Merging Multiple Secondary Data for Collocated Cokriging (recall of *super secondary variable*)

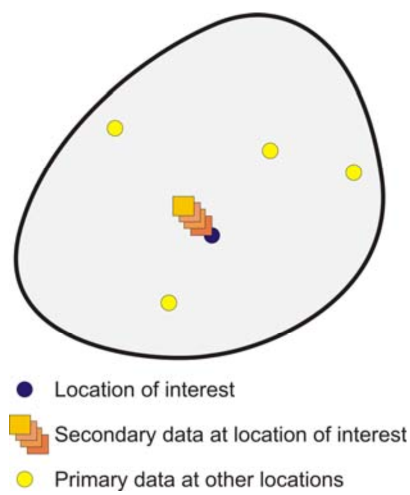
Olena Babak and Clayton V. Deutsch

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

Most collocated cokriging (and cosimulation) programs are based on the use of one secondary variable; however, it is common to have multiple secondary data. Under a multivariate Gaussian model, multiple secondary data can be merged into one secondary variable for use in standard cosimulation software. The idea is presented with examples. The super secondary variable must be restandardized to unit variance and the correct correlation coefficient calculated. The theoretical validity of the approach is established.

Problem Setting

The context of this note relates to estimating a primary variable in presence of multiple secondary data. We could be applying the procedure in a sequential fashion, that is, a primary variable in one pass of simulation could be a secondary variable in a second step. In this context, we could consider n_{sec} secondary data consisting of exhaustively sampled secondary data and previously simulated primary data and one primary variable that is being predicted at the current step. The secondary data are denoted $y_{s,i}$, $i=1,\ldots,n_{sec}$ (the colored squares in the schematic to the right). The primary variable is denoted y_0 or y. We are interested in predicting the uncertainty in y at a location of interest (the blue circle in the schematic to the right). There are a number of primary data at other locations $y(\mathbf{u}_i)=y_i$, $i=1,\ldots n$ (the yellow circles in the schematic to the right). What is not communicated in the schematic to the right is that there are secondary data at all locations; however, we only use the collocated ones when considering a location of interest.



All secondary variables and the primary variable are standardized with means of 0 and variances of 1, which is the case when we adopt a multivariate Gaussian model or if we simply scale the data according to $y=(z-m)/\sigma$. We are working within the classic collocated cokriging paradigm, that is, we only use secondary data at the location being estimated and we adopt an intrinsic model for the cross covariances, that is, the cross covariances are proportional to the covariance structure of the primary variable. We could be estimating at the unsampled location, predicting uncertainty in `PostMG` fashion or, more commonly, using the conditional distribution for simulation in a sequential Gaussian paradigm.

Collocated Cokriging

Collocated cokriging can be implemented with multiple secondary data. The form of the estimator and estimation variance:

$$y^* = \sum_{i=1}^{n} a_i \bullet y_i + \sum_{i=1}^{n_{sec}} b_i \bullet y_{s,i} \quad (1)$$

$$\sigma_K^2 = 1 - \sum_{i=1}^{n} a_i \bullet \rho_{i,0} - \sum_{i=1}^{n_{\sec}} b_i \bullet \rho_{s_i,0} \tag{2}$$

The a's and b's are weights applied to the primary and secondary data, respectively. The ρ values are correlation coefficients between each primary and secondary data and the primary variable at the location being estimated. The secondary data are at the location being estimated and the primary data are at other locations. The estimate and estimation variance are, of course, the mean and variance of the conditional distribution in a multivariate Gaussian context. The equations to compute the $n+n_{sec}$ weights:

$$\begin{aligned} &\sum_{j=1}^{n} a_j \bullet \rho_{j,i} - \sum_{j=1}^{n_{\sec}} b_j \bullet \rho_{s_j,i} = \rho_{i,0} \qquad i = 1,\ldots,n \\ &\sum_{j=1}^{n} a_j \bullet \rho_{s_i,j} - \sum_{j=1}^{n_{\sec}} b_j \bullet \rho_{s_j,s_i} = \rho_{s_i,0} \quad i = 1,\ldots,n_{\sec} \end{aligned} \tag{3}$$

These equations can be shown in matrix form (below). Note the parts of the system measuring redundancy between data sources (left) and closeness to what is being estimated (right).

$$\begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ n \\ s_1 \\ s_2 \\ \vdots \\ s_{nsec} \end{array}
\begin{array}{c} \begin{array}{cccccccc} 1 & 2 & \ldots & n & s_1 & s_2 & \ldots & s_{nsec} \end{array} \\
\begin{bmatrix} \rho_{1,1} & \rho_{2,1} & \cdots & \rho_{n,1} & \rho_{s1,1} & \rho_{s2,1} & \cdots & \rho_{snsec,1} \\ \rho_{1,2} & \rho_{2,2} & \cdots & \rho_{n,2} & \rho_{s1,2} & \rho_{s2,2} & \cdots & \rho_{snsec,2} \\ \vdots & & & \vdots & \vdots & & & \vdots \\ \rho_{1,n} & \rho_{2,n} & \cdots & \rho_{n,n} & \rho_{s1,n} & \rho_{s2,n} & \cdots & \rho_{snsec,n} \\ \rho_{1,s1} & \rho_{2,s1} & \cdots & \rho_{n,s1} & \rho_{s1,s1} & \rho_{s2,s1} & \cdots & \rho_{snsec,s1} \\ \rho_{1,s2} & \rho_{2,s2} & \cdots & \rho_{n,s2} & \rho_{s1,s2} & \rho_{s2,s2} & \cdots & \rho_{snsec,s2} \\ \vdots & & & \vdots & \vdots & & & \vdots \\ \rho_{1,nsec} & \rho_{2,snsec} & \cdots & \rho_{n,nsec} & \rho_{s1,nsec} & \rho_{s2,s2} & \cdots & \rho_{snsec,snsec} \end{bmatrix} \end{array}
\begin{array}{c} \\ \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \\ b_1 \\ b_2 \\ \vdots \\ b_{nsec} \end{bmatrix} \end{array}
\begin{array}{c} \\ = \end{array}
\begin{array}{c} \\ \begin{bmatrix} \rho_{1,0} \\ \rho_{2,0} \\ \vdots \\ \rho_{n,0} \\ \rho_{s1,0} \\ \rho_{s2,0} \\ \vdots \\ \rho_{snsec,0} \end{bmatrix} \end{array}$$

These are standard cokriging equations. The different correlation coefficients are computed from the following sources:

$\rho_{i,j},\ i, j = 0,\ldots,n$ correlation between primary data:

calculated from normal scores variogram: $\rho_{i,j} = 1 - \gamma\left(\mathbf{u}_i - \mathbf{u}_j\right)$

$\rho_{s_j,i},\ i = 0,\ldots,n;\ j = 1,\ldots,n_{\sec}$ correlation between primary values and secondary data:

calculated from Markov model: $\rho_{s_j,i} = \rho_{s_j,0} \bullet \rho_{i,0}$

$\rho_{s_j,s_i},\ i, j = 0,\ldots,n_{\sec}$ correlation between secondary data:

calculated directly from the secondary data

Note that the only variogram requires is the variogram of the primary variable, which is used to compute the correlation between the primary data at other locations to the location being estimated. The spatial correlation structure of the secondary data is not required because the secondary data are only used at the location being estimated. The cross spatial correlation between primary data at other locations and the secondary data at the location being estimated are estimated through a Markov-type assumption, that is, the cross variograms are assumed to have the same shape – the sill is scaled to the correct cross correlation.

Software could be modified to solve these equations. There is a need, however, to use legacy code such as the standard GSLIB program or commercial software such as Petrel© and Pangeos© that commonly permit only one variable at a time. These are the situations when it is convenient to merge the secondary data.

Merging Multiple Secondary Variables (the *Super Secondary* Variable)

All secondary data can be merged as a linear combination into a single secondary variable that can be used in the conventional collocated cokriging.

$$y_{\substack{\text{super}\\\text{secondary}}}(\mathbf{u}) = \frac{\sum_{i=1}^{n_{\text{sec}}} c_i \bullet y_{s,i}}{\rho_{\substack{\text{super}\\\text{secondary}}}} \tag{4}$$

Where the c_i weights are calculated from the well known cokriging equations:

$$\sum_{j=1}^{n_{\text{sec}}} c_j \bullet \rho_{i,j} = \rho_{i,0}, \quad i = 1,\ldots,n_{\text{sec}} \tag{5}$$

The left hand side $\rho_{i,j}$ values represent the redundancy between the secondary data. The right hand side $\rho_{i,0}$ values represent the relationship between each secondary data and the primary variable being predicted. The correlation coefficient of the super secondary value with the primary variable being estimated is based on the cokriging variance:

$$\rho_{\substack{\text{super}\\\text{secondary}}} = \sqrt{\sum_{i=1}^{n_{\text{sec}}} c_i \bullet \rho_{i,0}} \tag{6}$$

The expression inside the square root is one minus the estimation variance, which would is precisely the correlation coefficient is one data is being used. Recall that $\sigma_k^2 = 1 - \rho_{i,0} \bullet \rho_{i,0}$ in presence of one data, thus $\rho_{i,0} = \sqrt{1-\sigma_k^2}$ given that the estimation variance is known, as is the case here.

The single super secondary variable is used with the primary data in the well known collocated cokriging equations:

$$y^* = \sum_{i=1}^{n} a_i \bullet y_i + c \bullet y_{\substack{\text{super}\\\text{secondary}}} \tag{7}$$

$$\sigma_K^2 = 1 - \sum_{i=1}^{n} a_i \bullet \rho_{i,0} - c \bullet \rho_{\substack{\text{super}\\\text{secondary}}} \tag{8}$$

The results of equations (7) and (8) are exactly the same as that of equations (1) and (2).

Proof of the Super Secondary Approach

Let us now prove that the two collocated cokriging estimators presented in (1)-(3) and (4)-(9) are exactly the same. First, let us rewrite both collocated cokriging systems into matrix format. Specifically, the system (1)-(3) for cokriging with multiple secondary data can be rewritten (using Markov model) as

$$y_1^* = a_1^T y + b^T y_s, \tag{10}$$

$$\sigma_{1,K}^2 = 1 - a_1^T C_0 - b^T \rho_0, \tag{11}$$

where $a_1 = (a_{11},\ldots,a_{1n})^T$ and $b = (b_1,\ldots,b_{n_{\text{sec}}})^T$ are the weights applied in estimation to the primary $y = (y_1,\ldots,y_n)^T$ and secondary $y_s = (y_{s1},\ldots,y_{sn})^T$ data, respectively; $C_0 = (\rho_{1,0},\ldots,\rho_{n,0})^T$ is the covariance between primary data and the estimation location; $\rho_0 = (\rho_{s1,0},\ldots,\rho_{sn_{\text{sec}},0})^T$ is the vector of correlations between primary and multiple collocated secondary data. The weights $a_1 = (a_{11}, \ldots, a_{1n})^T$ and $b = (b_1, \ldots, b_{n_{\text{sec}}})^T$ are given by the following system:

$$\begin{aligned} &C a_1 + C_0 \rho_0^T b = C_0 \\ &\rho_0 C_0^T a_1 + C_{\text{sec}} b = \rho_0 \end{aligned} \tag{12}$$

where C is *n* by *n* data-to-data covariance matrix for the primary data ($C_{ij} = \rho_{i,j}, \quad i, j = 1, \ldots, n$) and C_{sec} is n_{sec} by n_{sec} matrix of correlations between multiple secondary data ($C_{\text{sec},ij} = \rho_{si,sj}, \quad i, j = 1, \ldots, n_{\text{sec}}$).

The system (4)-(9) for cokriging with one super secondary data can be rewritten (using Markov model) as

$$y_2^{\,*} = a_2^T y + c y_{\substack{\text{super}\\\text{secondary}}}, \tag{13}$$

$$\sigma_{2,K}^2 = 1 - a_2^T C_0 + c \rho_{\substack{\text{super}\\\text{secondary}}}, \tag{14}$$

where $a_2 = (a_{21}, \ldots, a_{2n})^T$ and c are the weights applied in estimation to the primary and super secondary data, respectively; as before $C_0 = (\rho_{1,0}, \ldots, \rho_{n,0})^T$ is the covariance between primary data and the estimation location; $\rho_{\substack{\text{super}\\\text{secondary}}}$ is given (see equations (5)-(6)) by

$$\rho_{\substack{\text{super}\\\text{secondary}}}^2 = \rho_0^T C_{\text{sec}}^{-1} \rho_0 \tag{15}$$

Note that C_{sec} is the covariance matrix – positive definite, thus invertible.

The weights for the primary and super secondary data, $a_2 = (a_{21}, \ldots, a_{2n})^T$ and *c*, are found from the following system:

$$\begin{aligned} &C a_2 + C_0 \rho_{\substack{\text{super}\\\text{secondary}}} c = C_0 \\ &\rho_{\substack{\text{super}\\\text{secondary}}} C_0^T a_2 + c = \rho_{\substack{\text{super}\\\text{secondary}}} \end{aligned}, \tag{16}$$

where as before C is *n* by *n* data-to-data covariance matrix for the primary data ($C_{ij} = \rho_{i,j}, \quad i, j = 1, \ldots, n$). Using the fact that (see equations (4)-(6)),

$$y_{\substack{\text{super}\\\text{secondary}}} = \frac{1}{\rho_{\substack{\text{super}\\\text{secondary}}}} (C_{\text{sec}}^{-1} \rho_0)^T y_s, \tag{17}$$

we can rewrite equation for the estimate of the collocated cokriging approach with one super secondary data as

$$y_2^{\,*} = a_2^T y + c \frac{1}{\rho_{\substack{\text{super}\\\text{secondary}}}} (C_{\text{sec}}^{-1} \rho_0)^T y_s. \tag{18}$$

Looking at systems (10)-(12) and (13)-(18), we can conclude that in order to prove that they result in the same outcome; we need to show that the weights received by primary and secondary data in both systems are the same. That is, we need to show that all of the following equalities hold

1) $a_{1i} = a_{2i}, \quad i = 1, \ldots, n;$

and

2) $b_i = c \dfrac{1}{\rho_{\substack{\text{super}\\\text{secondary}}}} (C_{\text{sec}}^{-1}\rho_0)_i, \quad i = 1, \ldots, n_{\text{sec}}.$

Proof of 1):

Let us first consider second matrix equation of system (12), that is,

$$\rho_0 C_0^T a_1 + C_{\text{sec}} b = \rho_0$$

Multiplying both sides of this matrix equation by $\rho_0^T C_{\text{sec}}^{-1}$, we obtain

$$\rho_0^T C_{\text{sec}}^{-1}[\rho_0 C_0^T a_1 + C_{\text{sec}} b] = \rho_0^T C_{\text{sec}}^{-1}\rho_0 \,,$$

or,

$$(\rho_0^T C_{\text{sec}}^{-1}\rho_0) C_0^T a_1 + \rho_0^T (C_{\text{sec}}^{-1} C_{\text{sec}}) b = \rho_0^T C_{\text{sec}}^{-1}\rho_0 \,, \tag{18}$$

Using (15) and the fact that $C_{\text{sec}}^{-1} C_{\text{sec}} = I,$ where I is identity matrix of size n by n, equation (18) reduces to

$$\rho^2_{\substack{\text{super}\\\text{secondary}}} C_0^T a_1 + \rho_0^T b = \rho^2_{\substack{\text{super}\\\text{secondary}}} \,. \tag{19}$$

Now let us consider second matrix equation of system (16), that is,

$$\rho_{\substack{\text{super}\\\text{secondary}}} C_0^T a_2 + c = \rho_{\substack{\text{super}\\\text{secondary}}}$$

Multiplying both sides of this equation by $\rho_{\substack{\text{super}\\\text{secondary}}}$, we obtain

$$\rho^2_{\substack{\text{super}\\\text{secondary}}} C_0^T a_2 + c\rho_{\substack{\text{super}\\\text{secondary}}} = \rho^2_{\substack{\text{super}\\\text{secondary}}} \,. \tag{20}$$

Subtracting from equation (19) equation (20), we get

$$\rho^2_{\substack{\text{super}\\\text{secondary}}} C_0^T a_1 + \rho_0^T b - \left[\rho^2_{\substack{\text{super}\\\text{secondary}}} C_0^T a_2 + c\rho_{\substack{\text{super}\\\text{secondary}}} \right] = \rho^2_{\substack{\text{super}\\\text{secondary}}} - \rho^2_{\substack{\text{super}\\\text{secondary}}} \,,$$

or,

$$\rho^2_{\substack{\text{super}\\\text{secondary}}} C_0^T [a_1 - a_2] + [\rho_0^T b - c\rho_{\substack{\text{super}\\\text{secondary}}}] = 0. \tag{21}$$

Recall that it follows from first matrix equations of systems (12) and (16) that

$$C_0 \rho_0^T b = C_0 - C a_1, \tag{22}$$

and

$$C_0 \rho_{\substack{\text{super}\\\text{secondary}}} c = C_0 - C a_2, \tag{23}$$

respectively. Thus, after subtracting equation (23) from (22), we obtain

$$C_0 \rho_0^T b - C_0 \rho_{\substack{\text{super}\\\text{secondary}}} c = C_0 - C a_1 - [C_0 - C a_2],$$

or,

$$C_0 [\rho_0^T b - c \rho_{\substack{\text{super}\\\text{secondary}}}] = -[C a_1 - C a_2]. \tag{24}$$

Let us now return to equation (21). If we multiply both sides of equation (21) by C_0, we obtain

$$C_0 \rho^2_{\substack{\text{super}\\\text{secondary}}} C_0^T [a_1 - a_2] + C_0 [\rho_0^T b - c \rho_{\substack{\text{super}\\\text{secondary}}}] = 0. \tag{25}$$

Due to (24), we can rewrite (25) as

$$C_0 \rho^2_{\substack{\text{super}\\\text{secondary}}} C_0^T [a_1 - a_2] - [C a_1 - C a_2] = 0,$$

or,

$$\left[C_0 \rho^2_{\substack{\text{super}\\\text{secondary}}} C_0^T - C \right] [a_1 - a_2] = 0. \tag{26}$$

Thus, it follows from equation (26) that $a_1 = a_2$ provided that determinant of matrix $C_0 \rho^2_{\substack{\text{super}\\\text{secondary}}} C_0^T - C$ is not equal to zero. So, let us examine whether determinant of matrix $C_0 \rho^2_{\substack{\text{super}\\\text{secondary}}} C_0^T - C$ could be zero. Consider system (16), rewritten below,

$$\begin{aligned} &C a_2 + C_0 \rho_{\substack{\text{super}\\\text{secondary}}} c = C_0 \\ &c = \rho_{\substack{\text{super}\\\text{secondary}}} [1 - C_0^T a_2] \end{aligned} \tag{27}$$

If we substitute expression for *c* in system (27) into first matrix equation of this system, we will obtain the following

$$\begin{aligned} &C a_2 + C_0 \rho^2_{\substack{\text{super}\\\text{secondary}}} [1 - C_0^T a_2] = C_0 \\ &c = \rho_{\substack{\text{super}\\\text{secondary}}} [1 - C_0^T a_2] \end{aligned},$$

or,

$$\begin{aligned} &-\left[C_0 \rho^2_{\substack{\text{super}\\\text{secondary}}} C_0^T - C \right] a_2 = C_0 \left[1 - \rho^2_{\substack{\text{super}\\\text{secondary}}} \right] \\ &c = \rho_{\substack{\text{super}\\\text{secondary}}} [1 - C_0^T a_2] \end{aligned} \tag{28}$$

Thus, clearly, if determinant of matrix $C_0 \rho^2_{\substack{\text{super}\\\text{secondary}}} C_0^T - C$ is equal to zero, the system (28) will have either multiple or no solution depending on the vector $C_0 \left[1 - \rho^2_{\substack{\text{super}\\\text{secondary}}} \right]$ (see Cramer's rule). The collocated cokriging system with super secondary variable (16) (or equivalently (28)) has only one solution for the primary weights a_2 if and only if determinant of matrix $C_0 \rho^2_{\substack{\text{super}\\\text{secondary}}} C_0^T - C$ is not equal to zero. Similarly, using simple matrix manipulations, we can show that the collocated cokriging system (12) will have unique solution for the primary weights if and only if determinant of matrix $C_0 \rho^2_{\substack{\text{super}\\\text{secondary}}} C_0^T - C$ is not equal to zero. As results, we can conclude that provided that the collocated cokriging systems have unique solution, the weights given by both cokriging systems (1)-(3) and (4)-(9) to the primary variable are exactly the same. That is,

$$a = a_{1i} = a_{2i}, \quad i = 1, \ldots, n.$$

Thus, proof of 1) is completed.

Proof of 2):

Let us now prove that

$$b_i = c \frac{1}{\rho_{\substack{\text{super}\\\text{secondary}}}} (C_{\sec}^{-1} \rho_0)_i, \quad i = 1, \ldots, n_{\sec}.$$

It follows from the second matrix equation of system (12) for the collocated cokriging weights with multiple secondary data that

$$C_{\sec} b = \rho_0 - \rho_0 C_0^T a = \rho_0 [1 - C_0^T a].$$

Also, it follows from system (28) (rewritten system (16)) for the collocated cokriging weights with super secondary data that

$$c = \rho_{\substack{\text{super}\\\text{secondary}}} [1 - C_0^T a].$$

As result,

$$\begin{aligned} C_{\sec} \left[b - c \frac{1}{\rho_{\substack{\text{super}\\\text{secondary}}}} (C_{\sec}^{-1} \rho_0) \right] &= C_{\sec} b - C_{\sec} c \frac{1}{\rho_{\substack{\text{super}\\\text{secondary}}}} (C_{\sec}^{-1} \rho_0) \\ &= C_{\sec} b - c \frac{1}{\rho_{\substack{\text{super}\\\text{secondary}}}} (C_{\sec} C_{\sec}^{-1} \rho_0) = C_{\sec} b - c \frac{1}{\rho_{\substack{\text{super}\\\text{secondary}}}} \rho_0 \\ &= \rho_0 [1 - C_0^T a] - \rho_{\substack{\text{super}\\\text{secondary}}} [1 - C_0^T a] \frac{1}{\rho_{\substack{\text{super}\\\text{secondary}}}} \rho_0 \\ &= \rho_0 [1 - C_0^T a] - [1 - C_0^T a] \rho_0 = [1 - C_0^T a] \rho_0 - [1 - C_0^T a] \rho_0 = 0. \end{aligned} \tag{29}$$

Since matrix C_{sec} is a positive-definite matrix of correlations between collocated secondary data, its determinant is not equal to zero. As result, it follows (29) that $b - c\frac{1}{\rho_{\substack{\text{super}\\\text{secondary}}}}(C_{\text{sec}}^{-1}\rho_0) = 0$. Therefore, we proved that

$$b_i = c\frac{1}{\rho_{\substack{\text{super}\\\text{secondary}}}}(C_{\text{sec}}^{-1}\rho_0)_i, \quad i = 1,\ldots,n_{\text{sec}}.$$

And, thus, we have shown that the two collocated cokriging estimators presented in (1)-(3) and (4)-(9) are exactly the same.

Collocated Cokriging with a Super Secondary Variable: Example

The following example is based on the Amoco data set of Chu and Xu (1995). A 10500 by 10500 ft reservoir layer is considered. There are 62 wells in the study area. For each well, the data on the porosity and thickness are available. Water saturation for each of the 62 wells is simulated. The location maps of the three variables of interest, that is, porosity, thickness and water saturation, together with their univariate declustered distributions are shown in Figure 1. The normal score variograms of the porosity, thickness and water saturation and crossplots between them in normal space are shown in Figures 2 and 3, respectively.

Let us now establish the procedure to compute the mean and variance of the local conditional distributions of water saturation in a multivariate Gaussian context using the collocated cokriging with super secondary variable calculated based on porosity and thickness for the 10500 by 10500 ft Amoco reservoir layer.

Step 1: Based on the exhaustive gridded data for the 10500 by 10500 ft Amoco reservoir layer on porosity and thickness, calculate the super secondary variable to be used in collocated cokriging for water saturation.

1.1. For the Amoco data example Sequential Gaussian Simulation (SGS) is performed to obtain one full realization of porosity for the study domain;

1.2. Then, Sequential Gaussian Simulation with collocated cokriging is performed to obtain one realization of thickness collocated to porosity.

1.3. Both simulated realizations are transformed to the normal space.

1.4. Super secondary random variable to be used in collocated cokriging for water saturation can be established as (see equations (4))

$$y_{\substack{\text{super}\\\text{secondary}}}(u) = \frac{c_1 y_{\text{ns porosity}} + c_2 y_{\text{ns thickness}}}{\rho_{\substack{\text{super}\\\text{secondary}}}}, \quad (30)$$

where $y_{\text{ns porosity}}$ and $y_{\text{ns thickness}}$ are the normal score transformed porosity and thickness realizations and c_1 and c_2 are the weights assigned to them by equation (5) and $\rho_{\substack{\text{super}\\\text{secondary}}}$ is given by equation (6). For the Amoco data, the weights c_1 and c_2 are fond from the following system

$$\begin{bmatrix} 1 & -0.345 \\ -0.345 & 1 \end{bmatrix}\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} -0.68 \\ 0.179 \end{bmatrix}. \quad (30)$$

Thus, $c_1 = -0.7018$, $c_1 = -0.0631$. And $\rho_{\substack{\text{super}\\\text{secondary}}} = \sqrt{c_1(-0.68) + c_2(0.179)} = 0.6826$. Therefore, the super secondary random variable to be used in collocated cokriging for water saturation is given by

$$y_{\substack{\text{super}\\\text{secondary}}}(u) = -1.0281 y_{\text{ns porosity}} - 0.0925 y_{\text{ns thickness}}. \qquad (31)$$

Step 2: Calculate the mean and variance of the local conditional distributions of water saturation using the super secondary variable in the collocated cokriging for the 10500 by 10500 ft Amoco reservoir layer.

2.1 'Traditional' Cokriging was performed for the water saturation. The map of the means and variances of the local conditional distributions of the nscore water saturation (see equations (7) and (8)) are given in Figure 4.

Conclusion

Multiple secondary data and previous primary variables are merged into a single *super* secondary variable. Consideration of this single variables amounts to using each constituent variable with the correct correlation. The variable can be used in collocated cokriging with a typical Markov model or under an intrinsic model to avoid variance inflation (see CCG paper 107 in this report).

The suggested workflow consists of proceeding one at a time through the primary variables. At each step, all secondary data and previously simulated primary variables are merged into a super secondary variable. This variable would be different for each realization. This greatly simplified the cosimulation of multiple variables and makes it easy to check the models as modeling proceeds.

This paper presents a proof that using a single merged super secondary variable and appropriate correlation coefficient is exactly the same as considering all constituent variables individually in a more complicated cosimulation. Although the notion of a super secondary variable was proposed earlier, it is reassuring to have the proof that the approach is theoretically justified.

References

Almeida, A.S., 1993, *Joint Simulation of Multiple Variables with a Markov-Type Coregionalization Model*, Ph.D. Thesis, Stanford University.

Almeida, A.S. and Journel, A.G., 1994, Joint Simulation of Multiple Variables with a Markov-Type Coregionalization Model, Math Geology, Vol 26, pages 565-588.

Deutsch, C.V. and Journel, A.G., 1998, *GSLIB: Geostatistical Software Library: and User's Guide*. Oxford University Press, New York, 2nd Ed.

Deutsch, C.V., 2002, *Geostatistical Reservoir Modeling*. Oxford University Press, New York.

Xu, W. Tran, T.T., Srivastava, R.M., and Journel, A.G., 1992, Integrating Seismic Data in Reservoir Modeling: the Collocated Cokriging Alternative, 67th Annual Technical Conference and Exhibition, Society of Petroleum Engineers, Washington, DC, October 1992, SPE Paper Number 24742.

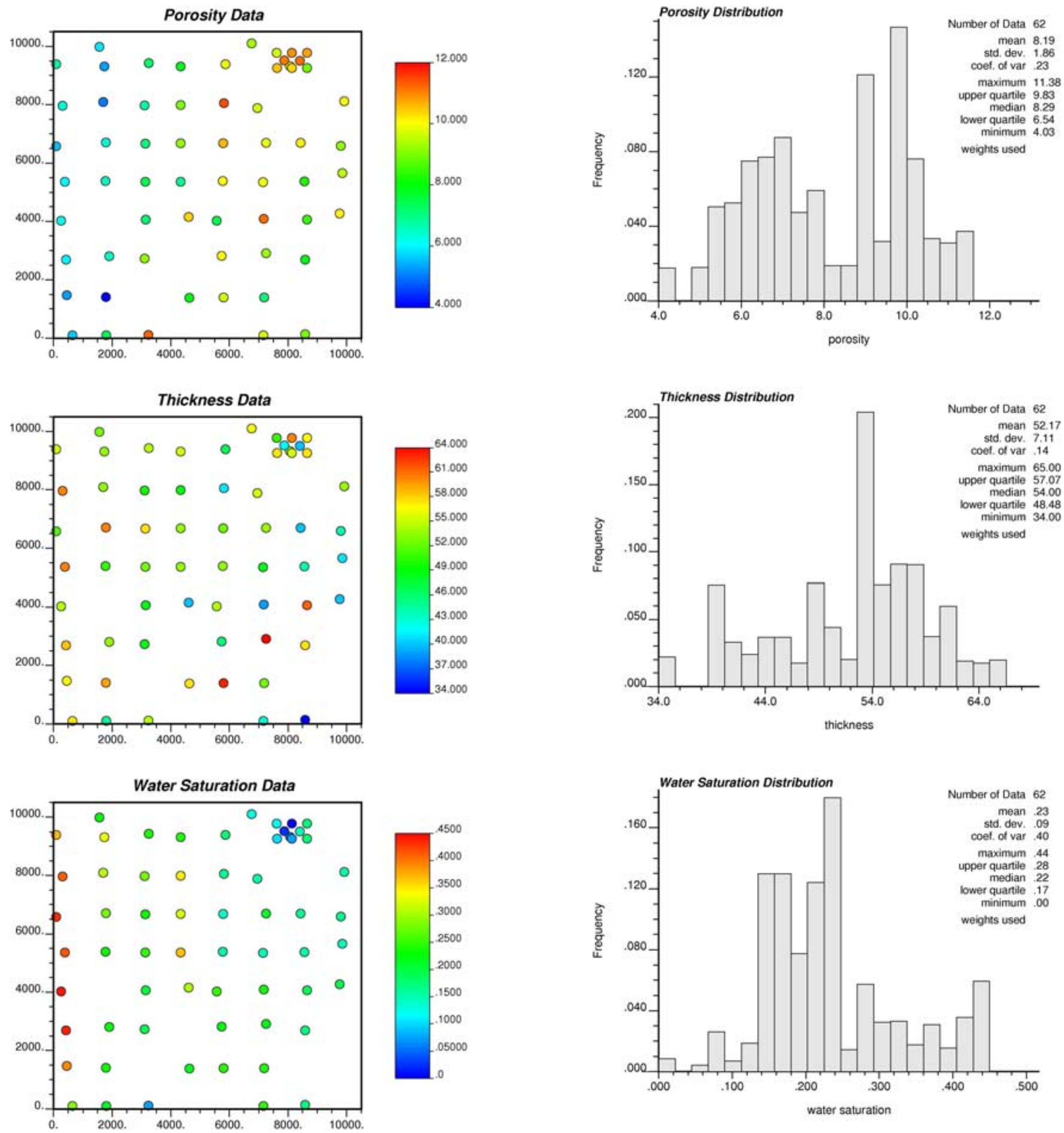


Figure 1: Location maps of the porosity (top), thickness (middle) and water saturation (bottom), together with their univariate declustered distributions for the set of 62 data.

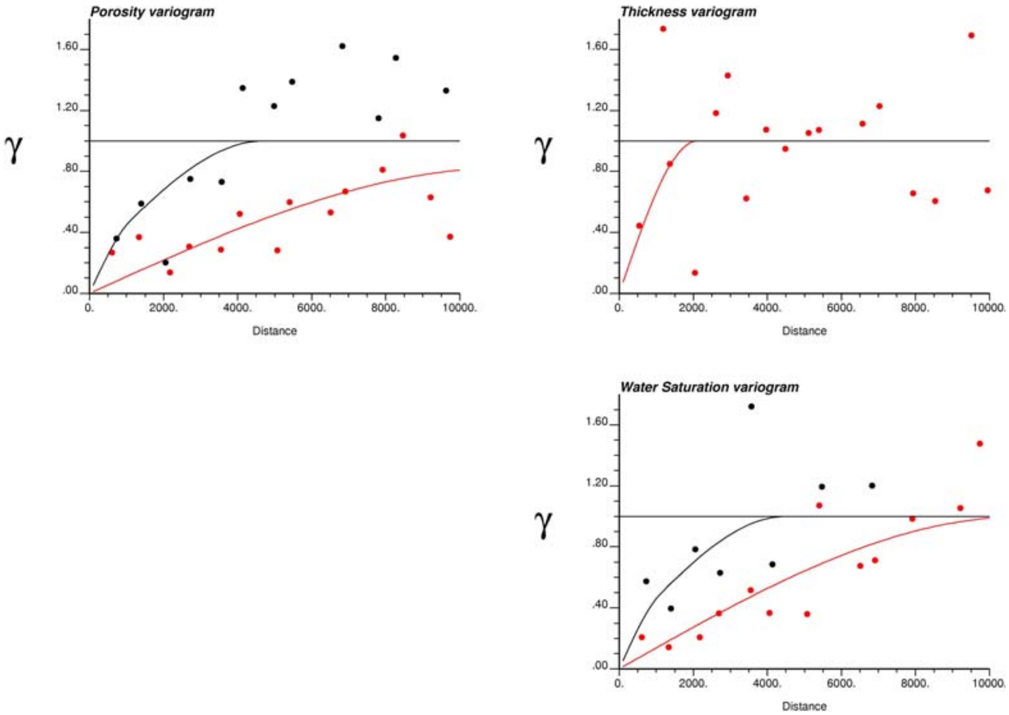


Figure 2: The normal score variograms of the porosity (top left), thickness (top right) and water saturation (bottom).

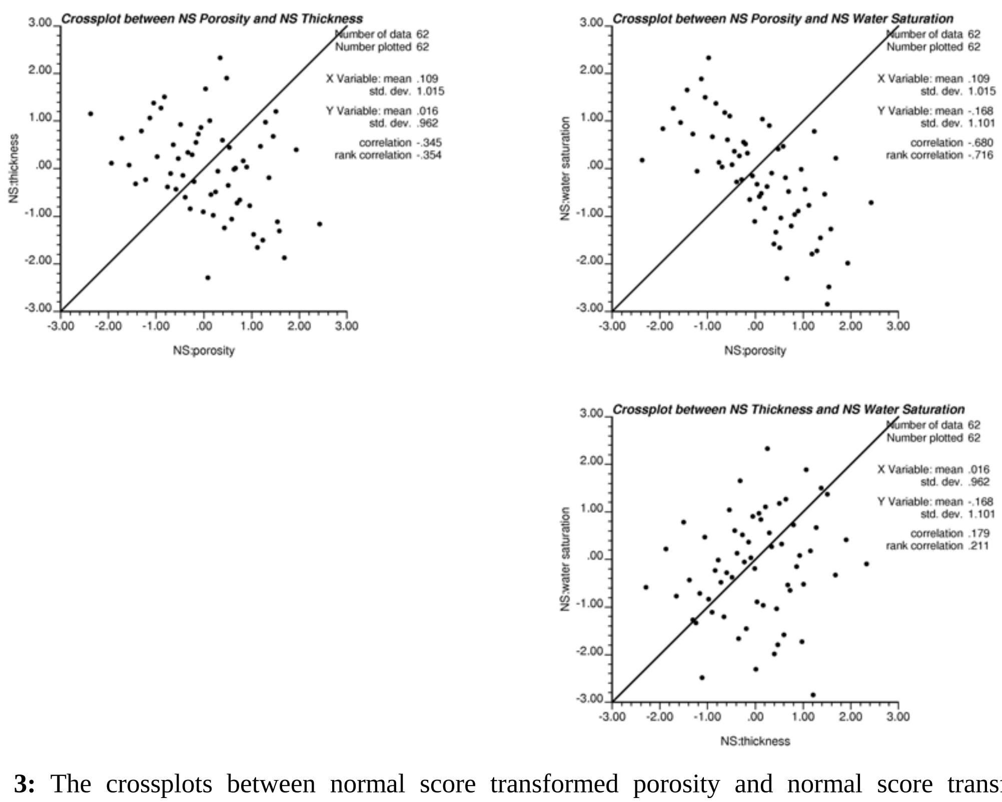


Figure 3: The crossplots between normal score transformed porosity and normal score transformed thickness (top left); normal score transformed porosity and normal score transformed water saturation (top right); and normal score transformed thickness and normal score transformed water saturation (bottom).

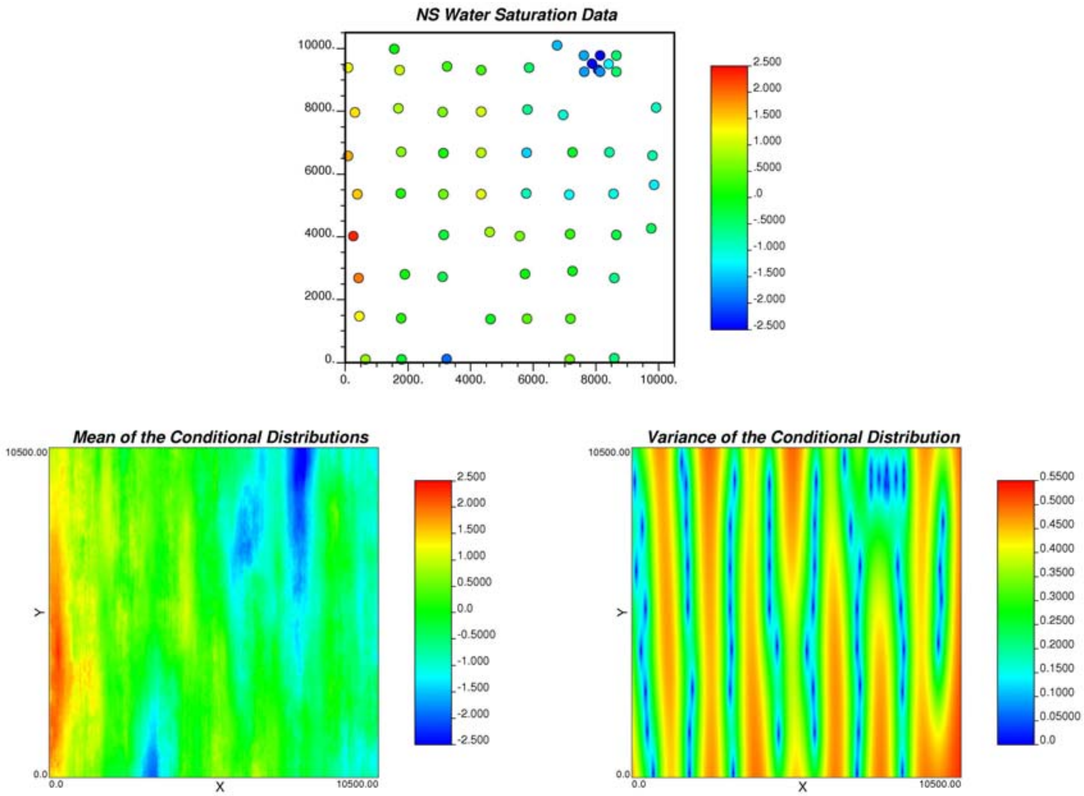


Figure 4: The normal score water saturation data (top); maps of the means (bottom left) and variances (bottom right) of the local conditional distributions of the nscore water saturation.