and the nugget effect.
- For each of the **nst** nested structures, one must define **it**: the type of the structure; **cc** the *c* parameter; **ang1**, **ang2**, **ang3**, the angles defining the geometric anisotropy; **aa_{hmax}**, the maximum horizontal range; **aa_{hmin}**, the minimum horizontal range and **aa_{vert}**, the vertical range. A detailed description of these parameters is given in section II.3 of the GSLIB manual.

The FORTRAN source code in file **bigaus2.f**, along with the corresponding parameter file **bigaus2.par**, can be downloaded from a searchable database www.iamg.org/CGEditor/index.htm.

5. An illustrative example

The following example illustrates the discrepancy incurred by generating the Gaussian RF $\{Y(\mathbf{u}), \mathbf{u} \in A\}$ using the standardized indicator variogram $\gamma_I(\mathbf{h})$ instead of the correct normal scores variogram $\gamma_Y(\mathbf{h})$. First, the difference between the standardized indicator variogram and the correct normal scores variogram model is shown for a hypothetical system with various proportions of two lithofacies.

The standardized indicator variogram modeling the spatial distribution of the two lithofacies is a spherical model with a 5% nugget and a range of 10 units; the field dimensions being 100×100 units. Four possible facies proportions are examined, $p = 0.5$, $p = 0.8$, $p = 0.9$ and $p = 0.97$. The corresponding normal scores variograms (case (2) in **bigaus2**) are shown in Fig. 5.

Fig. 5 illustrates the smaller nugget variance of the correct normal scores variogram (dashed line) and its parabolic behavior at the origin, as opposed to the linear behavior of the input standardized indicator variogram model (solid line). The mismatch is more pronounced as the proportion p deviates more from

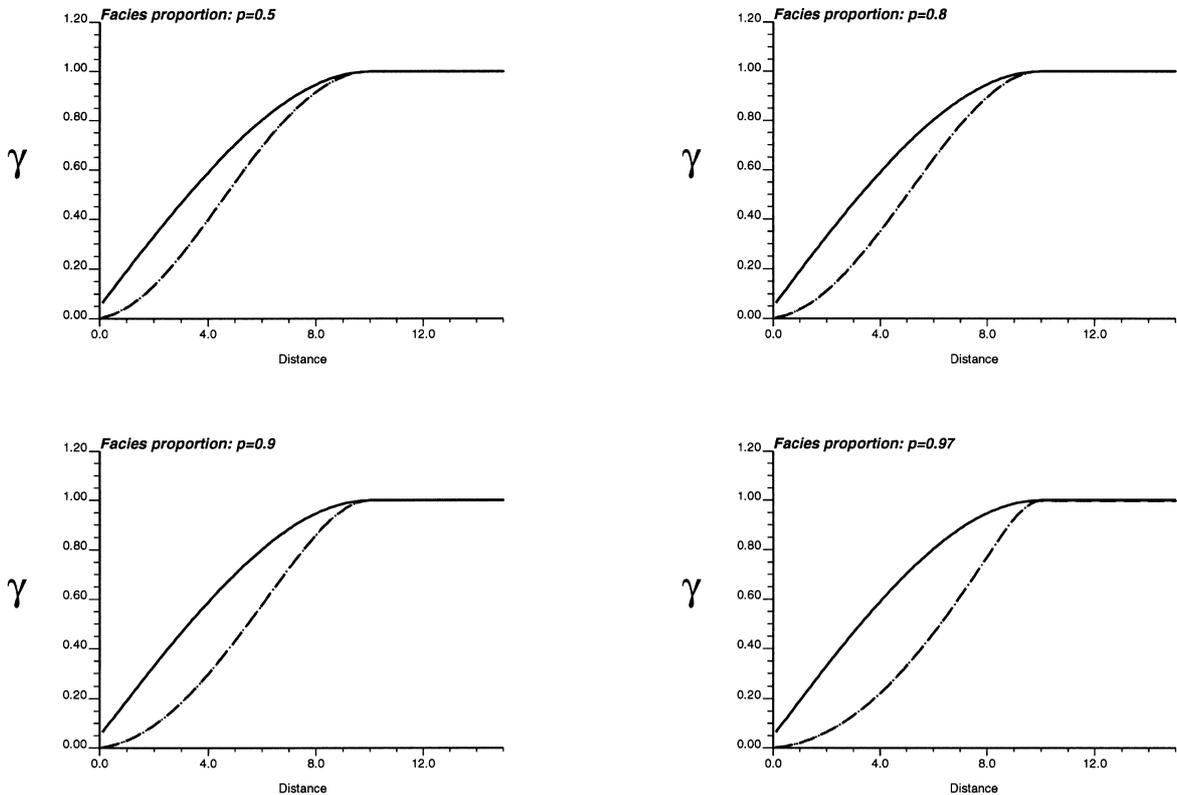


Fig. 5. Discrepancy between input standardized indicator variogram models (solid line) and correct normal scores variograms (dashed line) for various facies proportions p . Note that normal scores variograms do not have zero nugget effect; this can be detected if graphs were to be plotted on larger scale. Nugget effect values are 0.0031, 0.0026, 0.0021 and 0.0014, respectively, for $p = 0.5, 0.8, 0.9$ and 0.97 .

0.5, i.e. as one of the two lithofacies becomes more abundant. Everything else being equal, utilization of the correct normal scores variogram leads to a more continuous realization at short scales than had the standardized indicator variogram been used. Utilization of the correct normal scores variogram model becomes more critical when the data control does not suffice to impose the pattern of spatial variation. For this reason, the following example using unconditional simulation shows the discrepancies at their worst.

Consider the problem of generating a realization of the spatial distribution of the two lithofacies (Fig. 6). Let the desired proportion of facies A be $1 - p = 0.25$ and that of facies B be $p = 0.75$. Let the target indicator variogram be an isotropic spherical one with relative nugget of 5% and range of 10 units.

Two realizations of a Gaussian RF are generated using program **sgsim** (Deutsch and Journel, 1998) (see top of Fig. 6). The first (Fig. 6, top left) was generated using an isotropic spherical variogram model with 5%

nugget variance contribution and a range of 10 units. The second (Fig. 6, top right) was calculated using the correct normal scores variogram as generated from program **bigaus2** for $1 - p = 0.25$. These two Gaussian realizations were truncated via program **gtsim** (Deutsch and Journel, 1998) at the $1 - p = 0.25$ quantile. The two resulting indicator realizations (binary images) are shown in the middle of Fig. 6. Both indicator realizations have the correct proportion of lithofacies, yet the omnidirectional variograms of the simulated indicator values (Fig. 6, bottom) are quite different. Utilization of the standardized indicator variogram model results in a binary image exhibiting a higher nugget variance (more noise) than the target indicator variogram model. On the contrary, using the correct normal scores variogram model yields, after truncation at the $1 - p = 0.25$ quantile, a binary image whose indicator variogram is very close to the target indicator variogram model.

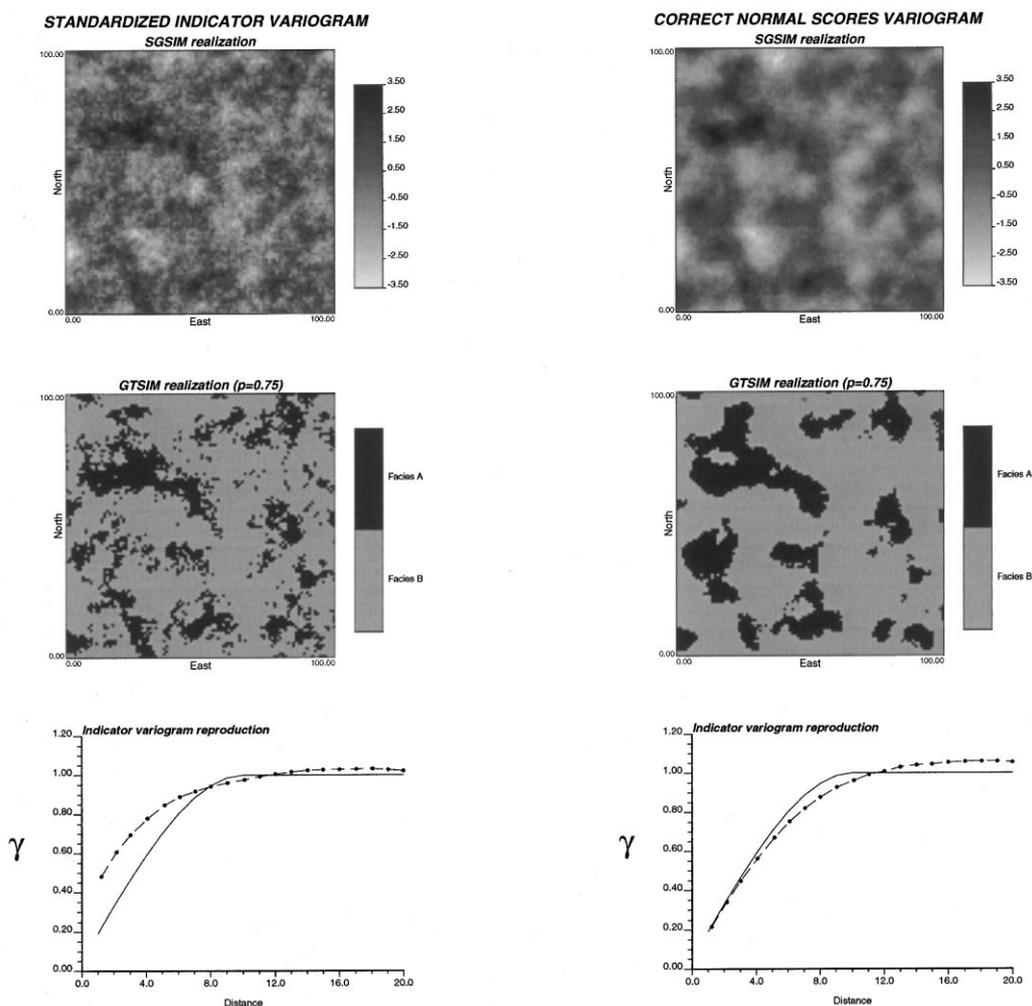


Fig. 6. Indicator variogram reproduction for two gtsim realizations using: (a) standardized indicator variogram model (left graphs) and (b) correct normal scores variogram model (right graphs). Model indicator variogram is shown in solid line.

6. Discussion

The theoretical basis for determining the correct normal scores variogram for truncated Gaussian simulation of categorical variables is briefly recalled. The one-to-one relationship between the normal scores correlogram $\rho_Y(\mathbf{h})$ and the indicator correlogram $\rho_I(\mathbf{h})$ after truncation at the $1-p$ standard normal quantile, is approximated using a series expansion.

Program **bigaus2** allows conversion between $\rho_Y(\mathbf{h})$ and $\rho_I(\mathbf{h})$ and vice versa. Two subroutines are used for the above conversion:

Subroutine **rhoy - j** calculates the indicator correlogram $\rho_I(\mathbf{h})$, which results after truncating a bivariate Gaussian RF (with correlogram $\rho_Y(\mathbf{h})$) at the $1-p$ quantile. This procedure is currently coded in program **bigaus** of GSLIB (Deutsch and Journel,

1998) using a formal numerical integration via Simpson's rule. The conversion from $\rho_Y(\mathbf{h})$ to $\rho_I(\mathbf{h})$ is used to check bivariate normality, i.e. theoretical $\rho_I(\mathbf{h})$ values are compared to experimentally available ones, and the hypothesis of bivariate normality regarding the RF $\{Y(\mathbf{u}), \mathbf{u} \in A\}$ may be refuted or not.

Subroutine **rhoj - y** calculates the correlogram $\rho_Y(\mathbf{h})$ between two Gaussian RVs constituting the bivariate Gaussian RF $\{Y(\mathbf{u}), \mathbf{u} \in A\}$, so that the resulting binary realization, after truncation of $\{Y(\mathbf{u}), \mathbf{u} \in A\}$ at the $1-p$ quantile, reproduces the inferred/target indicator correlogram $\rho_I(\mathbf{h})$. For this latter conversion, the better reproduction of the input variogram model has been illustrated via a simple example using unconditional stochastic simulation.

References

- Abramovitz, M., Stegun, I.A. (Eds.), 1972. Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables, 9th (revised) ed. Dover, New York, 1046 pp.
- Anderson, T., 1958. An Introduction to Multivariate Statistical Analysis. John Wiley & Sons, New York, 374 pp.
- Apostol, T.M., 1967. Calculus, vol. 1, 2nd ed. John Wiley & Sons, New York, 666 pp.
- Christakos, G., 1984. On the problem of permissible covariance and variogram models. *Water Resources Research* 20 (2), 251–265.
- Damsleth, E., Omre, H., 1997. Geostatistical approaches in reservoir evaluation. *Journal of Petroleum Technology* 49 (5), 498–501.
- Deutsch, C.V., Journel, A.G., 1998. *GSLIB: Geostatistical Software Library and User's Guide*, 2nd ed. Oxford University Press, New York, 368 pp.
- Journel, A.G., Alabert, F.G., 1988. Focusing on spatial connectivity of extreme valued attributes: stochastic indicator models of reservoir heterogeneities. Proceedings of the 63rd Annual Technical Conference and Exhibition of the Society of Petroleum Engineers, Volume on Reservoir Geology, Houston, Texas, October 2–5, 1988. SPE paper 18324, pp. 621–632.
- Journel, A.G., Isaaks, E., 1984. Conditional indicator simulation: application to a Saskatchewan uranium deposit. *Mathematical Geology* 16 (7), 685–718.
- Loc'h, G.L., Galli, A., 1997. Truncated pluri-Gaussian method: theoretical and practical points of view. In: Baafi, E., Schofield, N. (Eds.), *Geostatistics Wollongong '96*, vol. 1. Kluwer, Dordrecht, The Netherlands, pp. 211–222.
- Matheron, G., 1989. The internal consistency of models in geostatistics. In: Armstrong, M. (Ed.), *Geostatistics*, vol. 1. Kluwer, Dordrecht, pp. 21–38.
- Matheron, G., Beucher, H., de Fouquet, H., Galli, A., Guerillot, D., Ravenne, C., 1987. Conditional simulation of the geometry of fluvio-deltaic reservoirs. Proceedings of the 62nd Annual Technical Conference and Exhibition of the Society of Petroleum Engineers, Volume on Formation Evaluation and Reservoir Geology and Reservoir Engineering, Dallas, Texas, September 27–30, 1987. SPE paper 16753, pp. 123–131.
- Srivastava, R., 1995. An overview of stochastic methods for reservoir characterization. In: Yarus, J., Chambers, R. (Eds.), *Stochastic Modeling and Geostatistics: Principles, Methods and Case Studies*. AAPG Computer Applications in Geology, No. 3, pp. 3–16.
- Xu, W., Journel, A.G., 1993. Gtsim: Gaussian truncated simulation of lithofacies. In: Report 6. Stanford Center for Reservoir Forecasting, Stanford, CA, pp. 222–278.