On a Generalized Linear Model of Coregionalization

S. Elogne and O. Leuangthong

The practice of modeling the cross and direct semivariograms together to describe the covariation of p attributes can be a cumbersome task since the models that are fitted must lead to a positive semi definite matrix. This requirement is satisfied when linear model of coregionalization is used for the fitting process. One major limitation of the linear model of coregionalization is the constraint that all direct and cross semivariogram share the same set of basic structures, which can prove time-demanding and can produce a poor fit. In this paper, we generalize the linear model of coregionalization and prove that we can still ensure the positive semi definiteness requirement of the coregionalization matrices without necessarily requiring the same set of basic structures. We provide an example of such model of coregionalizalition where the classic linear model of coregionalization (LMC) cannot be applied. A real data example from the mining is also used for illustration.

Introduction

The widely used linear model of coregionalization (Journel and Huijbregts, 1978) comes to a limitation when one has to select the prior number and the types of the nested structures, which remain somewhat arbitrarily conducted and can cause a poor fit of the linear model of coregionalization. The enforcement that all cross semivariogram models must contain all the nestedstructures involved in all the directs semivariogram models is less realistic although in practice the user copes to justify this assumption by paying an 'human-consuming' effort and a 'time-demanding' process when selecting those priors models (Yao and Journel, 1998). Indeed, in addition to the fact the one attribute could follow from a differentiable process and the other could come from a non-differentiable model, it is unclear why all the attributes should share the same common pool of orthogonal random fields. Some alternatives approaches to overcome some limitations of the linear model of coregionalization have been proposed by several authors (Journel and Yao, 1998), (Vargas-Guzman, Warrick and Myers, 2002) who considered the Bochner's spectral representation of covariance functions.

A practical example in petroleum is that seismic data looks like Gaussian, porosity data looks like Spherical or exponential while the cross seems to follow the seismic data model, that is Gaussian. Another example of possible limitation of the LMC will be in any automatic fitting process with no prior specification of any nested models. For instance consider an automatic fitting approach of all the semivariogram estimates by a flexible model like a Mat\'ern model, where there is no need to consider more than two nested structures. Assuming the estimates are consistently given, there is no argument to justify why the fitted Matern models directs semivariogram should be proportional and that the fitted Matern cross semivariogram model will be expressed as a linear combination of the fitted ones.

The method described in this paper aims to generalize the linear model of coregionalization (LMC) by allowing a more freedom in the modeling process. It does not necessary allow a prior selection of a set of basic structures and may overcome the requirement that all the models presents in the cross semivariogram model should be found in all the direct semivariogram models. We discuss and propose an early approach for checking the positive definiteness property. The main results obtained in the paper include the following: (i) the generalized linear of coregionalization is formulated. (ii) A theoretical example where the classic LMC is limited is given. (iii) A real data example from the mining is carried out to illustrate the approach.

Theoretical Development of the Generalized LMC

Consider a set of p second order random fields $Z_1(\mathbf{u}),\ldots,Z_p(\mathbf{u})$ defined for points $\mathbf{u} \in \mathbb{R}^d$. For covariance or semivariogram modeling purposes the second order stationary and ergodic conditions are commonly assumed for all p random fields. The commonly used linear model of coregionalization (Journel and Huijbretjs, 1978, p. 171) writes each attribute as a linear combination of a same set of spatially orthogonal random fields $Y_1,\ldots,Y_L$, following

$$Z_k(\mathbf{u}) = \sum_{i=1}^{L} a_{k,i} Y_i(\mathbf{u}) \tag{1}$$

yielding the well known linear model of coregionalization (LMC) (Goovaerts, 1997) and all the limitations imported by this approach. In the following we consider a more flexible approach as follows.

- For each attribute Z_k, we consider a set of L orthogonal second-order stationary random fields $Y_{k1}(\mathbf{u}), \ldots, Y_{kL}(\mathbf{u})$, such that

$$\mathbb{C}ov\Big(Y_{ki}(\mathbf{u}), Y_{kj}(\mathbf{u}')\Big) = \delta_{ij}\Sigma_{ij}^{kk}(\mathbf{u}-\mathbf{u}') \tag{2}$$

  where δ_{ij} is the Kronecker symbol which takes the value one if i=j and zero otherwise.

- For two different attributes Z_k and $Z_{k'}$ with $k \neq k'$, the set of orthogonal random fields $\big(Y_{k1}(\mathbf{u}), \ldots, Y_{kL}(\mathbf{u})\big)$ and $\big(Y_{k'1}(\mathbf{u}), \ldots, Y_{k'L}(\mathbf{u})\big)$ are not necessarily the same and moreover are not orthogonal, that is there exists a covariance matrix $\Sigma_{ij}^{kk'}$ such that

$$\mathbb{C}ov\Big(Y_{ki}(\mathbf{u}), Y_{k'j}(\mathbf{u}')\Big) = \delta_{ij}\Sigma_{ij}^{kk'}(\mathbf{u}-\mathbf{u}'). \tag{3}$$

  Thus we write

$$Z_k(\mathbf{u}) = \sum_{i=1}^{L} a_{ki} Y_{ki}(\mathbf{u}) \tag{4}$$

  where the random field Y_{ki} depends on k and has covariance $\Sigma_i^{kk}(\mathbf{h})$. It follows that the covariance between two attributes $Z_k(\mathbf{u})$ and $Z_{k'}(\mathbf{u}+\mathbf{h})$ is given for any h by

$$\begin{aligned} C_{kk'}(\mathbf{h}) &= \mathbb{C}ov\Big(Z_k(\mathbf{u}), Z_{k'}(\mathbf{u}+\mathbf{h})\Big) \\ &= \mathbb{C}ov\left(\sum_{i=1}^{L} a_{ki}Y_{ki}(\mathbf{u}), \sum_{j=1}^{L} a_{k'j}Y_{k'j}(\mathbf{u}+\mathbf{h})\right) \\ &= \sum_{i=1}^{L}\sum_{j=1}^{L} a_{ki}a_{k'j}\ \underbrace{\mathbb{C}ov\Big(Y_{ki}(\mathbf{u}), Y_{k'j}(\mathbf{u}+\mathbf{h})\Big)}_{\delta_{ij}\Sigma_{ij}^{kk'}} \end{aligned}$$

Using Equation (3), we get

$$\begin{aligned} C_{kk'}(\mathbf{h}) &= \sum_{i=1}^{L}\sum_{j=1}^{L} a_{ki}\, a_{k'j}\Big[\delta_{ij}\Sigma_{ij}^{kk'}(\mathbf{h})\Big] \\ &= \sum_{i=1}^{L} a_{ki}\, a_{k'i}\ \Sigma_i^{kk'}(\mathbf{h}). \end{aligned} \tag{5}$$

and

$$C_{kk'}(\mathbf{h}) = \sum_{i=1}^{L} b_i^{kk'}\ \Sigma_i^{kk'}(\mathbf{h}) \tag{6}$$

where $b_i^{kk'} = a_{ki}\, a_{k'j}$. This generalized linear model of coregionalization means that the nested structures involved in the cross-covariance models are not necessary the same those involved in the direct as they depend on k and k'. Thus the direct/cross covariance matrix is as

$$\mathbf{C}(\mathbf{h}) = \sum_{i=1}^{L} \left[b_i^{k,k'} \, \Sigma_i^{kk'}(\mathbf{h}) \right]_{k,k'} \tag{7}$$

Thus $\mathbf{C}(\mathbf{h})$ is positive semi definite for any I, the matrix $M_i(\mathbf{h}) = \left[b_i^{k,k'} \, \Sigma_i^{kk'}(\mathbf{h}) \right]_{k,k'}$ is positive semi definite. We have

$$M_i(\mathbf{h}) = \begin{pmatrix} b_i^{11} & b_i^{12} & \cdots & b_i^{1p} \\ \vdots & \vdots & & \vdots \\ b_i^{p1} & b_i^{p2} & \cdots & b_i^{pp} \end{pmatrix} \bullet \begin{pmatrix} \Sigma_i^{11}(\mathbf{h}) & \Sigma_i^{12}(\mathbf{h}) & \cdots & \Sigma_i^{1p}(\mathbf{h}) \\ \vdots & \vdots & & \vdots \\ \Sigma_i^{p1}(\mathbf{h}) & \Sigma_i^{p2}(\mathbf{h}) & \cdots & \Sigma_i^{pp}(\mathbf{h}) \end{pmatrix} \tag{9}$$

where for two matrices A and B, the product $A \bullet B$ is the entry-wise product of A and B or the Schur/Hadamard product of matrices. It is proved in (Horn and Johnson, 1985, page 458) that a sufficient condition for $A \bullet B$ to be a positive definite matrix is that both A and B are positive semi definite matrices. In the classic linear model of coregionalization, the second matrix B is the same for all the entries, while in this generalization, there are not asked to be the same. In terms of semivariogram, the matrix of coregionalization is as

$$\mathbf{\Gamma}(\mathbf{h}) = \sum_{i=1}^{L} \begin{pmatrix} b_i^{11} & b_i^{12} & \cdots & b_i^{1p} \\ \vdots & \vdots & & \vdots \\ b_i^{p1} & b_i^{p2} & \cdots & b_i^{pp} \end{pmatrix} \bullet \begin{pmatrix} \Gamma_i^{11}(\mathbf{h}) & \Gamma_i^{12}(\mathbf{h}) & \cdots & \Gamma_i^{1p}(\mathbf{h}) \\ \vdots & \vdots & & \vdots \\ \Gamma_i^{p1}(\mathbf{h}) & \Gamma_i^{p2}(\mathbf{h}) & \cdots & \Gamma_i^{pp}(\mathbf{h}) \end{pmatrix} \tag{10}$$

The main rule is that the direct cross semivariogram models should have more nested structures than the cross semivariogram. We do not ask for each nested presents in the cross to be presents in both direct, which is an extension. One important which can be pointed out is checking the positive definite property. One obvious solution can be a numerical approach alongside with the one proposed by Goulard and Voltz (1992). Another approach (probably the simpler) will rely on the fact that only few models (Spherical, Gaussian, Exponential) are the most widely used in any geostatistics modeling. Thus writing a simple algorithm for plotting the determinant as a function of the lag $\mathbf{h}$ using those basics models can be an efficient method for checking the positive definite requirement, unless the practitioner can proceed by hands.

Theoretical Example of the Limitation of the LMC

In this section, we provide a theoretical model example where the linear model of coregionalization cannot be applied. Consider two attributes Z and Y and consider the following model of coregionalization as

$$\begin{cases} \gamma_Z(\mathbf{h}) & = 0.2 + 0.8 \text{ Expo}_8(\mathbf{h}) \\ \gamma_Y(\mathbf{h}) & = 0.3 + 0.7 \text{ Expo}_{12}(\mathbf{h}) \\ \gamma_{ZY}(\mathbf{h}) & = 0.02 + 0.48 \text{ Gaus}_{15}(\mathbf{h}) \end{cases} \tag{11}$$

This constitutes a typical example where the cross semivariogram is modeled through a Gaussian model. The classic LMC cannot be applied to such a model since the cross semivariogram $\gamma_{ZY}(\mathbf{h})$ has a Gaussian structure with a long dependence range, while the two directs semivariogram do not incorporate this Gaussian model. This model coregionalization can be written in a general form as

$$\begin{cases} \gamma_Z(\mathbf{h}) &= a_1 + a_2 \text{ Expo}_A(\mathbf{h}) \\ \gamma_Y(\mathbf{h}) &= b_1 + b_2 \text{ Expo}_B(\mathbf{h}) \\ \gamma_{ZY}(\mathbf{h}) &= c_1 + c_2 \text{ Gaus}_C(\mathbf{h}) \end{cases} \quad (12)$$

where the range A and B which are involved in the direct semivariogram are not necessarily equal. In this case, the GLMC (12) involves three different semivariogram models, where none of the models presents in the direct semivariograms is in the cross semivariogram, except the nugget. System (11) is obviously recovered by taking (A,B,C)=(8,12,15) and (a_1,a_2)=(0.2,0.8), (b_1,b_2)=(0.3,0.7) (c_1,c_2)=(0.02,0.48). Figure (1) gives the plots of all this generalized model of coregionalization.

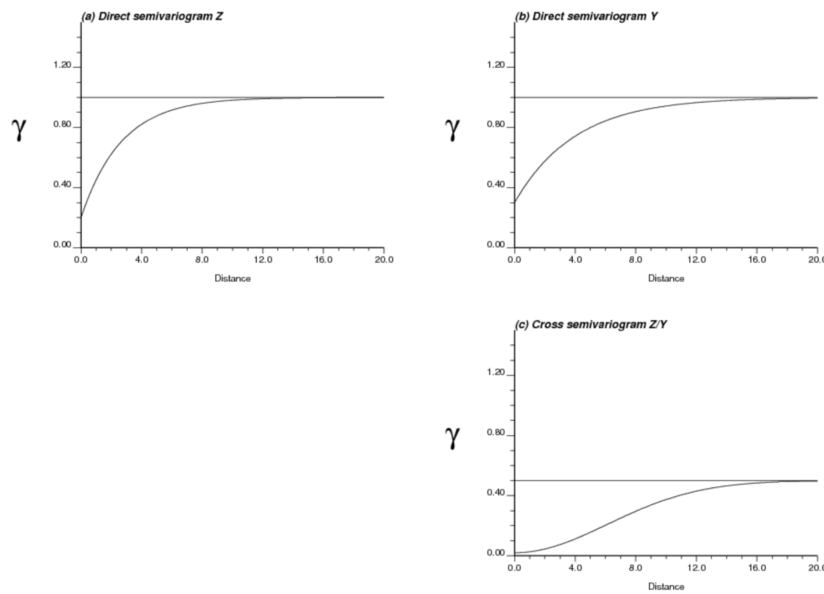

Figure 1: Theoretical example of positive semidefinite generalized linear model of coregionalization where the classical linear model of coregionalization finds itself limited.

Below, we prove that the GLMC given in Equation (12) is positive semi definite. Write that

$$\begin{aligned} \mathbf{\Gamma}(\mathbf{h}) &= \begin{bmatrix} \gamma_Z(\mathbf{h}) & \gamma_{Z,Y}(\mathbf{h}) \\ \gamma_{ZY}(\mathbf{h}) & \gamma_Y(\mathbf{h}) \end{bmatrix} \\ &= \begin{bmatrix} a_1 & c_1 \\ c_1 & b_1 \end{bmatrix} + \begin{bmatrix} a_2 \text{ Expo}_A(\mathbf{h}) & c_2 \text{ Gaus}_C(\mathbf{h}) \\ c_2 \text{ Gaus}_C(\mathbf{h}) & b_2 \text{ Expo}_B(\mathbf{h}) \end{bmatrix} \end{aligned} \quad (13)$$

or equivalently

$$\mathbf{\Gamma}(\mathbf{h}) = \begin{bmatrix} a_1 & c_1 \\ c_1 & b_1 \end{bmatrix} + \begin{bmatrix} a_2 & c_2 \\ c_2 & b_2 \end{bmatrix} \bullet \underbrace{\begin{bmatrix} \text{Expo}_A(\mathbf{h}) & \text{Gaus}_C(\mathbf{h}) \\ \text{Gaus}_C(\mathbf{h}) & \text{Expo}_B(\mathbf{h}) \end{bmatrix}}_{M}. \quad (14)$$

The conventional linear model of coregionalization will instead enforce the above matrix M to have the same entries and will rely on the semi definiteness condition as

$$a_1\, b_1 \geq c_1^2 \quad \text{and} \quad a_2\, b_2 \geq c_2^2. \tag{15}$$

Write that Equation (15) as

$$a_1\, b_1 \geq c_1^2 \quad \text{and} \quad \frac{a_2\, b_2}{c_2^2} \geq 1.$$

Assume the following

1. C > A
2. A > B
3. Consider that

$$\frac{a_2\, b_2}{c_2^2} \geq 2.43 \quad \text{that is} \quad p = \sqrt{\frac{a_2\, b_2}{c_2^2}} \geq 1.55. \tag{16}$$

Let us prove under the inequality (16) that the following matrix

$$\begin{bmatrix} a_2\, \text{Expo}_A(\mathbf{h}) & c_2\, \text{Gaus}_C(\mathbf{h}) \\ c_2\, \text{Gaus}_C(\mathbf{h}) & b_2\, \text{Expo}_B(\mathbf{h}) \end{bmatrix} \tag{17}$$

is positive definite. It is enough to show that

$$a_2 b_2\, \text{Expo}_A(\mathbf{h})\text{Expo}_B(\mathbf{h}) \geq c_2^2 \Big(\text{Gaus}_C(\mathbf{h})\Big)^2. \tag{18}$$

Since A> B, then $\text{Expo}_B(\mathbf{h}) \geq \text{Expo}_A(\mathbf{h})$ for all $\mathbf{h}$. Then to prove Equation (18), it suffices to prove that

$$a_2 b_2 \Big(\text{Expo}_A(\mathbf{h})\Big)^2 \geq c_2^2 \Big(\text{Gaus}_C(\mathbf{h})\Big)^2$$

or equivalently

$$\sqrt{\frac{a_2 b_2}{c_2^2}}\text{Expo}_A(\mathbf{h}) \geq \text{Gaus}_C(\mathbf{h}).$$

Since C>A, it follows that $\text{Gaus}_A(\mathbf{h}) \geq \text{Gaus}_C(\mathbf{h})$. Thus to prove Equation (18), it is sufficient to prove that

$$p\, \text{Expo}_A(\mathbf{h}) \geq \text{Gaus}_A(\mathbf{h}). \tag{19}$$

We have

$$\begin{aligned} r_c &= p\, \text{Expo}_A(\mathbf{h}) - \text{Gaus}_A(\mathbf{h}) \\ &= p\Big(1 - e^{-(3h/A)}\Big) - \Big(1 - e^{-(3h/A)^2}\Big) \end{aligned}$$

If h is such that 3h/A <1, it is clear that $r_c \geq 0$. Now consider the case where 3h/A>1. We get

$$r_c = p\Big(1 - e^{-(3h/A)}\Big) - 1 + e^{-(3h/A)^2}$$

Since 3h/A>1, it follows that

$$p\Big(1 - e^{-(3h/A)}\Big) - 1 > p(1 - e^{-1}) - 1 > 0$$

since by condition p>1, which concludes the proof. Different examples can be provided as valid generalized models of coregionalization. We believe that an efficient way to check for positive definiteness will be for sure a numerical consideration.

Mining Example

A real data example in the mining is considered to illustrate the flexibility of the generalized linear model of coregionalization and the limitation of the LMC and its variant like the two type of Markov models assumptions. In all the examples, the data will be centered and standardized through the empirical statistics. We consider an example from the Jura data set (Goovaerts, 1997) involving the metals Cr and Zn. We use the validation data set with a small sample of size n=100. The generalized linear model of coregionalization is plotted in Figure (2) below. The direct semivariogram of the Chromium looks like Gaussian, the one of the Zinc looks like exponential which a long range while the cross semivariogram looks like exponential with a shorter range realitevly close to the range of the Chrominum. This model then involves three different semivariogram models and thus cannot be modeled by the LMC. The answered is then beard by the generalized linear model of coregionalization. This model is written as

$$\begin{aligned}\mathbf{\Gamma}(\mathbf{h}) &= \begin{bmatrix} \gamma_Z(\mathbf{h}) & \gamma_{Z,Y}(\mathbf{h}) \\ \gamma_{Z,Y}(\mathbf{h}) & \gamma_Y(\mathbf{h}) \end{bmatrix} \\ &= \begin{bmatrix} 0.02 & 0.02 \\ 0.02 & 0.15 \end{bmatrix} + \begin{bmatrix} 0.98\,\mathrm{Gaus}_{0.6}(\mathbf{h}) & 0.46\,\mathrm{Gaus}_{0.5}(\mathbf{h}) \\ 0.46\,\mathrm{Gaus}_{0.5}(\mathbf{h}) & 0.85\,\mathrm{Expo}_{0.75}(\mathbf{h}) \end{bmatrix}. \end{aligned} \tag{20}$$

The generalized model (20) is clearly different from the above example since the first direct semivariogram model is a Gaussian while in the previous case, it was an exponential model. This example is reminiscent to the petroleum example between the seismic and the porosity data. It is important to note that the fitting process of directs and cross semivariogram estimates has been conducted independently. Let's now ensure this model is positive definite. It’s enough to show that

$$(0.98 \cdot 0.85)\,\mathrm{Gaus}_{0.6}(\mathbf{h})\mathrm{Expo}_{0.75}(\mathbf{h}) > (0.46^2)\Big(\mathrm{Gaus}_{0.5}(\mathbf{h})\Big)^2.$$

The ratio $p^2 = (0.98 \cdot 0.85)/(0.46^2)$ is greater that 3.93, that is p>1.98 or again

$$p^2\,\mathrm{Gaus}_{0.6}(\mathbf{h})\mathrm{Expo}_{0.75}(\mathbf{h}) > \Big(\mathrm{Gaus}_{0.5}(\mathbf{h})\Big)^2 \tag{21}$$

It is then sufficient to show that

$$p\,\mathrm{Gaus}_{0.6}(\mathbf{h}) > \mathrm{Gaus}_{0.5}(\mathbf{h}) \quad \text{and} \quad p\,\mathrm{Expo}_{0.75}(\mathbf{h}) > \mathrm{Gaus}_{0.5}(\mathbf{h})$$

The proof can be obtained in the similar way as above by using the inequality p>1.98.

Discussion and Conclusion

A generalized linear model of coregionalization model is proposed as an intension to the linear model of coregionalization. This model is more flexible than the conventional linear model of coregionalization since it does not necessarily required the direct/cross semivariogram models to share the same nested structure models, thus to share the same shape, continuity and differentiability. This approach has the advantage to model 'accurately' directs and cross semivariogram 'independently' before the step of checking the positive definiteness requirement takes place. The only requirement is that only the number of nested structures in the cross semivariogram model should be less or equal than those present in the directs semivariogram models.

In continuing research, we aim to propose a consistent and computational method that automatically ensures the positive definiteness property through the spectral representation of conditionally negative definite functions.

Generalized LMC using 100 sample data

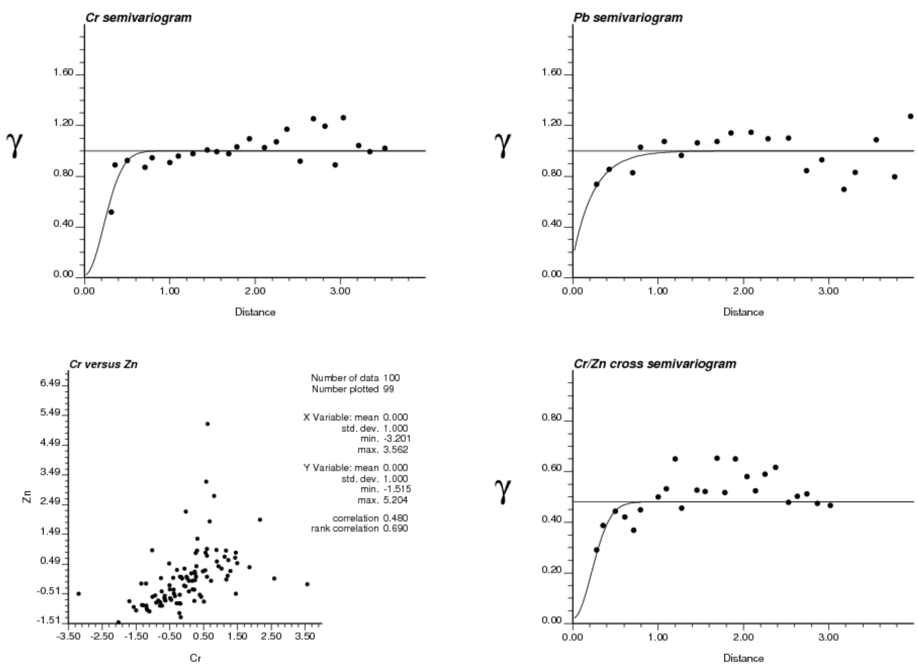

Figure 2: A mining example from the Jura data set using a sample of size n=100. The cross variogram shares the same shape as the primary variable (Chrominum data) but does not incorporate any exponential model which guides the spatial dependence of the secondary data.

References

Bourgault, G., and Marcotte, D., 1991, Multivariate variogram and its application to the linear model of coregionalization: Mathematical Geology, v. 23, no. 7, p. 899-927.

Chiles, J. P., and Delfiner, P., 1999, Geostatistics, modeling spatial uncertainty:Wiley, NewYork, 695 p.

Cressie, N., 1993, Statistics for spatial data: Wiley, New York, 900 p.

Deutsch, C., V. (2002) Geostatistical Reservoir Modeling. Oxford University Press, 2002.

Deutsch C.V. and Journel A.G., 1998, GSLIB: Geostatistical Software Library and User's Guide, 2nd edition, Oxford University Press, New YorkDeutsch, C.V. (2002) *Geostatistical Reservoir Modeling*. Oxford University Press, New York, 376 p.

Kunsch, H. R., Papritz, A., and Bassi, F., 1997, Generalized cross covariances and their estimation. Mathematical Geology, v. 29, no. 6, p.779-799.

Journel, A. and Huijbregts, J., 1978, Mining Geostatiscis, London: Academic Press, 600 p.

Goulard, M. and Voltz, M., Linear coregionalization model: tools for estimation and choice of crossvariogram matrix. Mathematical Geology, v. 24, no. 3, p.269-286.

Goovaerts, P., 1997, Geostatistics for natural resources evaluation: Oxford University Press, New York, 500 p.

Goulard, M., 1989, Inference in a coregionalization model, in Armstrong, M., ed., Geostatistics, vol. 1: Kluwer Academic, Dordrecht, p. 397–408.

Horn, R. and Johnson, C. (1985). Matrix Analysis. Cambridge University Press.

Matheron, G., The intrinsic random functions and their applications, Adv. Appl. Probab., 5 (1973), p. 437-468

Myers, D. E., 1991, Pseudo cross-variograms, positive-definiteness, and cokriging: Mathematical Geology, v. 23, no. 6, p. 805-816.

Myers, D. E., 1982, Matrix formulation of co-kriging: Math. Geol., v. 14, no. 3, p. 249–257.

Myers, D. E., 1983, Estimation of linear combinations and co-kriging: Mathematical Geology, v. 15, no. 5, p. 633-637

Vargas-Guzman, J., A., Warrick, A., W., and Myers, D., E., 2002, Coregionalization by Linear Combinaison of Nonorthogonal Components: Mathematical Geology, v. 34, no. 5, p. 406-419.

Yao, T., and Journel, A. G., 1998. Automatic modeling of (cross) covariance tables using fast fourier transform: Math. Geol. v. 30, no. 6, p. 589–616.

Zimmerman, D. L., and Zimmerman, M. B., 1991, A comparison of spatial semivariogram estimators and corresponding ordinary kriging predictors: Technometrics, v. 33, no. 1, p. 77–91.