A Close Look at Bivariate Gaussian Spatial Distributions

S. H. Derakhshan and C. V. Deutsch

The assumption of bivariate normal distribution is critical in many geostatistical applications such as simulation of a spatial variable. Direct and cross indicator variograms arise in many geostatistical applications such as indicator cokriging, truncated Gaussian simulation, sequential Indicator simulation and checking for bivariate Gaussianity. This paper addresses theoretical properties and derivation of direct and cross indicator variogram for a Gaussian random function.

1. Introduction

Many geostatistical simulation techniques including Sequential Gaussian Simulation assume that the stationary spatial continuous variable being simulated, $Y(\boldsymbol{u})$, follows a multivariate Gaussian spatial distribution.

The direct and cross indicator variograms are calculated based on different Gaussian cut-offs on a multivariate Gaussian random function. These Gaussian cut-offs are the basis for the indicator variables. The definition of indicator variable builds a useful relation between the continuous variable $Y(\boldsymbol{u})$ and the indicator variable $i_{p_k}(\boldsymbol{u})$. This definition helps to derive different spatial properties of the multivariate Gaussian random function. The indicator variable is defined as:

$$I_{p_k}(\boldsymbol{u}) = \begin{cases} 1, & if\ Y(\boldsymbol{u}) \le y_{p_k} \\ 0, & if\ not \end{cases}$$

Where, $k = 1, \dots, K$ is the index for indicator (category or facies) variable and p_k is its proportion. $y_{p_k} = G^{-1}(p_k)$ is the univariate Gaussian value or cut-off. The function $G(\cdot)$ is the univariate standard Gaussian cumulative distribution function (CDF):

$$E\{I_{p_k}(\boldsymbol{u})\} = Prob\{Y(\boldsymbol{u}) \le y_{p_k}\} = p_k = G(y_{p_k}) = \int_{-\infty}^{y_{p_k}} g(y) \cdot dy = \frac{1}{2\pi}\int_{-\infty}^{y_{p_k}} e^{\frac{-y^2}{2}} \cdot dy$$

$Y(\boldsymbol{u})$ is the multivariate Gaussian random function with zero mean, unit variance and covariance $\rho_Y(\boldsymbol{h})$. The bivariate distribution of $Y(\boldsymbol{u})$and $Y(\boldsymbol{u}+\boldsymbol{h})$ is fully characterized by the bivariate CDF, $K_I^{p_k,p_{k'}}(\boldsymbol{h})$:

$$\begin{aligned} K_I^{p_k,p_{k'}}(\boldsymbol{h}) &= E\left\{I_{p_k}(\boldsymbol{u}) \cdot I_{p_{k'}}(\boldsymbol{u}+\boldsymbol{h})\right\} = Prob\left\{Y(\boldsymbol{u}) \le y_{p_k}, Y(\boldsymbol{u}+\boldsymbol{h}) \le y_{p_{k'}}\right\} \\ &= \int_{-\infty}^{y_{p_{k'}}} \int_{-\infty}^{y_{p_k}} g\big(Y(\boldsymbol{u}), Y(\boldsymbol{u}+\boldsymbol{h}); \rho_Y(\boldsymbol{h})\big) \cdot dY(\boldsymbol{u}+\boldsymbol{h}) \cdot dY(\boldsymbol{u}) \end{aligned}$$

$K_I^{p_k,p_{k'}}(\boldsymbol{h})$ is also referred to as non-centered cross indicator covariance. $g\big(Y(\boldsymbol{u}), Y(\boldsymbol{u}+\boldsymbol{h}); \rho_Y(\boldsymbol{h})\big)$ is the bivariate Gaussian probability density function (PDF):

$$g\big(Y(\boldsymbol{u}), Y(\boldsymbol{u}+\boldsymbol{h}); \rho_Y(\boldsymbol{h})\big) = \frac{1}{2\pi\sqrt{1-\rho_Y^2(\boldsymbol{h})}} \cdot exp\left[\frac{-1}{2}\left(\frac{Y^2(\boldsymbol{u}) - 2\rho_Y(\boldsymbol{h}) \cdot Y(\boldsymbol{u}) \cdot Y(\boldsymbol{u}+\boldsymbol{h}) + Y^2(\boldsymbol{u}+\boldsymbol{h})}{1-\rho_Y^2(\boldsymbol{h})}\right)\right]$$

The centered cross indicator covariance is defined as:

$$C_I^{p_k,p_{k'}}(\boldsymbol{h}) = Cov\left\{I_{p_k}(\boldsymbol{u}) \cdot I_{p_{k'}}(\boldsymbol{u}+\boldsymbol{h})\right\} = E\left\{\left[I_{p_k}(\boldsymbol{u}) - p_k\right] \cdot \left[I_{p_{k'}}(\boldsymbol{u}+\boldsymbol{h}) - p_{k'}\right]\right\} = K_I^{p_k,p_{k'}}(\boldsymbol{h}) - p_k \cdot p_{k'}$$

Following the first and second order stationary assumptions, the cross indicator variogram is defined as below:

$$\begin{aligned} \gamma_I^{p_k,p_{k'}}(\boldsymbol{h}) &= \frac{1}{2}E\left\{\left[I_{p_k}(\boldsymbol{u}) - I_{p_k}(\boldsymbol{u}+\boldsymbol{h})\right] \cdot \left[I_{p_{k'}}(\boldsymbol{u}) - I_{p_{k'}}(\boldsymbol{u}+\boldsymbol{h})\right]\right\} = K_I^{p_k,p_{k'}}(\boldsymbol{0}) - K_I^{p_k,p_{k'}}(\boldsymbol{h}) \\ &= C_I^{p_k,p_{k'}}(\boldsymbol{0}) - C_I^{p_k,p_{k'}}(\boldsymbol{h}) \end{aligned}$$

Where, $K_I^{p_k,p_{k'}}(\boldsymbol{0})$ and $C_I^{p_k,p_{k'}}(\boldsymbol{0})$ are calculated as below:

$$K_I^{p_k,p_{k'}}(\boldsymbol{0}) = min(p_k, p_{k'})$$

$$C_I^{p_k,p_{k'}}(\mathbf{0}) = min(p_k, p_{k'}) \cdot [1 - max(p_k, p_{k'})]$$

$\gamma_I^{p_k,p_{k'}}(\boldsymbol{h})$ can also be determined experimentally from actual spatial data, based on theoretical definition it can be concluded that:

$$\hat{\gamma}_I^{p_k,p_{k'}}(\boldsymbol{h}) = \frac{1}{2N(\boldsymbol{h})} \sum_{\alpha=1}^{N(\boldsymbol{h})} \left[i_{p_k}(\boldsymbol{u}_\alpha) - i_{p_k}(\boldsymbol{u}_\alpha + \boldsymbol{h})\right] \cdot \left[i_{p_{k'}}(\boldsymbol{u}_\alpha) - i_{p_{k'}}(\boldsymbol{u}_\alpha + \boldsymbol{h})\right]$$

Capital I corresponds to the indicator random variable while small i correspond to indicator data.

2. Properties of Cross Indicator Variogram of Gaussian Random Function

The following properties hold for cross indicator variogram for Gaussian random function as a function of cut-offs:

1. $\gamma_I^{p_k,p_{k'}}(\boldsymbol{h})$ is always theoretically positive that is $\gamma_I^{p_k,p_{k'}}(\boldsymbol{h}) \geq 0$. Assume that $y_{p_{k'}} < y_{p_k}$. Based on the two cut-offs of $y_{p_k}$and $y_{p_{k'}}$, there are 9 regions in the 2 dimensional Cartesian plane. The 9 regions are numbered from 1 to 9. The cross indicator variogram is non-zero and positive in regions 3 and 7 while it is zero in other regions. Table 1 and Figure 1 show the non-negative property of direct and cross indicator variograms.

Table 1 the signs of the direct and cross indicator variograms

Region	Tail; $Y(\boldsymbol{u})$	Head; $Y(\boldsymbol{u}+\boldsymbol{h})$	$I_{p_{k'}}(\boldsymbol{u})$	$I_{p_{k'}}(\boldsymbol{u}+\boldsymbol{h})$	$I_{p_{k'}}(\boldsymbol{u})$	$I_{p_{k'}}(\boldsymbol{u}+\boldsymbol{h})$	$\gamma_I^{p_k,p_{k'}}(\boldsymbol{h})$
1	$Y(\boldsymbol{u}) < y_{p_{k'}} < y_{p_k}$	$Y(\boldsymbol{u}+\boldsymbol{h}) < y_{p_{k'}} < y_{p_k}$	0	0	0	0	0
2	$y_{p_{k'}} \leq Y(\boldsymbol{u}) < y_{p_k}$	$Y(\boldsymbol{u}+\boldsymbol{h}) < y_{p_{k'}} < y_{p_k}$	1	0	0	0	0
3	$y_{p_{k'}} < y_{p_k} \leq Y(\boldsymbol{u})$	$Y(\boldsymbol{u}+\boldsymbol{h}) < y_{p_{k'}} < y_{p_k}$	1	0	1	0	+
4	$Y(\boldsymbol{u}) < y_{p_{k'}} < y_{p_k}$	$y_{p_{k'}} \leq Y(\boldsymbol{u}+\boldsymbol{h}) < y_{p_k}$	0	1	0	0	0
5	$y_{p_{k'}} \leq Y(\boldsymbol{u}) < y_{p_k}$	$y_{p_{k'}} \leq Y(\boldsymbol{u}+\boldsymbol{h}) < y_{p_k}$	1	1	0	0	0
6	$y_{p_{k'}} < y_{p_k} \leq Y(\boldsymbol{u})$	$y_{p_{k'}} \leq Y(\boldsymbol{u}+\boldsymbol{h}) < y_{p_k}$	1	1	1	0	0
7	$Y(\boldsymbol{u}) < y_{p_{k'}} < y_{p_k}$	$y_{p_{k'}} < y_{p_k} \leq Y(\boldsymbol{u}+\boldsymbol{h})$	0	1	0	1	+
8	$y_{p_{k'}} \leq Y(\boldsymbol{u}) < y_{p_k}$	$y_{p_{k'}} < y_{p_k} \leq Y(\boldsymbol{u}+\boldsymbol{h})$	1	1	0	1	0
9	$y_{p_{k'}} < y_{p_k} \leq Y(\boldsymbol{u})$	$y_{p_{k'}} < y_{p_k} \leq Y(\boldsymbol{u}+\boldsymbol{h})$	1	1	1	1	0

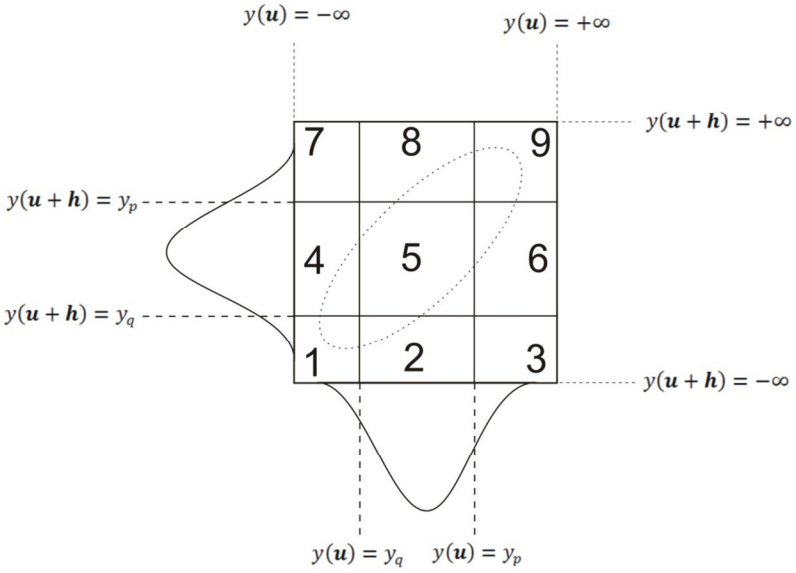

Figure 1 the nine different regions based on the two Gaussian cut-offs

2. $\gamma_I^{p_k,p_{k'}}(\boldsymbol{h})$ has the following symmetric properties:

$$\gamma_I^{p_k,p_{k'}}(\boldsymbol{h}) = \gamma_I^{p_{k'},p_k}(\boldsymbol{h}) = \gamma_I^{1-p_k,1-p_{k'}}(\boldsymbol{h}) = \gamma_I^{1-p_{k'},1-p_k}(\boldsymbol{h})$$

3. $\gamma_I^{p_k,p_{k'}}(\boldsymbol{h})$ has no closed form analytical formula in general case. A single or double integral should be solved to get the values for $\gamma_I^{p_k,p_{k'}}(\boldsymbol{h})$. There exist closed form analytical formula when $p_k = p_{k'} = \frac{1}{2}$:

$$\gamma_I^{\frac{1}{2},\frac{1}{2}}(\boldsymbol{h}) = \frac{1}{4} - \frac{ArcSin[\rho_Y(\boldsymbol{h})]}{2\pi}$$

The double integration for $\gamma_I^{p_k,p_{k'}}(\boldsymbol{h})$ can be reduced to a single integral analytically:

$$\begin{aligned}\gamma_I^{p_k,p_{k'}}(\boldsymbol{h}) &= min(p_k,p_{k'}) - \int_{-\infty}^{y_{p_{k'}}}\int_{-\infty}^{y_{p_k}} g\big(Y(\boldsymbol{u}),Y(\boldsymbol{u}+\boldsymbol{h});\rho_Y(\boldsymbol{h})\big)\cdot dY(\boldsymbol{u}+\boldsymbol{h})\cdot dY(\boldsymbol{u}) \\ &= min(p_k,p_{k'})\cdot[1-max(p_k,p_{k'})] \\ &\quad -\frac{1}{2\pi}\int_0^{ArcSin[\rho_Y(\boldsymbol{h})]} exp\left(-\frac{y_{p_k}^2 - 2y_{p_k}y_{p_{k'}}Sin\theta + y_{p_{k'}}^2}{2Cos^2\theta}\right)\cdot d\theta\end{aligned}$$

The integral has the simpler below forms if $p_k = p_{k'}$:

$$\gamma_I^{p_k,p_k}(\boldsymbol{h}) = p_k\cdot(1-p_k) - \frac{1}{2\pi}\int_0^{ArcSin[\rho_Y(\boldsymbol{h})]} exp\left(\frac{-y_{p_k}^2}{1+Sin\theta}\right)\cdot d\theta$$

Or if $p_k + p_{k'} = 1$:

$$\gamma_I^{p_k,1-p_k}(\boldsymbol{h}) = [min(p_k,1-p_k)]^2 - \frac{1}{2\pi}\int_0^{ArcSin[\rho_Y(\boldsymbol{h})]} exp\left(\frac{-y_{p_k}^2}{1-Sin\theta}\right)\cdot d\theta$$

Journel and Posa (1990) introduced an approximation for the direct indicator variogram (where the two cut-off are equal), i.e. $\gamma_I^{p_k,p_k}(\boldsymbol{h})$, with acceptable precision for a specific range of Gaussian cut-offs:

$$\left[\gamma_I^{p_k,p_k}(\boldsymbol{h})\right]^* = A\big(y_{p_k}\big)\cdot Exp_a(\boldsymbol{h}) + B\big(y_{p_k}\big)\cdot Exp_{0.3a}(\boldsymbol{h})$$

Where

$$\begin{cases} Exp_a(\boldsymbol{h}) = 1 - e^{\frac{-3\boldsymbol{h}}{a}} \\ A\big(y_{p_k}\big) + B\big(y_{p_k}\big) = p_k\cdot(1-p_k) \\ \dfrac{A\big(y_{p_k}\big)}{B\big(y_{p_k}\big)} = 5.26 + 0.123y_{p_k} - 4.65y_{p_k}^2 + 2.46y_{p_k}^3 - 0.37y_{p_k}^4 \\ -3.0 \le y_{p_k} \le +3.0 \end{cases}$$

The maximum relative error for this approximation is less than 7 % when $\boldsymbol{h} \ge 0.1a$:

$$\varepsilon_R = max_{\boldsymbol{h}\ge 0.1a}\left|\frac{\gamma_I^{p_k,p_k}(\boldsymbol{h}) - \left[\gamma_I^{p_k,p_k}(\boldsymbol{h})\right]^*}{\gamma_I^{p_k,p_k}(\boldsymbol{h})}\right| < 0.07$$

4. The derivatives of $\gamma_I^{p_k,p_{k'}}(\boldsymbol{h})$ with respect to $\boldsymbol{h}$ is:

$$\frac{\partial}{\partial\boldsymbol{h}}\left[\gamma_I^{p_k,p_{k'}}(\boldsymbol{h})\right] = \frac{d\gamma_Y(\boldsymbol{h})}{d\boldsymbol{h}}\cdot g\left(y_{p_k},y_{p_{k'}};\rho_Y(\boldsymbol{h})\right)$$

Which implies that $\frac{\partial}{\partial\boldsymbol{h}}\left[\gamma_I^{p_k,p_{k'}}(\boldsymbol{h})\right]$ and $\frac{d\gamma_Y(\boldsymbol{h})}{d\boldsymbol{h}}$ have the same sign since $g\left(y_{p_k},y_{p_{k'}};\rho_Y(\boldsymbol{h})\right)$ is positive. This relation takes advantage the stationary assumption of $\rho_Y(\boldsymbol{h}) + \gamma_Y(\boldsymbol{h}) = 1$ or $\frac{d\rho_Y(\boldsymbol{h})}{dh} + \frac{d\gamma_Y(\boldsymbol{h})}{dh} = 0$. The derivatives with respect to other variable $p_k, p_{k'}$ are as follows:

$$\frac{\partial}{\partial p_k}\left[\gamma_I^{p_k,p_{k'}}(\boldsymbol{h})\right] = \delta[p_k, min(p_k,p_{k'})] - \frac{1}{\sqrt{2\pi}}\cdot G\left(y_{p_k,p_{k'}}\right)$$

$$\frac{\partial}{\partial p_{k'}}\left[\gamma_I^{p_k,p_{k'}}(\boldsymbol{h})\right] = \delta[p_{k'}, min(p_k, p_{k'})] - \frac{1}{\sqrt{2\pi}} \cdot G\left(y_{p_{k'},p_k}\right)$$

Where $\delta[\cdot,\cdot]$ is Kronecker delta function and defined as below:

$$\delta[a,b] = \begin{cases} 1 & if \quad a = b \\ 0 & if \quad a \neq b \end{cases}$$

$y_{p_k,p_{k'}}$ and $y_{p_{k'},p_k}$ are auxiliary variables and defined as:

$$y_{p_k,p_{k'}} = \frac{y_{p_{k'}} - \rho_Y(\boldsymbol{h}) \cdot y_{p_k}}{\sqrt{1-\rho_Y^2(\boldsymbol{h})}}$$

$$y_{p_{k'},p_k} = \frac{y_{p_k} - \rho_Y(\boldsymbol{h}) \cdot y_{p_{k'}}}{\sqrt{1-\rho_Y^2(\boldsymbol{h})}}$$

5. It can be shown that $\gamma_I^{p_k,p_{k'}}(\boldsymbol{h})$ has Gaussian (in terms of variogram shape) type behavior if $p_k \neq p_{k'}$ and Exponential behavior if $p_k = p_{k'}$. The cross indicator variogram is hyper Gaussian (destruction effect) and has extraordinary low values when the Gaussian cut-offs (y_{p_k} and $y_{p_{k'}}$) are extreme low and high values. Figure 2 shows the cross indicator variogram for different probability cut-offs of 0.1, 0.5 and 0.9 from a spherical normal score variogram.

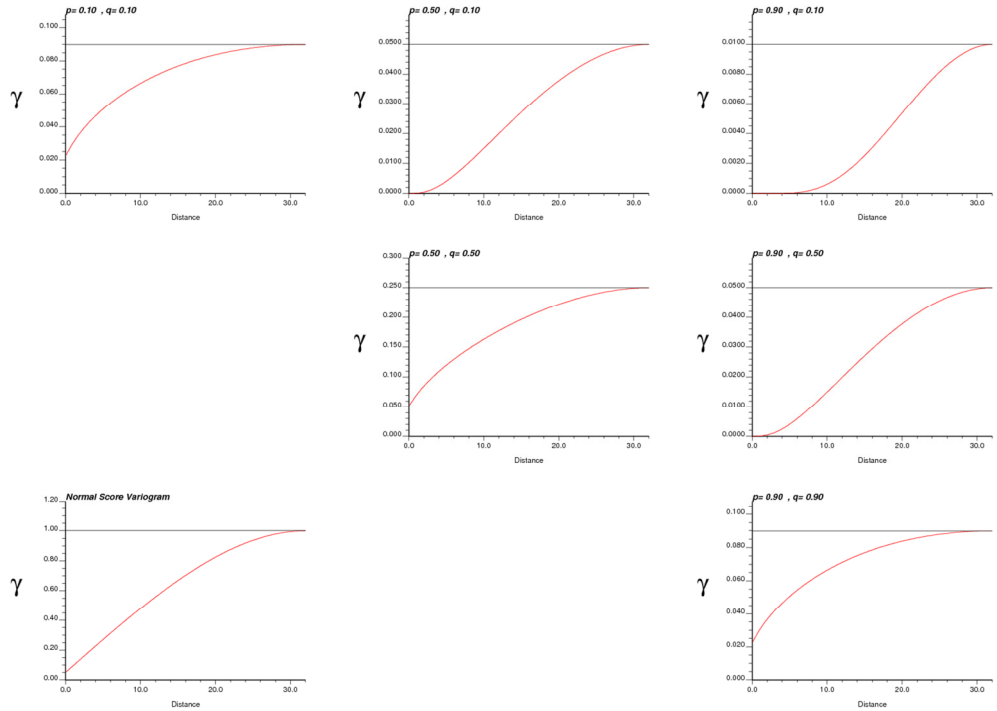

Figure 2 the cross indicator variograms for different probability cut-offs of 0.1, 0.5 and 0.9 from a single structure spherical normal score variogram

The normalized variogram values (values divided by associated sills) are also plotted in two dimensional planes for the three dependent variable of probability cut-offs (p and q) and the standardized lags, $\frac{\boldsymbol{h}}{a}$. Figure 3 shows these plots. The variograms values are calculated using the algorithm introduced in Derakhshan and Deutsch (2011).

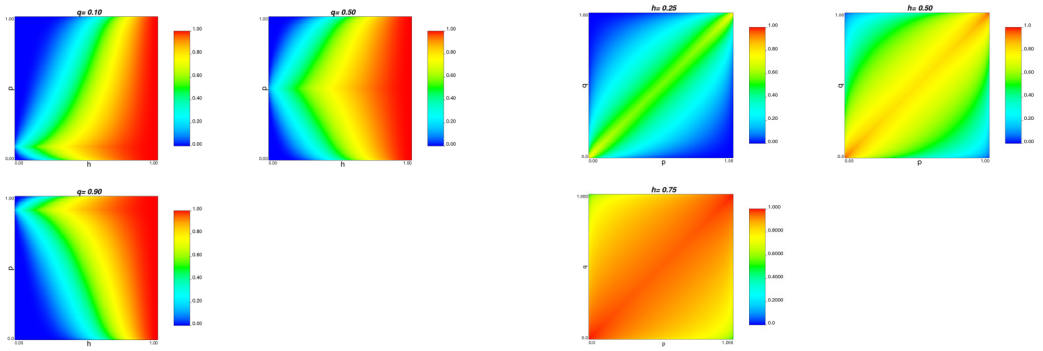

Figure 3 direct and cross indicator variograms for different probability cut-offs of 0.1, 0.5 and 0.9 at standardized lag distances of 0.25,0.50 and 0.75

6. All of the direct and cross indicator variograms have the same *range* as the original normal score variogram of Gaussian random field. The *sill* value for each of the indicator variograms is defined as below:

$$C_I^{p_k,p_{k'}}(\mathbf{0}) = min(p_k, p_{k'}) \cdot [1 - max(p_k, p_{k'})]$$

7. The direct and cross indicator variograms and covariances must satisfy below general order relations (Journel and Posa, 1990) because of the behavior and properties of bivariate normal CDF that underlay them (for simplicity p_k is replaced by p , and $p_{k'}$ is replaced by q):
   a. Property #1 entitle the following:
   $$0 \le \gamma_I^{p,q}(\boldsymbol{h}) \le min(p,q) \cdot [1 - max(p,q)] \le min(p,q) \le max(p,q)$$
   Or
   $$0 \le \gamma_I^{p,q}(\boldsymbol{h}) \le min(p,q) \cdot [1 - max(p,q)] \le 1 - max(p,q) \le 1 - min(p,q)$$
   b. The symmetric properties in #2 and integral definition in #3 entitle (for all $p, q, \boldsymbol{h}$):
   $$\gamma_I^{1,q}(\boldsymbol{h}) = \gamma_I^{p,1}(\boldsymbol{h}) = \gamma_I^{0,q}(\boldsymbol{h}) = \gamma_I^{p,0}(\boldsymbol{h}) = 0$$
   c. The bivariate normal CDF is non-decreasing (for all $p, q, \boldsymbol{h}$)::
   $$C_I^{p,p'}(\boldsymbol{h}) - C_I^{q,q'}(\boldsymbol{h}) \ge qq' - pp' \le 0$$
   Specifically when $p = p'$ and $q = q'$;
   $$C_I^{p,p}(\boldsymbol{h}) - C_I^{q,q}(\boldsymbol{h}) \ge q^2 - p^2 \le 0$$
   $$\gamma_I^{q,q}(\boldsymbol{h}) - \gamma_I^{p,p}(\boldsymbol{h}) \ge q - p \le 0$$
   d. For all $p, q, \boldsymbol{h}$ such that $q - p \le 0$, the following inequality holds true:
   $$-C_I^{p,p}(\boldsymbol{h}) + 2C_I^{p,q}(\boldsymbol{h}) - C_I^{q,q}(\boldsymbol{h}) \le (p-q)^2 \ge 0$$
   $$\gamma_I^{p,p}(\boldsymbol{h}) - 2\gamma_I^{p,q}(\boldsymbol{h}) + \gamma_I^{q,q}(\boldsymbol{h}) \le p - q \ge 0$$

3. Direct and Cross Indicator Covariance Models of Coregionalizations:

Machuka-Mory and Deutsch (2006) showed that Linear Models of Coregionalization (LMC) cannot be fitted to the direct and cross indicator variograms for a Gaussian random function field. This result is mostly because of both hyper-Gaussianity feature of cross indicator variogram (for two extreme low and high cut-offs) and exponentiality behavior of direct indicator variogram (in the case of equal cut-offs). This difference causes the un-fitness of LMC for direct and cross indicator variograms (see Figure 3). Basically LMC states that every variogram structure existing on a cross variogram must also exist in the corresponding direct variograms. For indicator case, the hyper-Gaussianity behavior of cross indicator variogram does not match with the exponentiality behavior of direct indicator variogram.

The theoretical derivation and the bivariate integral formulation of direct and cross indicator variograms which were presented in this paper can be considered as Non-Linear (Analytical) Models of Coregionalizations (NLMC). The derived model cannot be fitted or modeled simultaneously with basic variogram models (i.e.

spherical, Gaussian, Exponential, etc.) simultaneously. The bivariate formulation satisfies the positive semi-definiteness of covariance matrix; therefore this NLMC can be used directly in indicator co-kriging/co-simulation of continuous variables.

Positive semi-definiteness of the full indicator covariance

Characterization of the spatial relationship between the K indicator variables needs a matrix of stationary indicator covariance function, $C_I^{p_k,p_{k'}}(\boldsymbol{h}); k,k' = 1,\dots,K$:

$$C_I(\boldsymbol{h}) = \begin{bmatrix} C_I^{p_1,p_1}(\boldsymbol{h}) & \cdots & C_I^{p_1,p_K}(\boldsymbol{h}) \\ \vdots & \ddots & \vdots \\ C_I^{p_K,p_1}(\boldsymbol{h}) & \cdots & C_I^{p_K,p_K}(\boldsymbol{h}) \end{bmatrix}_{K\times K}, \forall \boldsymbol{h}$$

Theoretically in case of any isofactorial random functions (including Gaussian random function), $C_I(\boldsymbol{h})$ is a symmetric matrix, that is $C_I^{p_k,p_{k'}}(\boldsymbol{h}) = C_I^{p_{k'},p_k}(\boldsymbol{h})$. $C_I(\boldsymbol{h})$ is positive semi-definite for any h, the necessary and sufficient conditions for $C_I(h)$ to be positive semi-definite are, (1) the matrix of $C_I(\boldsymbol{h})$ must be symmetric for any h, (2) The matrix of $C_I(\boldsymbol{h})$ must have nonnegative eigenvalues for any h. Suro Perez (1995) considered that $C_I(\boldsymbol{h})$ is positive definite. He benefited the positive definiteness of $C_I(\boldsymbol{h})$ to create the indicator principal component values and perform indicator principal component kriging (IPCK). The positive semi-definiteness is verified through an example. The steps to show this verification are as below:

1. Consider 100 realizations of a Gaussian random function with an isotropic spherical covariance function:

$$\rho_Y(\boldsymbol{h}) = 1.0 - 1.5\left(\frac{\boldsymbol{h}}{32.0}\right) + 0.5\left(\frac{\boldsymbol{h}}{32.0}\right)^3$$

For each realization;

2. Generate 10 uniform random numbers between 0 and 1 (consider these 10 random numbers as probability cut-offs); $p_1,\dots,p_{10}$
3. For each lag distance
   - I. Calculate the corresponding covariance value from $\rho_Y(\boldsymbol{h})$
   - *II.* For each pair of cut-offs, $\{(p_k,p_{k\prime}); k,k' = 1,\dots,10\}$, calculate the direct and cross indicator covariance $\{C_I^{k,k\prime}(\boldsymbol{h}); k,k' = 1,\dots,10\}$
   - III. Calculate the eigenvalues of $C_I(h)$ (see Numerical Reciepes)

Results with 100 realizations show that all of the eigenvalues are positive definite. Figure 4 shows the histogram for each of the eigenvalues:

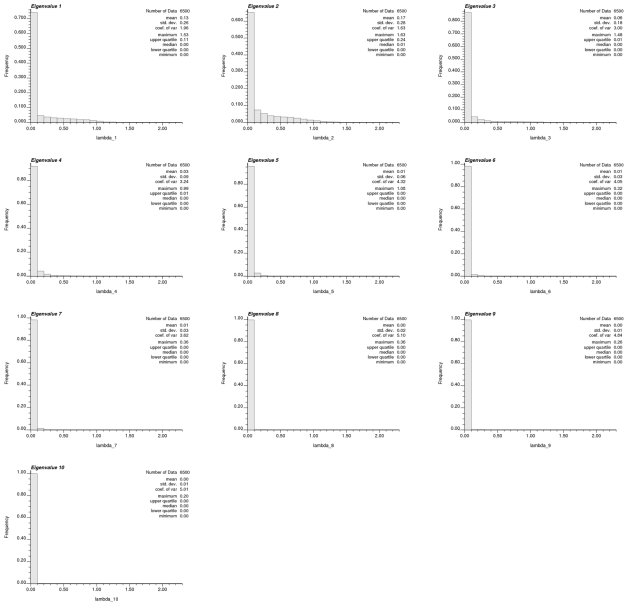

Figure 4 verification of non-negetaiveness of eigenvalues of full indicator covariance function

Before using NLMC in estimation and/or simulation modes, the bivariate Gaussianity should be checked by comparing the experimental direct and cross indicator variograms with the theoretical derivation of $\gamma_I^{p_k,p_{k'}}(\boldsymbol{h})$. As an example, a Gaussian random field is generated using a known single structure spherical variogram (with nugget effect of 5 % and range of 30 % of the field size), the experimental indicator variograms are calculated and are plotted against the theoretical ones from numerical integration.

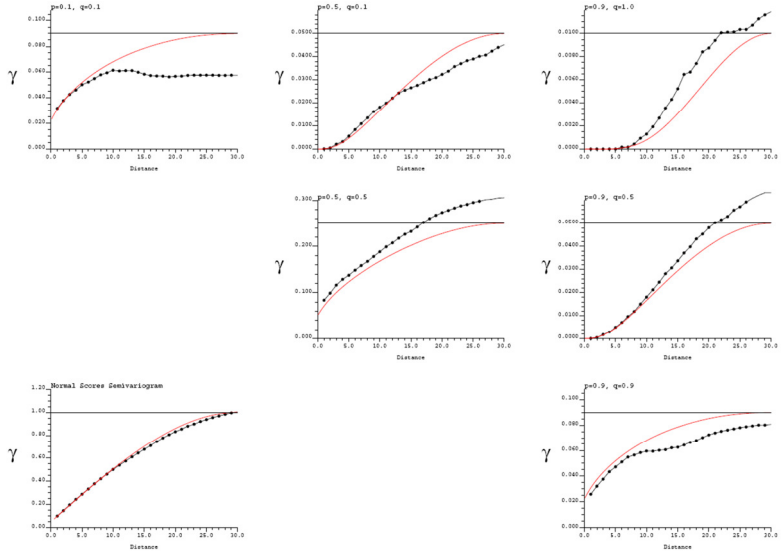

Figure 5 the check for bivariate Gaussianity for a random field

Figure 5 shows the checks for bivariate Gaussianity, although the normal score variogram (bottom left) perfectly reproduces the input original variogram but there is no control on cross and indicator variograms. For generating a perfect Gaussian random field all of the univariate, bivariate,…, multivariate Gaussianity should be checked but it is not possible in reality.

In order to assess uncertainty in both direct and cross indicator variograms at each lag using spatial bootstrap (Deutsch, 2004), a two dimensional Gaussian random field is generated with exponential single structure variogram (with nugget effect of 5 % and range of 12.5 % of the field size) as follows:

1. Performing unconditional LU simulation to get Gaussian random field in a 64 x 64 field using an isotropic exponential variogram with one structure, 5% nugget effect and correlation range of 8.0 (one eighth of the field size). Figure 6 shows the spatial distribution, histogram and variogram reproductions.

$$\gamma_Y(\boldsymbol{h}) = 0.05 + 0.95e^{\frac{-3\boldsymbol{h}}{8}}$$

2. Adding coordinate to the unconditional LUSIM realization, it is necessary to perform Spatial Bootstrap
3. Performing spatial bootstrap to resample data at data locations. Note that all of the simulated points in step 1 are used for resampling. 100 realizations are generated.
4. Calculating experimental normal score variogram, direct and cross indicator variograms for different cut-offs (e.g. 0.1, 0.5 and 0.9) for all of the realizations (see Figure 7).
5. Checking the bivariate plots of the outlier points in matrix of indicator variograms,
6. Table 2 summarizes the outlier point information based on the two extreme cut-offs of 0.1 and 0.9 (the one that is circled in each of the plots in Figure 7) along with exact theoretical values for variograms. The histograms for cross indicator variogram (with cut-offs of 0.9 and 0.5) values at each lag distance are shown in Figure 8.

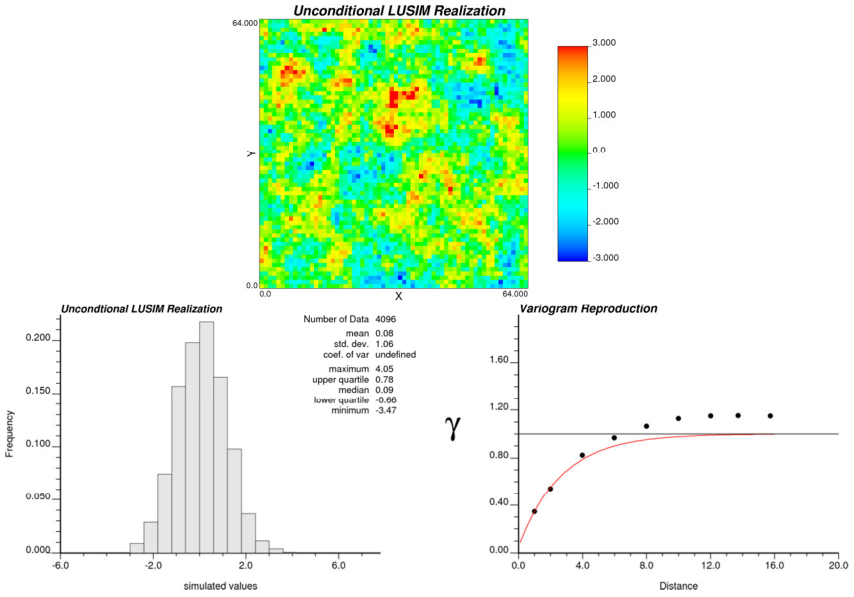

Figure 6 spatial distribution, histogram and variogram reproductions of the synthetic Gaussian random field

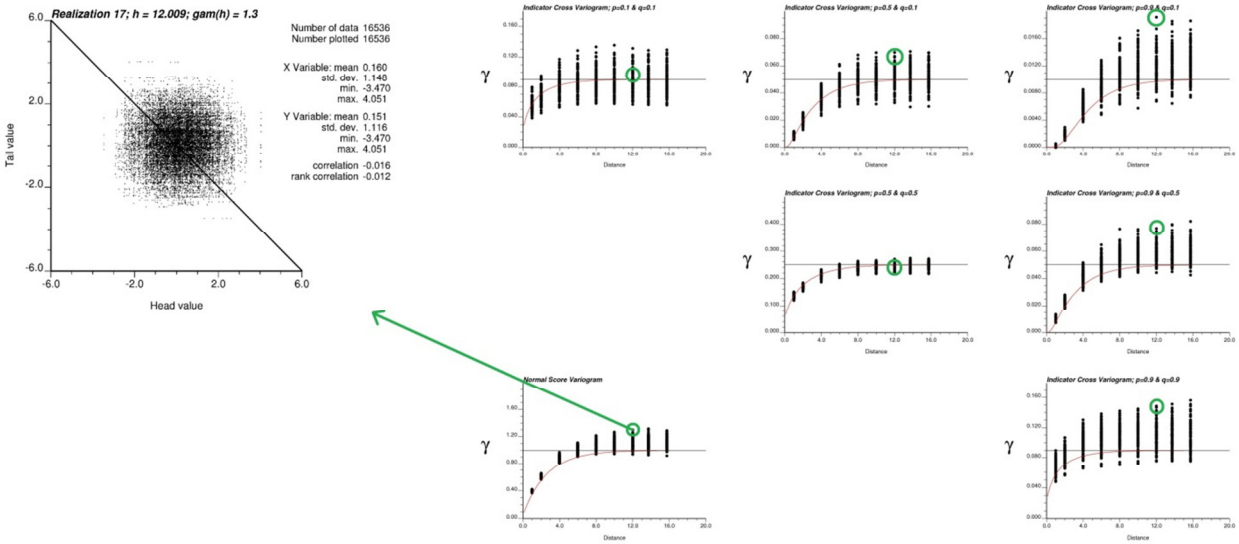

Figure 7 bivariate Gaussianity checks for a Gaussian random field using spatial bootstrap, the scatter plot in the left corresponds to the Head values versus Tail values for the outlier datum

Table 2 outlier datum statistics in bivariate Gaussianity check using spatial bootstrap

h	p	q	Experimental $\hat{\gamma}$	Theoretical
12.009	0.1	0.1	0.09482	0.08697
12.009	0.5	0.1	0.05881	0.04926
12.009	0.9	0.1	0.01191	0.00968
12.009	0.5	0.5	0.24773	0.24833
12.009	0.9	0.5	0.07683	0.04926
12.009	0.9	0.9	0.14257	0.08967
12.009	-- (normal score)	-- (normal score)	1.30189	0.98948

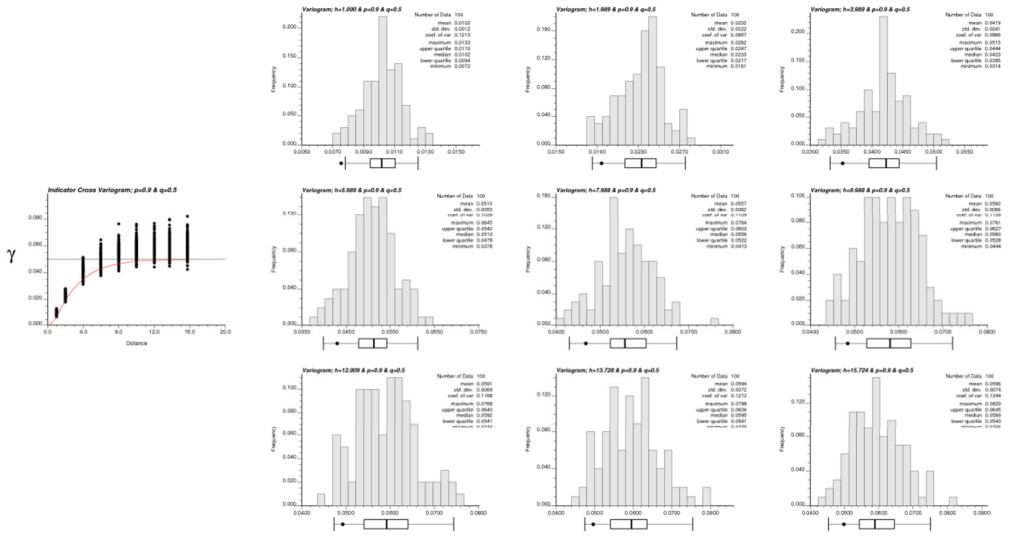

Figure 8 histogram of cross indicator variograms at each lag using spatial bootstrap with cut-offs of 0.9 and 0.5

4. Full Indicator CoKriging

The indicator approach to conditional distribution modeling has applications in both continuous variables (evaluating the conditional cumulative distribution function (ccdf)) and categorical variables (estimating indicators at unsampled locations). Assume the same definition for indicator variable as presented in previous sections. The number of spatial hard data is n and the index for that is α. The general formulism for indicator approach to conditional distribution modeling is as below:

$$F\big(\boldsymbol{u}; y_k|(n)\big) = Prob\big(Y(\boldsymbol{u}) \le y_k|(n)\big) = E\big\{I_{p_k}(\boldsymbol{u})|(n)\big\}$$

The formulism can be written as below in order to derive the indicator cokriging approach:

$$F\left(\boldsymbol{u}; y_{k_0}\middle|(n)\right) = Prob\big(Y(\boldsymbol{u}) \le y_{k_0}|Y(\boldsymbol{u_\alpha}) = y_\alpha, \alpha = 1, \dots, n\big)$$
$$= E\left\{I_{p_{k_0}}(\boldsymbol{u})\middle|I_{p_k}(\boldsymbol{u_\alpha}) = i_{p_k}(\boldsymbol{u_\alpha}), k = 1, \dots, K; \alpha = 1, \dots, n\right\}$$

y_{k_0} is one of the Kcut-offs considered. Indicator kriging (IK) ignores the cross correlation between indicators at different cut-offs and location, this assumption in derivation of indicator kriging can be written as:

$$E\left\{I_{p_{k_0}}(\boldsymbol{u})\middle|I_{p_k}(\boldsymbol{u_\alpha}) = i_{p_k}(\boldsymbol{u_\alpha}), k = 1, \dots, K; \alpha = 1, \dots, n\right\} = E\left\{I_{p_{k_0}}(\boldsymbol{u})\middle|I_{p_{k_0}}(\boldsymbol{u_\alpha}) = i_{p_{k_0}}(\boldsymbol{u_\alpha}), \alpha = 1, \dots, n\right\}$$

It means that in order to estimate conditional distribution using indicator kriging, it is only needed to use the indicator data for the same cut-offs y_{k_0}. In other words, the indicators for the same cut-offs y_{k_0} are more correlated with the indicator that is being estimated than the indicators for other cut-offs. Therefore the indicator kriging approach can be writing as:

$$\big[F\big(\boldsymbol{u}; y_k|(n)\big)\big]^*_{IK} = \sum_{\alpha=1}^{n} \lambda_\alpha^k \cdot i_{p_k}(\boldsymbol{u}_\alpha)$$

With error variance of:

$$Error\ Variance = E\left\{\left[\big[F\big(\boldsymbol{u}; y_k|(n)\big)\big]^*_{IK} - I_{p_k}(\boldsymbol{u})\right]^2\right\} = p_k - 2\sum_{\alpha=1}^{n} \lambda_\alpha^k \cdot \big[C_I^{p_k,p_k}(\boldsymbol{u}_\alpha - \boldsymbol{u}) + p_k^2\big]$$
$$+ \sum_{\alpha=1}^{n}\sum_{\alpha'=1}^{n} \lambda_\alpha^k \cdot \lambda_{\alpha'}^k \cdot \big[C_I^{p_k,p_k}(\boldsymbol{u}_\alpha - \boldsymbol{u}_{\alpha'}) + p_k^2\big]$$

The derivatives of error variance with respect to weights will give normal equations for IK.

Therefore, the IK algorithm needs K direct indicator variograms/covariances in order to calculate the kriging weights. If the assumption of indicator kriging is ignored, then the Indicator cokriging (coIK) is derived:

$$\big[F\big(\boldsymbol{u}; y_k|(n)\big)\big]^*_{coIK} = \sum_{k'=1}^{K}\sum_{\alpha=1}^{n} \lambda_\alpha^{k,k'} \cdot i_{p_{k'}}(\boldsymbol{u}_\alpha)$$

Indicator approach is nonparametric that is the distribution $Y(\boldsymbol{u})$ can have any form (Journel 1983). In order to calculate the indicator cokriging estimates, K^2 direct and cross indicator variograms/covariances are needed. The error variance in coIK approach can be derived as:

$$Error\ Variance = E\left\{\left[\left[F\left(\boldsymbol{u}; y_k\middle|(n)\right)\right]^*_{coIK} - I_{p_k}(\boldsymbol{u})\right]^2\right\} = p_k - 2\sum_{k'=1}^{K}\sum_{\alpha=1}^{n}\lambda_\alpha^{k,k'}\cdot\left[C_I^{p_k,p_{k'}}(\boldsymbol{u}_\alpha - \boldsymbol{u}) + p_k\cdot p_{k'}\right] + \sum_{k=1}^{K}\sum_{k'=1}^{K}\sum_{\alpha=1}^{n}\sum_{\alpha'=1}^{n}\lambda_\alpha^{k,k'}\cdot\lambda_{\alpha'}^{k,k'}\cdot\left[C_I^{p_k,p_{k'}}(\boldsymbol{u}_\alpha - \boldsymbol{u}_{\alpha'}) + p_k\cdot p_{k'}\right]$$

The same as IK, the derivatives of error variance with respect to weights will give normal equations for coIK. The size of the kriging matrix in IK is $n \times n$, and this matrix is evaluated for each of the indicator cut-offs, while the size of the kriging matrix in coIK is $nK \times nK$ (considering both cut-offs and spatial locations in the covariance matrix), and this matrix is evaluated for each of the indicator cut-offs. Indicator cokriging is the best estimate for the conditional distribution which can be found in vector space of estimates generated by hard data (Journel and Alabert, 1989). The materials presented in previous section regarding the direct and cross indicator variograms can be applied here for indicator cokriging.

5. Conclusion

Direct and cross indicator variograms/covariances are discussed for Gaussian random function field, the unique properties for this family of covariance functions are presented. It is impossible to fit a Linear Model of Coregionalization (LMC) to the full covariance matrix of bivariate Gaussian indicators. The full covariance model presented can be treated as an Analytic or Non-Linear model of Coregionalization by itself, as the full covariance matrix is positive semi-definite, but the problem of using it in estimation or simulation modes is hyper Gaussianity (extraordinary continuity) of indicator variograms at extreme cut-offs. The experimental and theoretical direct and cross variograms are also compared to ensure the multi-Gaussian assumption. If the theoretical and experimental do not match well, then the Indicator approach is preferred than the Gaussian approach for simulation. The full indicator cokriging approach to conditional distribution is also presented with simplifications for indicator kriging. The presented full matrix of indicator covariance is used in indicator cokriging approach.

References

Balakrishnan, N. and Lai C. D., 2009, *Continuous Bivariate Distributions*, Springer

Derakhshan, S. H. and Deutsch, C. V., 2011. *Numerical Integration of Bivariate Gaussian Distributions.* Annual Report 13, Centre for Computational Geostatistics (CCG), University of Alberta.

Deutsch, C. V., 2004. *A Statistical Resampling Program for Correlated Data : Spatial_Boostrap.* Annual Report 6, Centre for Computational Geostatistics (CCG), University of Alberta.

Journel, A.G., 1983. *Nonparametric estimation of spatial distributions*. Mathematical Geology 15(3): 445–468.

Journel, A.G., Alabert, F., 1989. *Non-Gaussian data expansion in the Earth Sciences*. Terra Nova 1: 123–134.

Jounel, A.G. and Posa, D., 1990. *Characteristic behaviour and order relations for indicator variograms*. Mathematical Geology, 22(8): 1011-1025

Kyriakidis, P.C., Deutsch, C.V. and Grant, M.L., 1999. *Calculation of the normal scores variogram used for truncated Gaussian lithofacies simulation: theory and FORTRAN code*. Computers & Geosciences, 25: 161-169.

Machuka-Mory, D. F. and Deutsch, C. V., 2006. *Deriving Indicator Direct and Cross Variograms from a Normal Scores Variogram Model (bigaus-full).* Annual Report 8, Centre for Computational Geostatistics (CCG), University of Alberta.

Pardo-Iguzquizaa, E., and Dowd, P. A., 2005. *Multiple indicator cokriging with application to optimal sampling for environmental monitoring*. Computers & Geosciences, 31: 1-13.

Suro-Pe´rez, V., Journel, A.G., 1991. *Indicator principal component kriging*. Mathematical Geology 23 (5): 759–788.