Comparison of Cokriging with an LMC versus the Intrinsic Model

Olena Babak and Clayton V. Deutsch

Centre for Computational Geostatistics
Department of Civil & Environmental Engineering
University of Alberta

The major obstacle to extensive use of cokriging for integration of multiple data types in estimation is the requirement of modeling a positive definite covariance matrix of size K by K including up to K^2 *different terms if K coregionalized variables are considered. In practice, this matrix is modeled using a linear model of coregionalization (LMC); each covariance is modeled by a different linear combination of the same basic covariance functions. Oftentimes, sampling of the secondary variable is much more extensive than that of the primary variable, often, the secondary variable is exhaustively sampled (sampled at each node where the primary variable is to be estimated). In such cases, only the secondary data at the estimation location ('colocated' value) could be retained in estimation. The underlying Markov-type hypothesis is that the collocated primary data screens out all further away primary information. Under Markov hypothesis, all that is needed to perform estimation is the primary covariance function and the correlation coefficient between primary and secondary data. An intrinsic correlation model (ICM) was proposed for modeling the covariance matrix. The intrinsic correlation model is obtained when the direct and cross covariance functions are all proportional to the same underlying spatial correlation function. Although ICM appears similar to a Markov models, it makes greater use of the secondary data and does not result in variance inflation. This paper is aimed at comparing and analyzing different correlation models, that is, linear model of coregionalization, intrinsic correlation model and Markov model based on a small comparative example.*

Introduction and Recalls

Simple Cokriging (CSK) is a natural extension of Simple Kriging to the case when multivariate data is available (Chiles and Delfiner, 1999). Simple Cokriging allows estimating an unknown value at the location of interest not only from data of the same type, but also based on the auxiliary variables in the neighborhood. Specifically, the Simple Cokriging estimator is the following weighted linear combination of the mean of the variable of interest (m_*) and the data from different variables located at sample points in the neighborhood of the estimation location $u*$

$$Z*_{CSK}(u*) = m_* + \sum_{i=1}^{N}\sum_{\alpha=1}^{n_i}\lambda_\alpha^i (Z_i(u_\alpha) - m_i),. \tag{1}$$

where the CSK weights $[\lambda_1^T,\ldots,\lambda_N^T]^T$ are found from a Simple Cokriging system given by

$$\begin{pmatrix} C_{11} & \cdots & C_{1j} & \cdots & C_{1N} \\ \vdots & \ddots & & & \vdots \\ C_{i1} & \cdots & C_{ii} & & C_{iN} \\ \vdots & & & \ddots & \vdots \\ C_{N1} & \cdots & C_{Nj} & \cdots & C_{NN} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_i \\ \vdots \\ \lambda_N \end{pmatrix} = \begin{pmatrix} c_{1*} \\ \vdots \\ c_{i*} \\ \vdots \\ c_{N*} \end{pmatrix}, \qquad 2)$$

where the left hand side covariance matrix is built up with square symmetric n_i by n_i blocks C_{ii} on the diagonal and with rectangular n_i by n_j blocs C_{ij} off the diagonal, with

$$C_{ij} = C_{ji}^{T}.$$

The blocks C_{ij} contain either direct ($i = j$) or cross ($i \neq j$) covariances between sample points. The vectors c_{i*} contain the covariances with the variable of interest, for a specific variable of the set, between sample points and the estimation location. The vectors λ_i represent the weight attached to the sample of the *i*-th variable.

We can see from system (1)-(2) that in order to perform Simple Cokriging we require a joint model for the matrix of covariance functions. In what follows we describe three possible types of models for the matrix of covariance functions, that is, linear model of coregionalization (LMC), intrinsic correlation model (ICM) and Markov models of coregionalization (MM I and MM II).

Linear Model of Coregionalization (LMC)

The Linear model of coregionalization is the most extensively used model for describing joint spatial continuity of two or more random variables. The LMC for the two random variables is given by the following system (Deutsch, 2002)

$$\begin{aligned}
\gamma_{Z,Z}(h) &= b_{Z,Z}^{0} + b_{Z,Z}^{1}\Gamma^{1}(h) + b_{Z,Z}^{2}\Gamma^{2}(h)\ldots \\
\gamma_{Z,Z}(h) &= b_{Z,Y}^{0} + b_{Z,Y}^{1}\Gamma^{1}(h) + b_{Z,Y}^{2}\Gamma^{2}(h)\ldots \\
\gamma_{Y,Y}(h) &= b_{Y,Y}^{0} + b_{Y,Y}^{1}\Gamma^{1}(h) + b_{Y,Y}^{2}\Gamma^{2}(h)\ldots
\end{aligned} \quad 3)$$

where the Γ^{i}, $i = 1, \ldots, n_{st}$, are nested structures made up from the common pool of variogram models (spherical, exponential, etc.). The linear model of corregionalization models each direct and cross variogrma with the same variogram nested structures, but the sill (contribution) parameters are allowed to change so that the following constraints are satisfied,

$$\left.\begin{aligned}
& b_{Z,Z}^{i} \geq 0 \\
& b_{Y,Y}^{i} \geq 0 \\
& b_{Z,Z}^{i} b_{Y,Y}^{i} \geq b_{Z,Y}^{i} b_{Z,Y}^{i}
\end{aligned}\right\} \forall i \quad (4)$$

The linear model of coregionalization (3) can be extended to any number of variables, however, it is rarely applied to large number of random variables.

Intrinsic Correlation Model (ICM)

Intrinsic correlation model is the simplest multivariate covariance model that can be adopted for a covariance function matrix. It describes the relationships between variables by the variance-covariance matrix **V** and the relations between points in space by a spatial correlation function $\rho(h)$ as follows (Wackernagel, 2003)

$$C(h) = V\rho(h). \quad (5)$$

Note that the spatial correlation function $\rho(h)$ is the same for all variables. The model (5) is called the intrinsic correlation model due to its property that the correlation between any two variables ρ_{ij} is independent of the spatial scale, that is,

$$\frac{\sigma_{ij}\rho(h)}{\sqrt{\sigma_{ii}\rho(h)\sigma_{jj}\rho(h)}} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}\rho_{ij}. \quad (6)$$

In practice the intrinsic correlation model is obtained when direct and cross covariance functions are all proportional to the same underlying spatial correlation function,

$$C_{ij}(h) = b_{ij}\rho(h), \tag{7}$$

where the coefficients b_{ij} represent the variances (i = j) and covariances ($i \neq j$) between variables. Similarly to (7), an intrinsic model can be defined in terms of variograms. Specifically, the intrinsic correlation model is a product of a positively coreligionalization matrix B of coefficients b_{ij} and a direct variogram $\gamma(h)$, that is

$$\Gamma(h) = B\gamma(h). \tag{8}$$

When comparing a linear model of correlation (3) with an intrinsic correlation model, we can observe that the IMC can be viewed as a limit of the LMC. This is because when direct and cross covariance functions of the variables under consideration are all proportional to the same underlying spatial correlation function the linear model of correlation reduces to the intrinsic correlation model.

Collocated Simple Cokriging

In order to perform Simple Cokriging we require a joint model for the matrix of covariance functions, Linear Model of Coregionalization. Thus, when *K* different variables are considered, the covariance matrix in the left hand side of Simple Cokriging equation (2) requires K^2 covariance functions. Such inference is very demanding in terms of data and subsequent joint modeling, therefore, more simple estimation technique called Collocated Simple Cokriging handling multiple data variables is frequently employed instead.

Collocated Simple Cokriging is a strategy in which the neighborhood of the auxiliary variable is reduced to only one point at the estimation location. This value of the auxiliary variable $Y(u^*)$ is said to be collocated with the variable of interest $Z(u)$ at the estimation location *u**. The Collocated Simple Cokriging estimator is given by (Goovaerts, 1997)

$$Z^*_{CCSK}(u^*) = m_Z + \lambda_0(Y(u^*) - m_Y) + \sum_{\alpha=1}^{N} \lambda_\alpha (Z(u_\alpha) - m_Z), \tag{9}$$

where Collocated Simple Cokriging weights $[\lambda_Z^T \ \lambda_Y]^T$ are found from the following system of equations

$$\begin{pmatrix} C_{ZZ} & c_{YZ} \\ c_{YZ}^T & \sigma_{YY} \end{pmatrix} \begin{pmatrix} \lambda_Z \\ \lambda_Y \end{pmatrix} = \begin{pmatrix} c_{ZZ} \\ \sigma_{YZ} \end{pmatrix}, \tag{10}$$

where C_{ZZ} is the left hand matrix of the Simple Kriging system of $Z(u)$ and c_{ZZ} is the corresponding right hand side covariance vector. The vector c_{YZ} contains the cross covariances between the *n* sample points of $Z(u)$ and the estimation location *u** with its collocated value $Y(u^*)$.

Markov Model I

The Markov Model I assumes that the primary *Z* data prevails over the secondary *Y* data. Formally, it can be written as (Almeida and Journel, 1996)

$$E\{Y(u) \mid Z(u) = z, Z(u+h) = z'\} = E\{Y(u) \mid Z(u) = z\}, \quad \forall h, \forall z'. \tag{11}$$

(That is, dependence of the secondary variable on the primary is limited to the collocated primary datum). The cross covariance C_{ZY} under the Markov model I is given by

$$C_{YZ}(h) = \sigma_{YZ} r_Z(h), \tag{12}$$

where σ_{YZ} denotes the covariance between *Z* and *Y*; and r_Z is the vector of spatial correlations $\rho(u_\alpha - u_0), \quad \alpha = 1, \ldots, n.$

Using the Markov model I for the cross covariance c_{YZ}, we can rewrite the Collocated Simple Cokriging system (4) for the weights $[\lambda_Z^T \ \lambda_Y]^T$ as

$$\begin{pmatrix} \sigma_{ZZ} R_Z & \sigma_{YZ} r_Z \\ \sigma_{YZ} r_Z^T & \sigma_{YY} \end{pmatrix} \begin{pmatrix} \lambda_Z \\ \lambda_Y \end{pmatrix} = \begin{pmatrix} \sigma_{ZZ} r_Z \\ \sigma_{YZ} \end{pmatrix}, \tag{13}$$

where R_Z is the matrix of spatial correlations $\rho(u_\alpha - u_\beta), \quad \alpha, \beta = 1, \ldots, n.$

In order to to perform the collocated cokriging with Markov model I, we only need to know the covariance function

$$C_{ZZ}(h) = \sigma_{ZZ} \rho(h),$$

the variance σ_{YY} of the auxiliary variable and the correlation coefficient $\rho_{YZ} = \rho_{YZ}(0)$. Retaining only the collocated secondary data, in general, does not affect the resulting estimate, since the close neighborhood data are usually very similar in values. However, it may affect the Cokriging estimation variance. Cokriging variances are overestimated, oftentimes significantly. This causes serious problem in sequential simulation (Deutsch, 2002).

Markov Model II

The Markov Model II assumes that the secondary *Y* data prevails over the primary *Z* data. Formally, it can be written as.

$$E\{Z(u) \mid Y(u) = y, Y(u+h) = y'\} = E\{Z(u) \mid Y(u) = y\}, \quad \forall h, \forall y' \tag{14}$$

(That is, dependence of the primary variable on the secondary is limited to the collocated secondary datum). The cross covariance C_{ZY} under the Markov model II is given by

$$C_{ZY}(h) = \sigma_{ZY} r_Y(h), \tag{15}$$

where σ_{ZY} denotes the covariance between *Z* and *Y*; and r_Z is the vector of spatial correlations $\rho(u_\alpha - u_0), \quad \alpha = 1, \ldots, n.$

Markov Models and Intrinsic Model of Correlation

It is interesting to note that despite both Markov models and Intrinsic correlation model assume that the cross covariance and crosscovariogram between primary and secondary variables are proportional to the spatial correlation function of the primary random variable (Markov Model I or secondary random variable in Markov model II), there is a significant difference between IMC and Markov models. On the contrary to ICM, Markov models do not lead to the linear model of coregionalization(LMC). This is because there is no assumption on the continuity of the other variable in Markov models. Intrinsic correlation model put more assumptions on the continuity of the variables under study; however, it results in better models of heterogeneity and no variance inflation (Babak and Deutsch, 2007).

Example

Let us consider the following Linear Model of Coregionalization (LMC) for the primary unit variance, zero mean random variable *Z* and secondary unit variance, zero mean random variable *Y*

$$\begin{aligned}\gamma_{YY}(h) &= 0.1 \cdot Sph_{16}(h) + 0.9 \cdot Gaus_{32}(h) \\ \gamma_{YZ}(h) &= 0.25 \cdot Sph_{16}(h) + 0.25 \cdot Gaus_{32}(h) \ , \\ \gamma_{ZZ}(h) &= 0.9 \cdot Sph_{16}(h) + 0.1 \cdot Gaus_{32}(h)\end{aligned} \tag{16}$$

where $Sph_{16}(h), Gaus_{32}(h)$ denote the Spherical variogram model with the range of 16 and Gaussian variogram model with the range of 32. The correlation at lag 0 between primary and secondary random variables can be calculated under stationarity as

$$\rho_{YZ} = 1 - \gamma_{YZ}(0) = 1 - [0.25 \cdot Sph_{16}(0) + 0.25 \cdot Gaus_{32}(0)] = 1 - 0.25 - 0.25 = 0.5 \, .$$

Now let us consider estimation of the domain 50 by 50 units based on the primary data and exhaustive secondary data. Figure 1 shows locations of 12 primary data and their distribution, the crossplot between primary data and collocated secondary data and the map of exhaustive secondary data. Three approaches for estimation are considered:

- the Simple Cokriging with the linear model of corregionalization (16);
- the Simple Cokriging with the intrinsic correlation model. The primary variable variogram of system (16) is taken as the underlying variogram for the intrinsic correlation model;
- and finally using Collocated Simple Cokriging with MMI.

For the fair comparison, the following parameters were used in estimation: all primary data were applied in estimation, in Simple Cokriging secondary data was assembled at the same locations as the primary data and at the estimation data location.

Results: Weights profiles

Let us first perform estimation of the 2 arbitrary locations in the study domain, say (10,10) and (35, 35), based on Simple Cokriging and Collocated Simple Cokriging and analyze the difference in profiles of Cokriging weights. Figures 2 and 3 for each of the two locations of interest show the estimation variances, accumulated weights, and the cokriging weights profiles. From both Figures we can clearly see that Intrinsic correlation model assigns the collocated secondary data weight equal to the correlation coefficient between primary and secondary data, the largest weight assigned to the collocated secondary data is obtained in Simple Cokriging with LMC and the smallest in the Collocated Simple Cokriging with Markov Model I. Note also that despite the LMC assigns the largest weight to the collocated data, the accumulated weight assigned to all secondary data is obtained in Simple Cokriging with LMC is the smallest, the largest accumulated weight assigned to secondary data (one - collocated) is obtained in Collocated Simple Cokriging. The largest accumulated weight assigned to the primary data is obtained using Linear Model of Correnalization; the smallest is obtained using Markov Model. Intrinsic Correlation model provides an intermediate case in-between the other two correlation models also in terms of estimation variance. The smallest estimation variance is obtained in Simple Cokriging with LMC, the largest in Collocated Simple Cokriging.

Results: Estimation

Now let us consider estimation of the entire domain of study. Figure 4 shows the maps of estimates (means of the local conditional distributions) and estimation variances (variances of the local conditional distributions) obtained based on the Collocated Simple Cokriging with Markov model I, Simple Cokriging with the intrinsic correlation model and Simple Cokriging with the linear model of corregionalization. From Figure 4 we can note that result of Collocated Simple Cokriging with Markov model I and Simple Cokriging with the intrinsic correlation model estimation are very similar; Simple Cokriging with the linear model of corregionalization results in more unsmooth estimates than obtained by the other two approaches and in smaller estimation variance. To assess further difference, the maps of the difference between Collocated Simple Cokriging with Markov model I and Simple Cokriging with the intrinsic correlation model and the difference between Simple Cokriging with the linear model of corregionalization and Simple Cokriging with the intrinsic correlation model are prepared for the estimates and their estimation variances. Figure 5 shows results. From Figure 5 we can confirm that indeed Collocated Simple Cokriging with Markov model I and Simple Cokriging with the intrinsic correlation model results in very similar estimates. The estimation variances are also close for these two methods, however Simple Cokriging with the intrinsic correlation model is always the same or slightly smaller than that of Collocated Simple Cokriging with Markov model. From Figure 5 we can also note that Simple Cokriging with the linear model of corregionalization can result in significantly different estimates and, moreover, the estimation variance of Simple Cokriging with the linear model of corregionalization is always the same or slightly smaller than the estimation variance of Simple Cokriging with the intrinsic correlation model. That is, the local conditional ditributions of uncertainty obtained by the Simple Cokriging with the linear model of corregionalization are almost always narrower than the local conditional distributions obtained in Simple Cokriging with the intrinsic correlation model. Therefore, we can see that there is a trade off between simplicity of the Simple Cokriging with the intrinsic correlation model and width of the local uncertainty distributions.

Conclusions

In this paper a comparative study of Simple Cokriging with three different correlation models were considered: Linear Model of coregionalization, Intrinsic Correlation Model and Markov model were applied in cokriging to obtained estimates and estimation variances for a small domain under study.

References

Almeida, A. S., 1993, Joint simulation of multiple variables with a Markov-type coregionalization model: Doctoral dissertation, Stanford University, Stanford.

Almeida, A. S., and Journel, A. G., 1996, Joint simulation of multiple variables with a Markov-type coregionalization model. Mathematical Geology, 26.

Babak, O., and Deutsch, C.,V., 2007. Solution to the Variance Inflation Problem in Collocated Cokriging, CCG Report 9.

Chiles, J.P. and Delfiner, P., 1999. Geostatistics, *Modeling Spatial Uncertainty*.

Deutsch, C.V., *Geostatisticaql Reservoir Modeling*, 2002.

Goovaerts, P., 1997, Geostatistics for Natural Resources Evaluation: Oxford Press, New York.

Journel, A. G., 1999, Markov Models for Cross-Covariances. Mathematical Geology, 31.

Journel, A. G., and Huijbregts, 1978, Mining Geostatistics: Academic Press, New York.

Wackernagel, H., 2003 *Multivariate Geostatistics*.

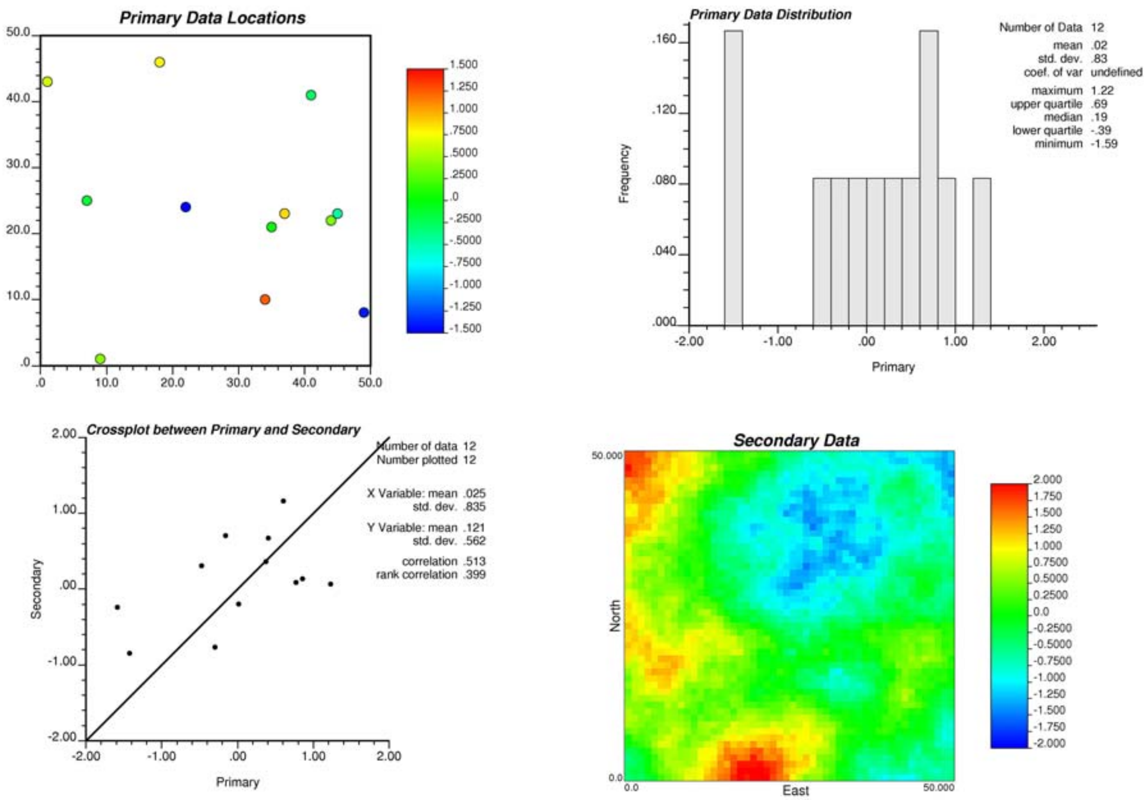
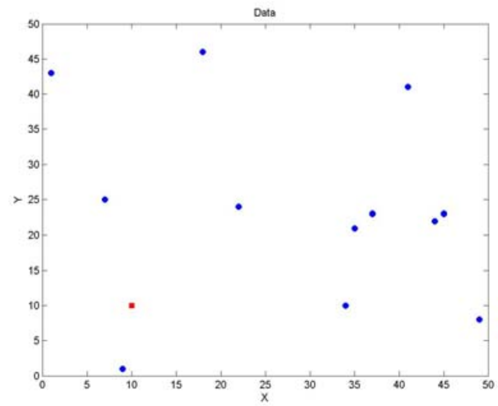
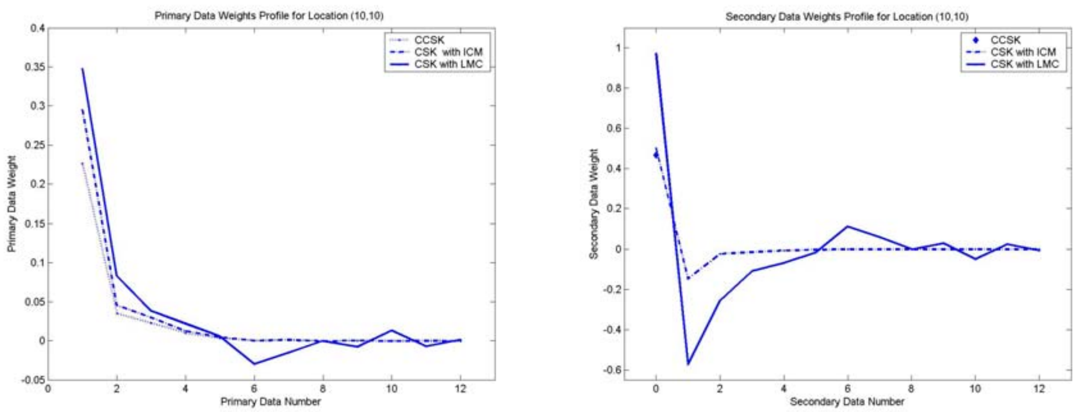


Figure 1: Locations of 12 primary data (top left) and their distribution (top right), the crossplot between primary data and collocated secondary data (bottom left) and the map of exhaustive secondary data (bottom right).



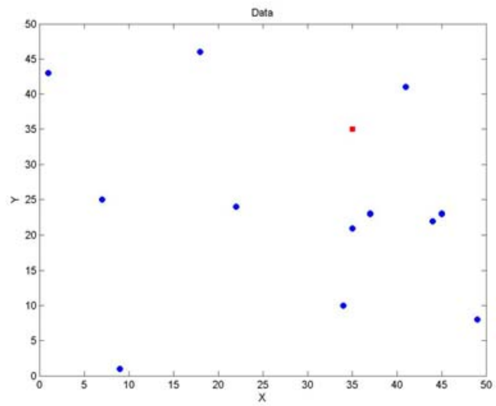
a)



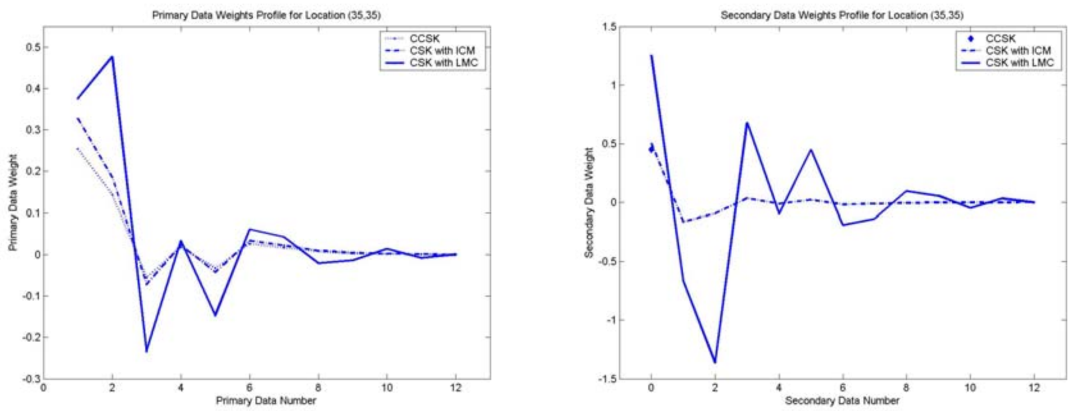
	Sum of Primary Variable Weights	Sum of Secondary Variable Weights	Weight received by Collocated Data	Sum of Cokriging Weights	Estimate	Estimation Variance
CCSK	0.2974	0.4650	0.4650	0.7625	0.2528	0.6976
CSK (Intrinsic)	0.3875	0.3062	0.5000	0.6938	0.1735	0.6817
CSK (LMC)	0.4511	0.1246	0.9715	0.5757	-0.0331	0.5961

b)

Figure 2: a) Study domain with a) conditioning data (circles) and the estimation location (10, 10) (square); Primary data weights as a function of the ordered conditioning data, ordered according to the closeness to the estimation location (first column) and secondary data weights as a function of the ordered conditioning data, zero stands for the estimation location (second column).



a)



	Sum of Primary Variable Weights	Sum of Secondary Variable Weights	Weight received by Collocated Data	Sum of Cokriging Weights	Estimate	Estimation Variance
CCSK	0.3824	0.4497	0.4497	0.8321	-0.5751	0.6746
CSK (Intrinsic)	0.4933	0.2534	0.5000	0.7466	-0.5270	0.6527
CSK (LMC)	0.5789	0.0683	1.2575	0.6472	-1.0345	0.5153

b)

Figure 3: a) Study domain with a) conditioning data (circles) and the estimation location (35, 35) (square); Primary data weights as a function of the ordered conditioning data, ordered according to the closeness to the estimation location (first column) and secondary data weights as a function of the ordered conditioning data, zero stands for the estimation location (second column).

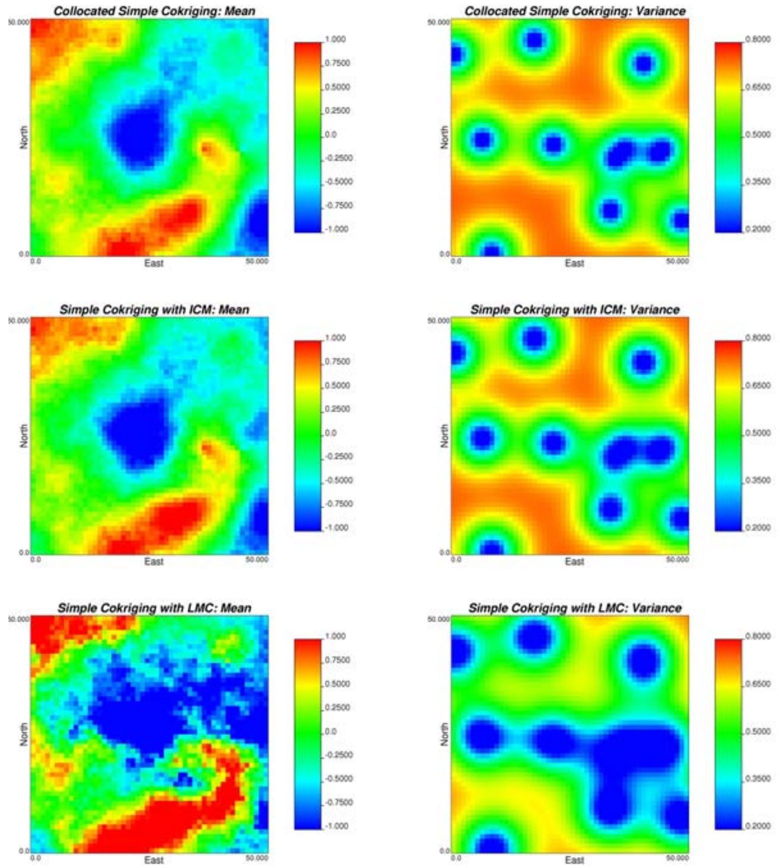


Figure 4: The maps of estimates (left) and estimation variances (right) obtained based on the Collocated Simple Cokriging with Markov model I (top), Simple Cokriging with the intrinsic correlation model (middle) and Simple Cokriging with the linear model of corregionalization (bottom).

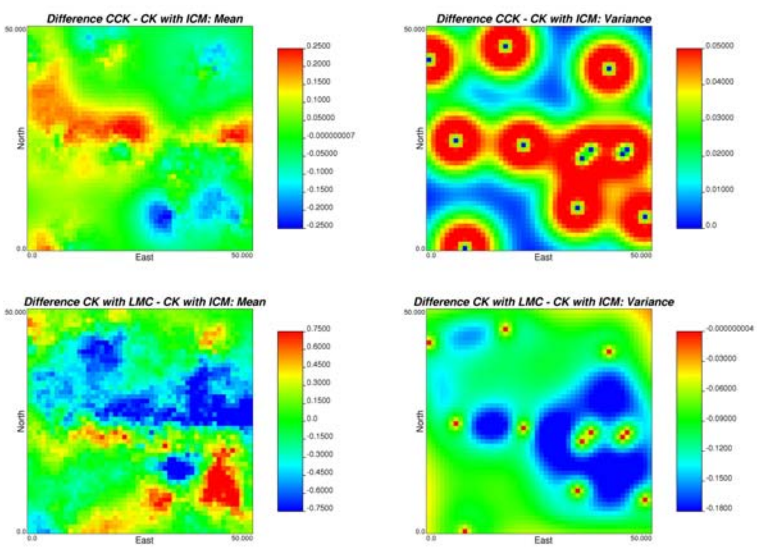


Figure 5: The maps of the difference in means (left) and variances (right) for Collocated Simple Cokriging with Markov model I and Simple Cokriging with the intrinsic correlation model (top) and for Simple Cokriging with the linear model of corregionalization and Simple Cokriging with the intrinsic correlation model (bottom).