Natural Resources Research, Vol. 11, No. 1, March 2002 (© 2002)

Size Scaling of Cross Correlation Between Multiple Variables

Bora Oz[1] and Clayton V. Deutsch[1,2]

Received 6 April 2000; accepted 14 July 2001

Reservoir models have large uncertainty because of spatial variability and limited sample data. The ultimate aim is to use simultaneously all available data sources to reduce uncertainty and provide reliable reservoir models for resource assessment and flow simulation. Seismic impedance or some other attribute provides a key source of data for reservoir modeling. These seismic data are at a coarser scale than the hard well data and it not an exact measurement of facies proportions or porosity. A requirement for data integration is the cross-covariance between the well and seismic data.

The size-scaling behavior of the cross correlation for different measurement scales was investigated. The size-scaling relationship is derived theoretically and validated by numerical studies (including an example with real data). The limit properties of the cross-correlation coefficient when the averaging volume becomes large is shown. After some averaging volume, the volume-dependent cross-correlation coefficient reaches a limit value. This plateau value is controlled mainly by the large-scale behavior of the cross and direct variograms.

The cross correlation can increase or decrease with volume support depending on the relative importance of long- and short-scale covariance structures. If the direct and cross variograms are proportional, there is no change in the cross correlation as the averaging volume changes. Our study shows that the volume-dependent cross-correlation coefficient is sensitive to the shape of the cross variogram and differences between the direct variograms of the well data and seismic data.

KEY WORDS: Data-integration; correlation coefficient; volume scaling; dispersion variance; dispersion covariance.

INTRODUCTION

Reconciling different data types for spatial modeling of reservoir properties is important because these different data provide complementary information about the reservoir architecture and heterogeneity. There are a variety of methods to integrate different data types; these include *External Drift, Locally Varying Mean, Block Kriging* (Behrens and others, 1996; Behrens and Tran, 1998; Deutsch and Journel, 1997; Journel and Huijbregts, 1978), *Block Cokriging* (Doyen, 1988), *Markov-Bayes* (or *Bayesian Updating Rule*) (Behrens and Tran, 1998; Doyen, den Boer, and Pillet, 1996; Doyen, Psaila, den Boer, and Jans, 1997; Doyen, Psaila, and Strandenes, 1994; Journel and Zhu, 1990; Tran, Wen, and Behrens, 1999) and *Collocated Cokriging* (Almeida, 1993; Xu, Tran, Srivastava, and Journel, 1992). Details and application examples of these methods are given in literature (Deutsch and Journel, 1997; Deutsch, Srinivasan, and Mo, 1996; Journel and Huijbregts, 1978). The main aim of this paper is to address a common requirement: the cross covariance between multiple data types.

The determination of the cross covariance between multiple variables with different measurement scales is an important step for all data integration techniques. Some approaches assume that the soft or secondary data provides information on large-scale trends of the primary variable; the external drift and locally differing mean algorithms assume that the spatial variability of the secondary variable gives

[1] Department of Civil & Environmental Engineering, University of Alberta.
[2] To whom correspondence should be addressed. (e-mail: cdeutsch @civil.ualberta.ca)

1520-7439/02/0300-0001/1

information on trends in the primary variable. This approach does not capture fully the spatial cross correlation.

A better approach to data integration is cokriging which calls for a cross covariance that explicitly measures the information content of the secondary data with regard to the primary variable (Journel, 1999a). A major problem is that conventional implementations of cokriging assume the secondary data are defined at the same volume support as the hard well data. It is possible theoretically to use data at different support; however, the computation and inference burden prevents application in practice.

A Markov-type assumption termed MM1 (Almeida, 1993; Journel, 1999b; Shmaryan and Journel, 1999) simplifies inference of the cross variogram, but is not valid when the secondary variable is defined on a support larger than that of the primary variable. For such situations, a second Markov hypothesis termed MM2 is proposed (Journel, 1999b; Shmaryan and Journel, 1999) leading to a different cross-covariance model. The MM2 cross covariance then is a function of the secondary variable covariance and the colocated or small-scale correlation coefficient. Under both approximations, the cross covariance is rescaled from either primary or secondary variable covariance by the small-scale correlation coefficient independently.

Kupfersberger, Deutsch, and Journel (1998) propose analytical equations to infer small-scale variograms from a combination of small-scale and large-scale data. The key idea is to downscale the large-scale variograms to small-scale, then complete the small-scale horizontal variogram with the more extensive secondary data.

Data sources have a wide range of measurement volumes and cross-correlation characteristics; a volume or size dependent cross-correlation structure is required. Vargas-Guzman, Warrick, and Myers (1999a, 1999b) have tackled a related question. They extend the concept of dispersion variance to the multivariate situation where the volume size or support affects dispersion covariances and the matrix of correlation between attributes. This leads to a correlation between attributes as a function of sample support and the size of the physical domain. They show that the correlation matrix asymptotically approaches a constant at two or three times the largest variogram range. They also analyzed the behavior of the cross covariance by keeping the data support at a point support and changing the field size.

In terms of data integration, changing the data support size for a fixed field size is more critical; this has not been tackled by previous workers. Therefore, our focus in this paper is on the effect of the data support on cross correlation. We organize the paper according to the following sections:

- **Theoretical Development:** The cross-correlation coefficient is defined from the dispersion variance and covariance. A general equation for the volume-dependent cross-correlation coefficient is presented and the critical terms forming this equation are interpreted.
- **Numerical Validation of Theory:** A numerical solution technique for the volume-dependent cross-correlation coefficient equation is described. We show how to calculate the required volume-average covariance values. An example illustrates the volume-dependent cross-correlation coefficient and its numerical calculation.
- **Sensitivity Cases:** Different positive definite cross-variogram direct variogram models and nugget effects are used to better understand the characteristics of the cross correlation with respect to "upscaling." In our study upscaling is a linear block averaging of the small-scale features (e.g. grid blocks etc . . .). Note that the functional relationship of the cross-correlation coefficient to scales is complex and depends on many factors.
- **Application:** Theory and practice are compared with real satellite data. Direct variograms and cross variograms are calculated and upscaled to estimate the cross correlation at different scales. The experimentally obtained results are close to the theoretical and numerical results for the large block volumes; however, some discrepancies are observed for the small-block averaging volumes. Both curves clearly approach to the same limit (plateau) value of cross-correlation coefficient as scaling volume increases.
- **Analytical Analysis:** The complexity and nonlinearity of the terms forming the volume-dependent cross-correlation coefficient equation are investigated. Closed-form equations for the estimation of volume-dependent cross-correlation coefficient are shown for some limited examples. The asymptotic values of the

cross correlation as the volume scale increases are considered; a good match is observed between the numerical and analytical results.

There are many benefits to better understanding the size-scaling relationship of the cross correlation: (1) the input parameters for conventional collocated cokriging applications (small-scale correlation coefficient) can be selected more correctly on the basis of the calculated large-scale correlation coefficient, (2) the value of seismic data can be more realistically appraised, and (3) correct variograms can be used for development of rigorous block cokriging.

THEORETICAL DEVELOPMENT

There exist two types of spatial variability in almost all natural phenomena; local *random* aspects and general *structured* aspects (Journel and Huijbregts, 1978). The concept of a "random function" provides a representation of both aspects of variability. A random variable (RV) Z is a variable that can take a series of outcome values z, according to some probability distribution. A random function (RF) is defined as a set of dependent variables $Z(\mathbf{u})$, one for each location $\mathbf{u}$ in the study area $\mathbf{A}$, $(Z(\mathbf{u}), \forall \mathbf{u} \in \mathbf{A})$.

Classically, the first-order moment of the function $Z(\mathbf{u})$ is its expected value, which is the probability-weighted sum of all possible occurences of the RV. "Stationarity," that is spatial homogeneity, removes the location-dependent nature of the expected value,

$$E\{Z(\mathbf{u})\} = m_Z, \quad \forall \mathbf{u} \in \mathbf{A} \tag{1}$$

where m_Z is the stationarity mean. The assumption of stationarity is critical. The stationary variance is defined as:

$$Var_Z\{Z(\mathbf{u})\} = \sigma^2 = E\{[Z(\mathbf{u}) - m_Z]^2\}, \quad \forall \mathbf{u} \in \mathbf{A} \tag{2}$$

Moving on from classical one-point statistics, we consider pairs of data a vector $\mathbf{h}$ apart, $[Z(\mathbf{u}), Z(\mathbf{u}+\mathbf{h})]$. Second-order stationarity amounts to assume that pairs of data do not depend on the location $\mathbf{u}$ within $\mathbf{A}$, but rather only on the distance, $\mathbf{h}$ separating them. The stationary covaraince is defined as:

$$C_Z(\mathbf{h}) = E\{Z(\mathbf{u}) \cdot Z(\mathbf{u}+\mathbf{h})\} - m_Z^2 \quad \forall \mathbf{u} \in \mathbf{A} \tag{3}$$

The variogram is defined as:

$$2\gamma_Z(\mathbf{h}) = E\{[Z(\mathbf{u}) - Z(\mathbf{u}+\mathbf{h})]^2\}, \quad \forall \mathbf{u} \in \mathbf{A} \tag{4}$$

The relation between the stationary semivariogram and the stationary covariance is straight-forwardly derived:

$$\gamma_Z(\mathbf{h}) = \sigma^2 - C(\mathbf{h}) \tag{5}$$

One important remark on Equations from (1) to (5) is that they are equally valid at the scale of the support of $Z(\mathbf{u})$. A challenge is to be able to calculate them at different scales; this is addressed in the next section. Another implementation detail is the estimation of the expected values in practical settings with limited data; however, this will not be addressed in this paper.

The elementary statistics described here could be calculated with a primary data variable, denoted Z, or a different secondary data variable, denoted Y. A familiar statistic relating two variables is the correlation coefficient, ρ, defined as:

$$\rho = \frac{E\{Z(\mathbf{u}) \cdot Y(\mathbf{u})\} - m_Z m_Y}{\sigma_Z \sigma_Y} = \frac{C_{ZY}}{\sqrt{C_{ZZ} C_{YY}}} \tag{6}$$

where two different notations are used; both notations are consistent with common practice and the introductions.

VOLUME-DEPENDENT CORRELATION COEFFICIENT

The equation for the data-scale (the scale of the support of $Z(\mathbf{u})$) correlation coefficient was presented in Equation (6). The correlation coefficient at a scale V different than the scale of the support of $Z(\mathbf{u})$ is defined as:

$$\rho(V, V) = (\bar{C}_{ZY}(V, V)) \cdot \left(\frac{1}{\sqrt{\bar{C}_{ZZ}(V, V)}}\right) \times \left(\frac{1}{\sqrt{\bar{C}_{YY}(V, V)}}\right) \tag{7}$$

where, $\bar{C}_{ZY}(V, V)$ is the volume-averaged cross-variogram, $\bar{C}_{ZZ}(V, V)$ and $\bar{C}_{YY}(V, V)$ are the volume-averaged direct variograms. These volume averaged covariances are defined classically as $\frac{1}{V^2}\int_V \int_V C(\mathbf{u}-\mathbf{u}')\, d\mathbf{u}\, d\mathbf{u}'$, which is approximated closely by numerical integration.

As a side note, the sill value of a direct semivariogram is the variance given it is calculated for a large volume. The variance at an arbitrary scale V also is termed "dispersion variance" and is equal to the $\bar{C}(V, V)$ when the area is large enough. The sill value of a cross semivariogram is similarly defined.

The dependence of correlation coefficient on volume is linked through Equation (7). A numerical approach to calculate this from the point-scale (point-scale is assumed for the scale of the support of data) variance and covariance is presented next.

NUMERICAL VALIDATION OF THEORY

Consider a semivariogram model at arbitrary scale V made up of a nugget effect and *nst* nested variogram structures:

$$\gamma_V(h) = C_V^0 + \sum_{i=1}^{nst} C_V^i \Gamma_V^i(h)$$

h¿0(8) where $\gamma_V(h)$ is the variogram model at the V scale, C_V^0 is the nugget effect, *nst* is the number of nested variogram structures, C_V^i is the variance contribution of each nested structure, $i = 1, \ldots, nst$, and $\Gamma_V^i(h)$ are nested structures consisting of analytical functions. The "sill" of each analytical function $\Gamma_V^i(h)$ is unity. The sum of the variance contribution is the variance at the V-scale and also is termed the dispersion variance:

$$D^2(V, A) = C_V^0 + \sum_{i=1}^{nst} C_V^i \quad \text{only if } A \gg V \tag{9}$$

where $D^2(V, A)$ is the variance (dispersion) of volumes of size V in the entire area of interest A. The variance decreases as the volume increases since high and low values are averaged out as the volume of investigation increases.

The variance contribution of each nested structure changes with volume in a well understood manner (Journel and Huijbregts, 1978):

$$C_V^i = C_v^i \frac{1 - \bar{\Gamma}(V, V)}{1 - \bar{\Gamma}(v, v)} \tag{10}$$

where C_V^i is the variance contribution of nested structure i at the large scale and the C_v^i is the variance contribution of nested structure i at the data scale, and $\bar{\Gamma}(V, V)$ and $\bar{\Gamma}(v, v)$ are the average variogram or "gamma-bar" values. Note that the change in the variance contribution is calculated separately for each nested structure. The "gamma-bar" value represents the mean value of $\Gamma(\mathbf{h})$ when one extremity of the vector $\mathbf{h}$ describes the domain $V(\mathbf{u})$ and the other extremity independently describes the same domain $V(\mathbf{u})$. In mathematical notation the "gamma-bar" value is expressed as:

$$\bar{\Gamma}(V, V) = \bar{\gamma}(V(\mathbf{u}), V(\mathbf{u})) = \frac{1}{V \cdot V} \int_{V(\mathbf{u})} \int_{V(\mathbf{u})} \gamma(y - y')\, dy\, dy' \tag{11}$$

Although there exist certain analytical solutions (David, 1977; Journel and Huijbregts, 1978) to $\bar{\gamma}(V(\mathbf{u}), V(\mathbf{u}))$, the value of "gamma-bar" usually is estimated numerically by discretizing the volume $V(\mathbf{u})$ and $V(\mathbf{u})$ into a number of points and simply averaging the variogram values:

$$\bar{\Gamma}(V, V) = \bar{\gamma}(V(\mathbf{u}), V(\mathbf{u})) \approx \frac{1}{n \cdot n'} \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma(\mathbf{u_i} - \mathbf{u'_j}) \tag{12}$$

where n is the number of regular spaced points discretizing the volume $V(\mathbf{u})$ with the same fractional volume of $V(\mathbf{u})$.

The same approach can be used to calculate the dispersion covariance using Equations (8) to (12); but instead of using auto or direct variograms, a cross-variogram should be used.

The values of dispersion variances and covariances allow calculation of the volume-dependent correlation coefficient. The volume dependent correlation coefficient is between $Z_v(\mathbf{u})$ and $Y_v(\mathbf{u})$ within the finite domain A:

$$\rho(V, A) = \left(D_{ZY}^2(V, A)\right) \cdot \left(\frac{1}{\sqrt{D_{ZZ}^2(V, A)}}\right) \times \left(\frac{1}{\sqrt{D_{YY}^2(V, A)}}\right) \tag{13}$$

where, $D_{ZY}^2(V, A)$ is the dispersion covariance and $D_{ZZ}^2(V, A)$ and $D_{YY}^2(V, A)$ are the dispersion variances at V-scale.

The `VarScale` program (Oz, Deutsch, and Frykman, 2000) can be used to calculate these dispersion variances and dispersion covariances. Consider an example to illustrate the numerical calculation of volume-dependent cross-correlation coefficient via Equation (13).

NUMERICAL VERIFICATION

A full cosimulation technique was used to simulate a prior-defined linear model of coregionalization

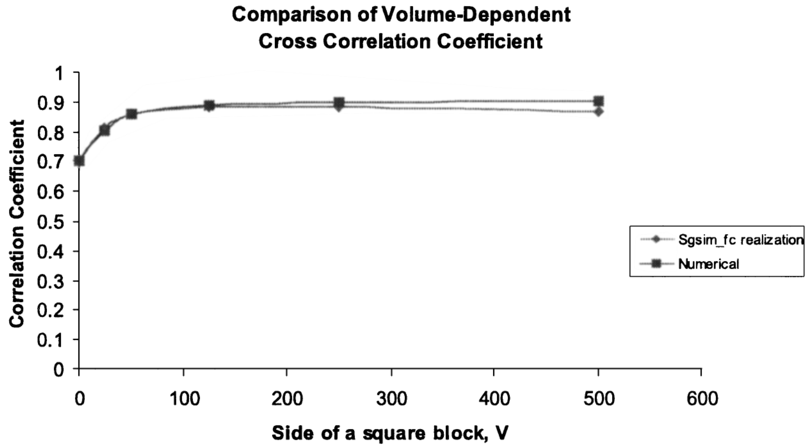

Figure 1. Illustration of cross-correlation coefficient from results obtained by numerical calculation and upscaling images generated by `Sgsimfc`.

(LMC) (Goovaerts, 1997; Journel and Huijbregts, 1978). The linear model of coregionalization provides a method for modeling the auto and cross-variograms of two or more variables. Each variable is characterized by its own variogram and each pair of variables with a cross-variogram.

An unconditional realization of 500 by 500 image was generated using the `Sgsim` program from GSLIB (Deutsch and Journel, 1997) and an isotropic variogram model given by Equation (14). Using this generated image as secondary data in `Sgsimfc` (Sequential Gaussian Full CoSimulation) (Deutsch and Journel, 1997) program, another image was created using the isotropic LMC model given by Equation (15). Two images were generated having correlation structure and direct variograms defined by Equation (15).

$$\gamma_Z(h) = 0.5Sph_{15} + 0.5Sph_{75} \tag{14}$$

$$\begin{aligned}\gamma_Z(h) &= 0.5Sph_{15} + 0.5Sph_{75}\\ \gamma_Y(h) &= 0.5Sph_{15} + 0.5Sph_{75}\\ \gamma_{ZY}(h) &= 0.2Sph_{15} + 0.5Sph_{75}\end{aligned} \tag{15}$$

Linear nonoverlapping block averaging (upscaling) have been applied for the two images and the corresponding volume-dependent cross-correlation coefficients were calculated. Next, using the LMC model in Equation (15) and the definition of volume-dependent cross-correlation coefficient in Equation (13), the values of correlation coefficient for different scaling ratios are calculated numerically by `VarScale` program (Oz, Deutsch, and Frykman, 2000). The numerically calculated volume-dependent cross-correlation coefficients and the ones obtained by upscaling the two images are illustrated on Figure 1.

The characteristics of the cross-correlation coefficient for different averaging volumes is now considered in more detail.

NUMERICAL EXPERIMENTATION

The main purpose of this study is to understand the general behavior of the cross correlation for different measurement scales. Some sensitivity runs are performed to understand the characteristics of the volume-dependent cross-correlation coefficient. These runs include sensitivity on the shape of the cross-variogram, nugget effect of cross-variogram and asymmetry of direct variograms.

Contribution of Nested Structures in Cross Variogram

Direct variograms, $\gamma_{ZZ}(\mathbf{h})$ and $\gamma_{YY}(\mathbf{h})$, for variables Z and Y were fixed and different situations of cross-variograms were considered. The direct variograms and all the considered cross-variograms are presented in Figure 2. The direct variograms are fixed at $\gamma_{ZZ}(\mathbf{h}) = \gamma_{YY}(\mathbf{h}) = 0.5Sph(|\mathbf{h}|/1) + 0.5Sph(|\mathbf{h}|/5)$. Two small-scale cross-correlation coefficients of 0.7 and 0.3 were considered. There are three scenarios

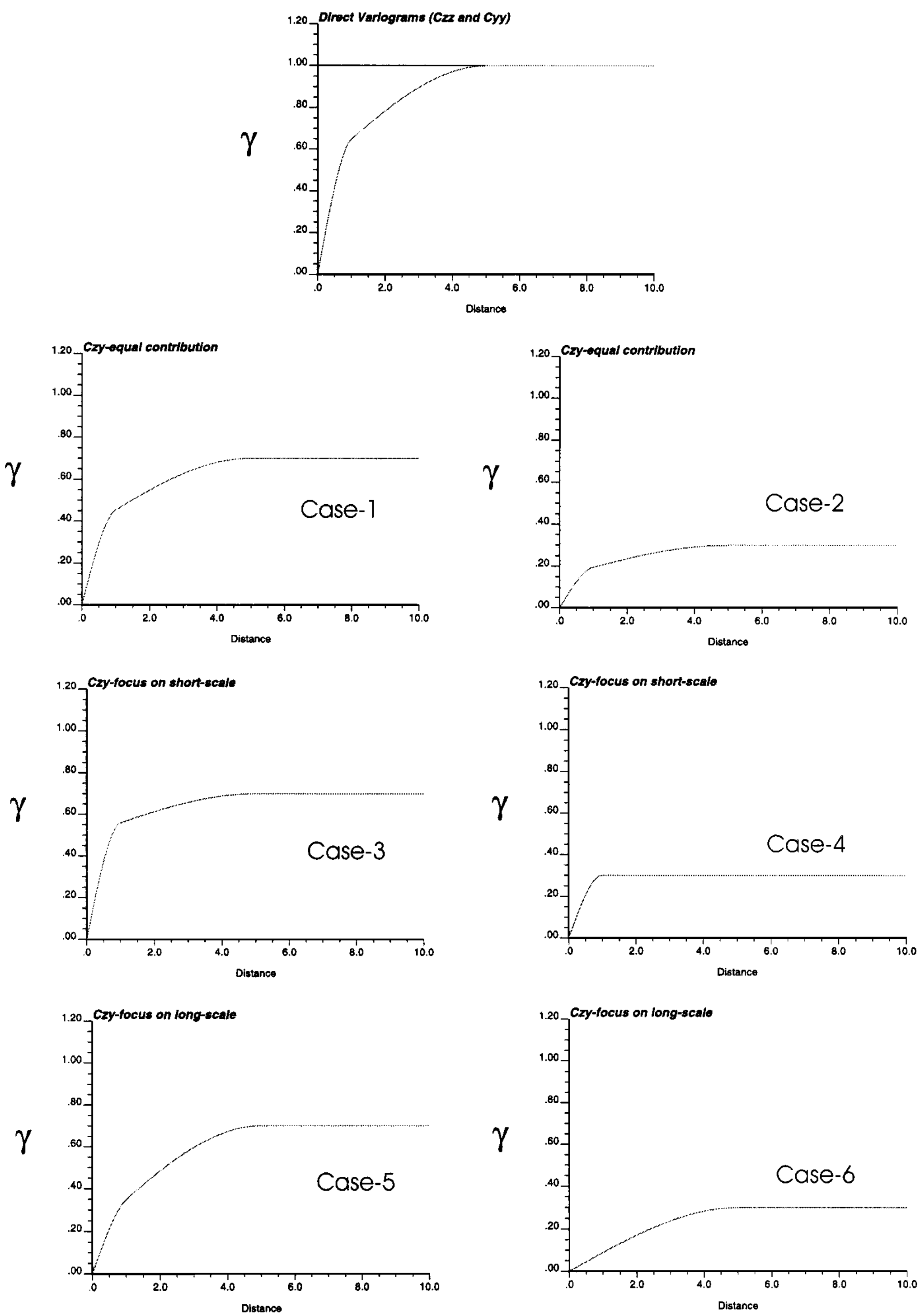

Figure 2. Fixed direct variogram and different cross-variograms for cases of cross variogram sensitivity.

for the cross variogram:

Equal contribution cases 1 and 2:

$$\gamma_{ZY}(\mathbf{h}) = 0.35Sph(|\mathbf{h}|/1) + 0.35Sph(|\mathbf{h}|/5)$$

$$\gamma_{ZY}(\mathbf{h}) = 0.15Sph(|\mathbf{h}|/1) + 0.15Sph(|\mathbf{h}|/5)$$

Focus on short-scale cases 3 and 4:

$$\gamma_{ZY}(\mathbf{h}) = 0.5Sph(|\mathbf{h}|/1) + 0.2Sph(|\mathbf{h}|/5)$$

$$\gamma_{ZY}(\mathbf{h}) = 0.3Sph(|\mathbf{h}|/1)$$

Focus on long-scale cases 5 and 6:

$$\gamma_{ZY}(\mathbf{h}) = 0.2Sph(|\mathbf{h}|/1) + 0.5Sph(|\mathbf{h}|/5)$$

$$\gamma_{ZY}(\mathbf{h}) = 0.3Sph(|\mathbf{h}|/5)$$

Cases 1 and 2 correspond to an "intrinsic" case where the shape of the cross-variogram is identical to the direct variograms. Because the ratio of short-scale contribution to the long-scale contribution is "1" in the direct variogram (i.e. $\gamma_{ZZ}(\mathbf{h})$, "Equal contribution" stated in Cases 1 and 2 also represents the ratio of 1. The short-scale contribution is increased to its maximum allowable under the linear model of coregionalization in Cases 3 and 4. The long-scale structure is maximum in Cases 5 and 6. The upscaled values of the correlation coefficient are given in Figure 3 for each case. The value of the correlation coefficient does not depend on volume scale for the equal contribution cases; however, increasing the contribution of short-scale decreases the correlation coefficient and increasing the contribution of long-scale increases the correlation coefficient. For large averaging volumes, the volume-dependent correlation coefficient stabilizes to a plateau-value.

Sensitivity on the Nugget Effect of Cross-Variogram

For this case, direct variograms again were fixed and different cases of nugget effects of the cross-variograms were considered. Cross-correlation coefficient of 0.7 is used. The direct variograms and all the cross-variograms are presented in Figure 4. The direct variograms:

$$\gamma_{ZZ}(\mathbf{h}) = \gamma_{YY}(\mathbf{h}) = 0.3 + 0.7Sph(|\mathbf{h}|/2.5)$$

Equal contribution case 7:

$$\gamma_{ZY}(\mathbf{h}) = 0.21 + 0.49Sph(|\mathbf{h}|/2.5)$$

Largest nugget cases 8 and 9:

$$\gamma_{ZY}(\mathbf{h}) = 0.3 + 0.4Sph(|\mathbf{h}|/2.5)$$

$$\gamma_{ZY}(\mathbf{h}) = 0.3$$

No nugget cases 10 and 11:

$$\gamma_{ZY}(\mathbf{h}) = 0.7Sph(|\mathbf{h}|/2.5)$$

$$\gamma_{ZY}(\mathbf{h}) = 0.3Sph(|\mathbf{h}|/2.5)$$

The same ratio of 0.3/0.7, used in the direct variogram (i.e. $\gamma_{ZZ}(\mathbf{h})$, is preserved in the "Equal contribution" cases. The upscaled values of the correlation coefficient are given in Figure 5. It is seen that, again, equal contribution does not effect the value of the correlation coefficient for successive volume scaling; however, increasing the contribution of the nugget effect decreases the correlation coefficient. Again, for large averaging volumes the volume-dependent correlation coefficient reaches a plateau-value.

Sensitivity on the Asymmetry of Direct Variogram Structures

In this sensitivity analysis the cross-variogram is fixed and different direct variograms are considered. Both the direct variograms and the cross variograms are presented in Figure 6. The cross variogram is fixed at $\gamma_{ZY}(\mathbf{h}) = 0.35Sph(|\mathbf{h}|/1) + 0.35Gauss(|\mathbf{h}|/5)$

No asymmetry case 12:

$$\gamma_{ZZ}(\mathbf{h}) = \gamma_{YY}(\mathbf{h}) = 0.5Sph(|\mathbf{h}|/1) + 0.5Gauss(|\mathbf{h}|/5)$$

High asymmetry case 13:

$$\gamma_{ZZ}(\mathbf{h}) = 0.15Sph(|\mathbf{h}|/1) + 0.85Gauss(|\mathbf{h}|/5)$$

$$\gamma_{YY}(\mathbf{h}) = 0.85Sph(|\mathbf{h}|/1) + 0.15Gauss(|\mathbf{h}|/5)$$

Partial asymmetry cases 14:

$$\gamma_{ZZ}(\mathbf{h}) = 0.3Sph(|\mathbf{h}|/1) + 0.7Gauss(|\mathbf{h}|/5)$$

$$\gamma_{YY}(\mathbf{h}) = 0.7Sph(|\mathbf{h}|/1) + 0.3Gauss(|\mathbf{h}|/5)$$

The term "asymmetry" represents the ratio of contribution of each nested structures. As this ratio decreases, asymmetry increases. Therefore, "no asymmetry" case corresponds to "equal contribution" which also corresponds to contribution ratio of 1. The upscaled values of the correlation coefficient are given in Figure 7. As we have seen before, the equal contribution (no asymmetry) does not effect the value of the correlation coefficient; however, increasing the asymmetry of direct variograms increases the correlation coefficient and this increase is proportional directly to the magnitude of the considered asymmetry ratio. Once more, the volume-dependent correlation coefficient reaches a stabilized value.

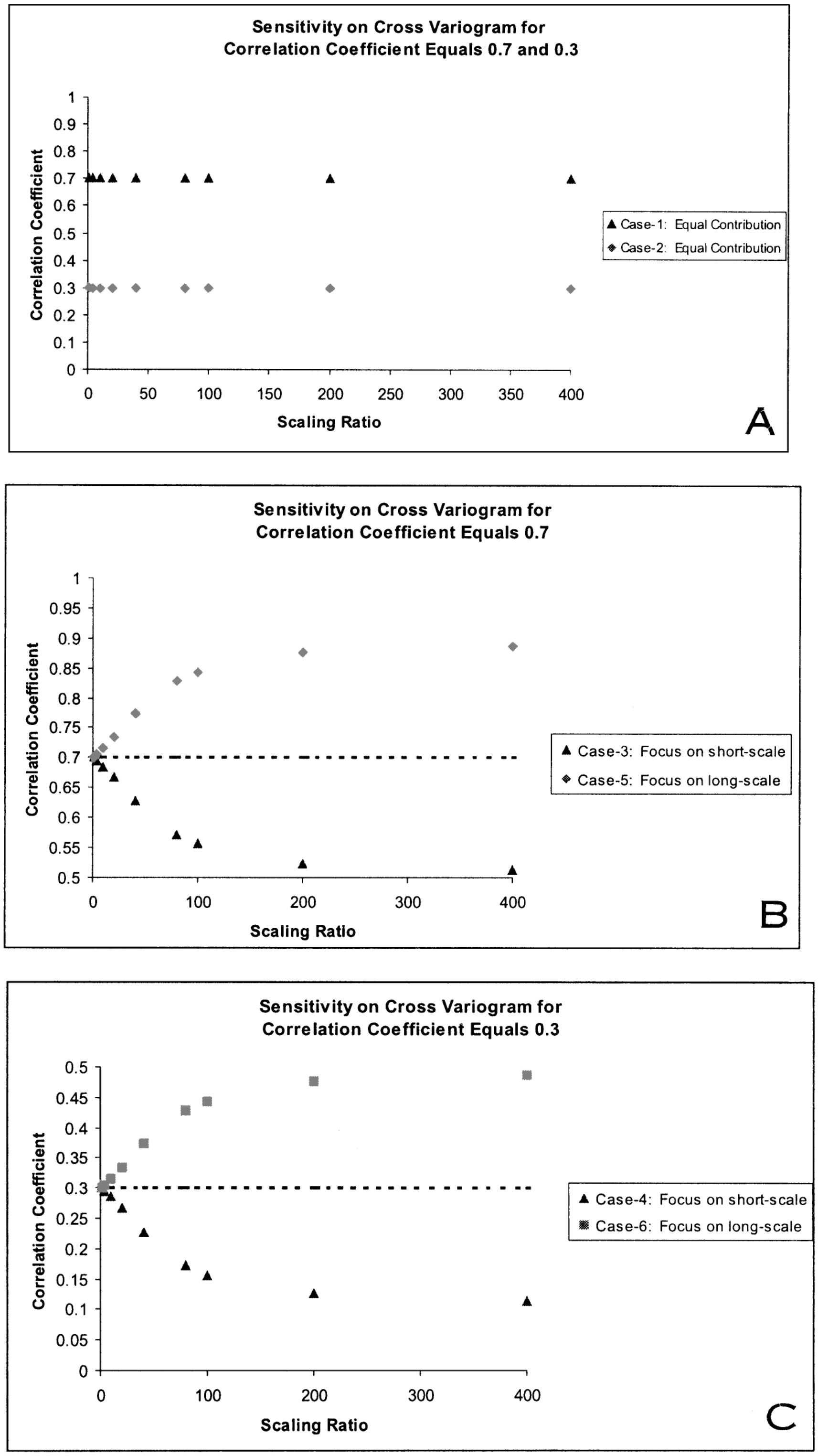

Figure 3. Sensitivity runs for cross variogram to illustrate effects of equal contribution, focusing on short-scale and long-scale. A, Sensitivity on cross-variogram to illustrate effect of equal contribution of each structure in cross-variogram. Correlation coefficient fixed to 0.7 for cross-variogram; B, Sensitivity on cross variogram to illustrate effect of focusing on long scale and short scale. Correlation coefficient fixed to 0.7 for cross-variogram; C, Sensitivity on cross-variogram to illustrate effect of focusing on long scale and short scale. Correlation coefficient fixed to 0.3 for cross-variogram.

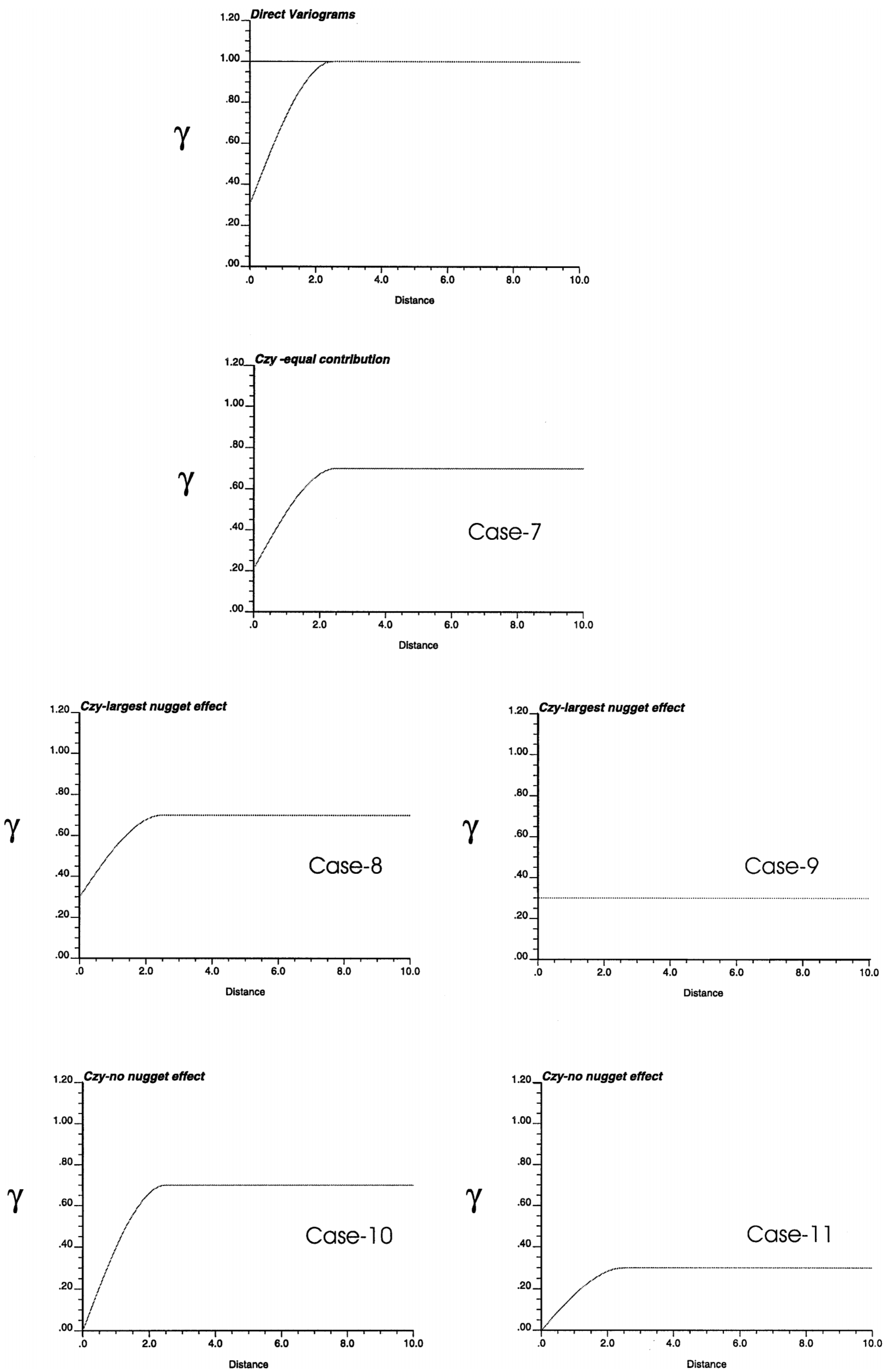

Figure 4. Fixed direct variograms and different cross variograms for cases of cross variogram nuggeteffect sensitivity.

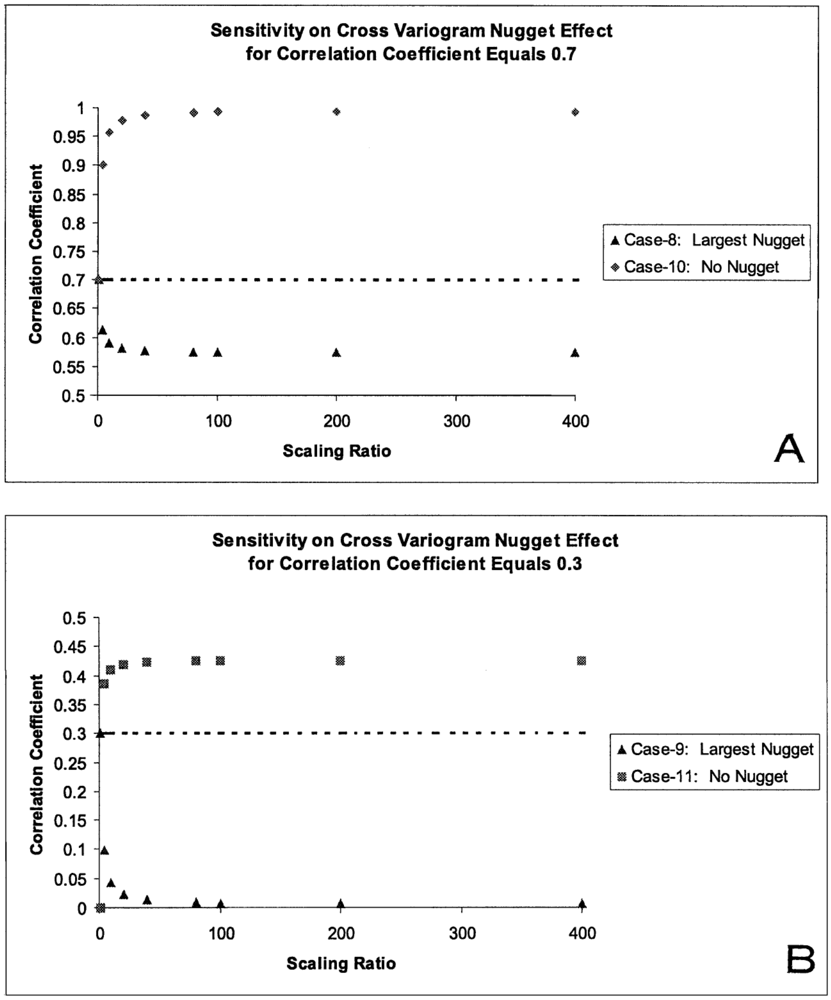

Figure 5. Sensitivity runs for cross-variogram nugget effect to illusrate effects of equal contribution, largest, and no nugget. A, Sensitivity on cross-variogram nugget effect to illustrate effect of largest and no nugget effect; B, Correlation coefficient fixed to 0.7 for cross-variogram.

AN APPLICATION TO REAL DATA

A real field example is investigated to see if the results from real data are consistent with theoretical results. This is an important step because this validation will identify shortcomings in current theory and prompt research into analytical relations.

A 500 by 500 pixel satellite image of Wadi Kufra, Libya was used. The "RGB" values of each pixel were extracted, but only, "red" and "blue" color values of were used. These two values are colocated and correlated. The histogram plots of both red and blue data are given by Figure 8. Note that the frequency distribution of red values is close to normal and the frequency distribution for blue values has a long tail more like a lognormal distribution. After the scatter plot of red data versus blue data, we determined the correlation coefficient as 0.708.

Direct and cross variograms were calculated and a linear model of coregionalization (LMC) was fitted. Recall that for a valid LMC, the auto and cross-variogram models must be constructed using the same basic variogram models. The experimental directional (points) and the modelled direct and cross variograms (solid lines) are given in Figure 9. Two nested

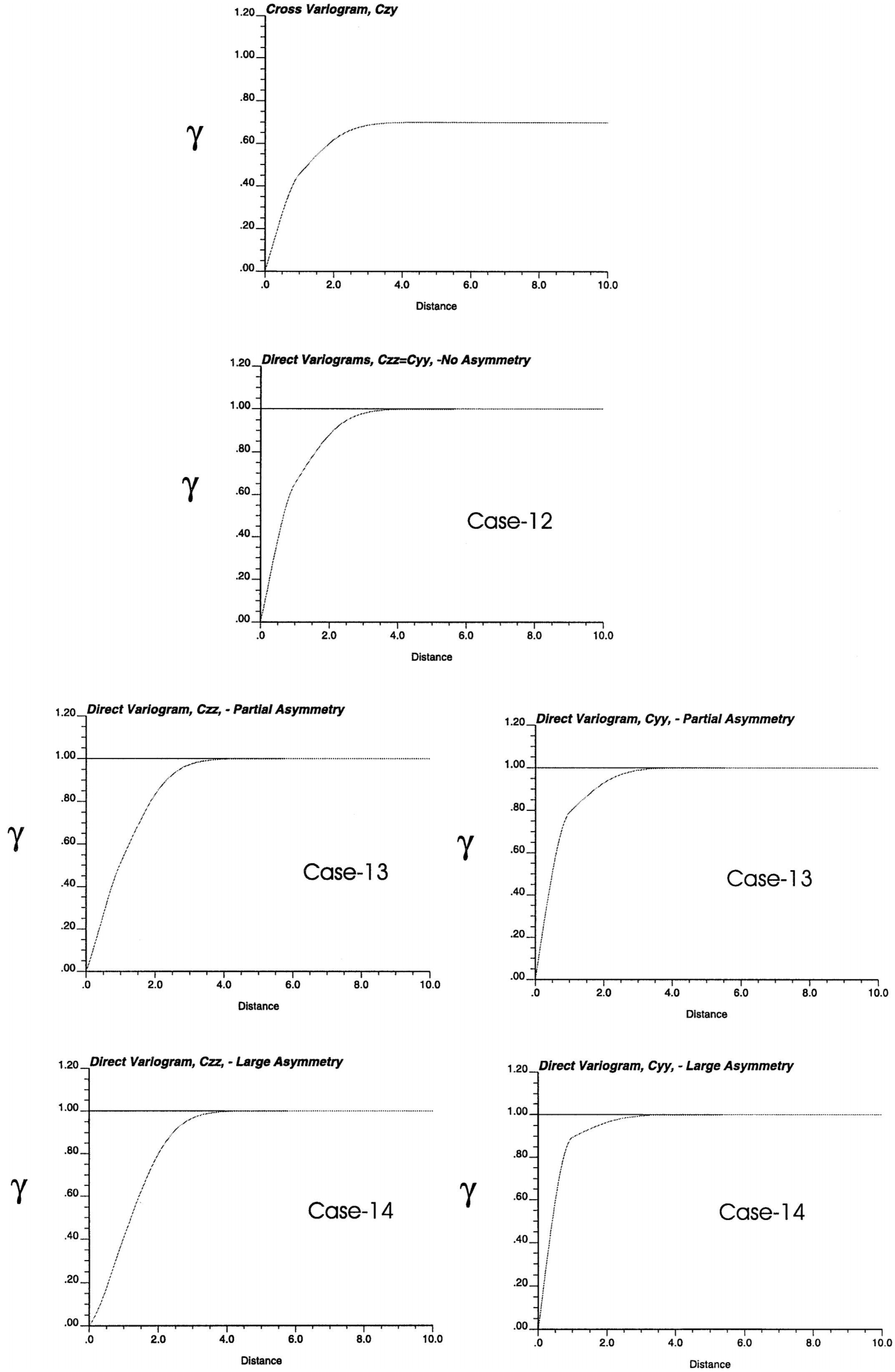

Figure 6. Fixed direct variogram and different cross-variograms for cases of cross-variogram asymmetry sensitivity.

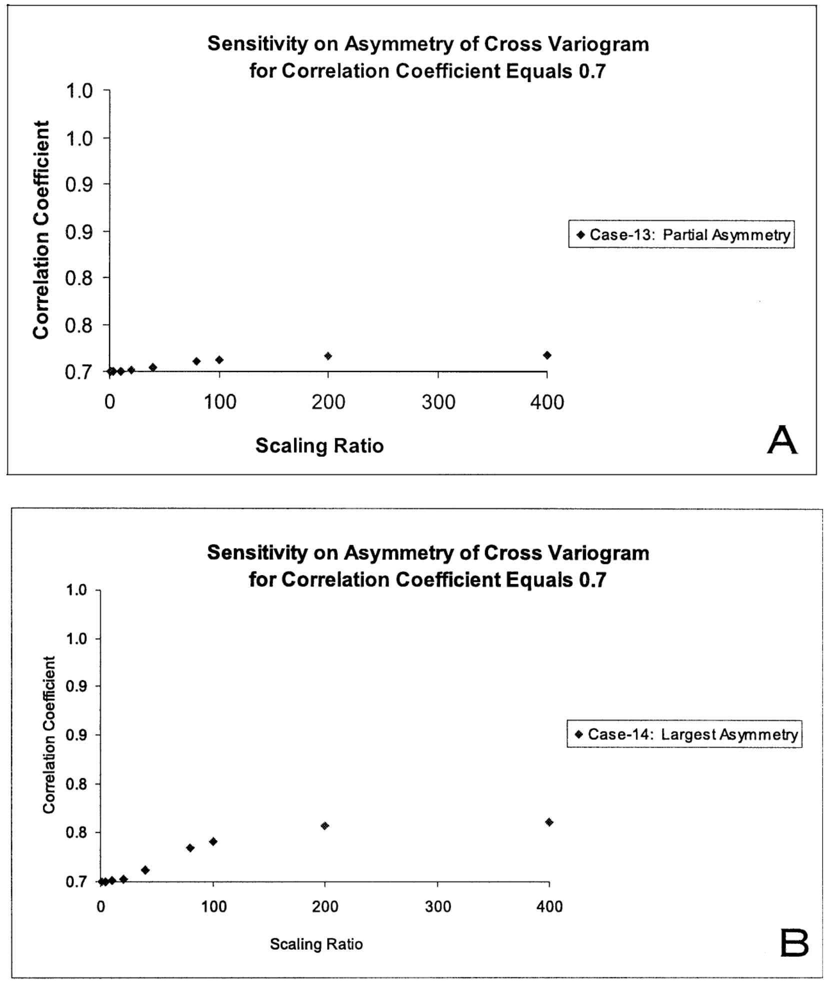

Figure 7. Sensitivity runs for cross-variogram asymmetry to illustrate effects of equal contribution, partial, and large asymmetry. A, Sensitivity on asymmetry of cross variogram to illustrate effect of partial asymmetry. Correlation coefficient fixed to 0.7 for cross variogram; B, Sensitivity on asymmetry of cross variogram to illustrate effect of largest asymmetry. Correlation coefficient fixed to 0.7 for cross variogram.

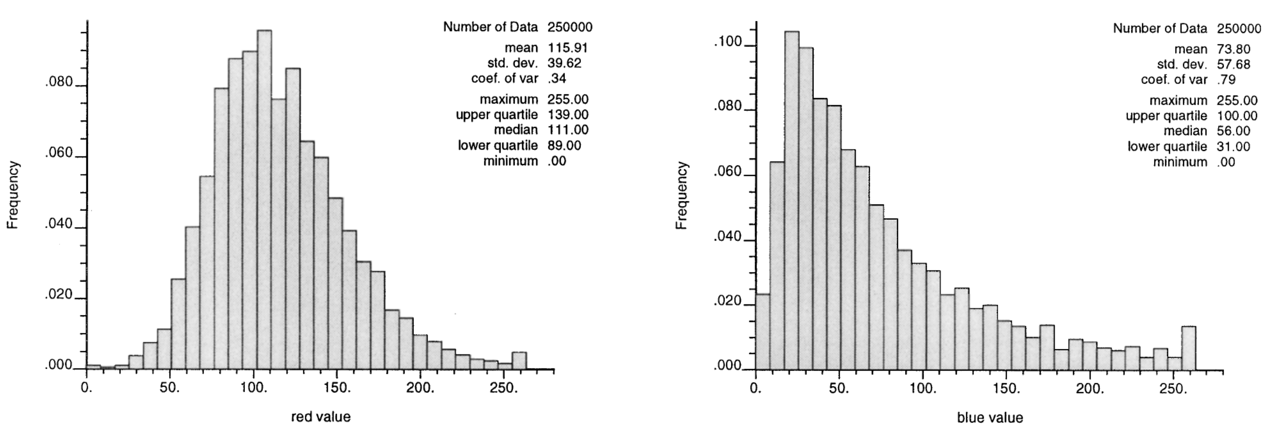

Figure 8. Histogram of two images: Right, histogram for red data; Left, histogram for blue data.

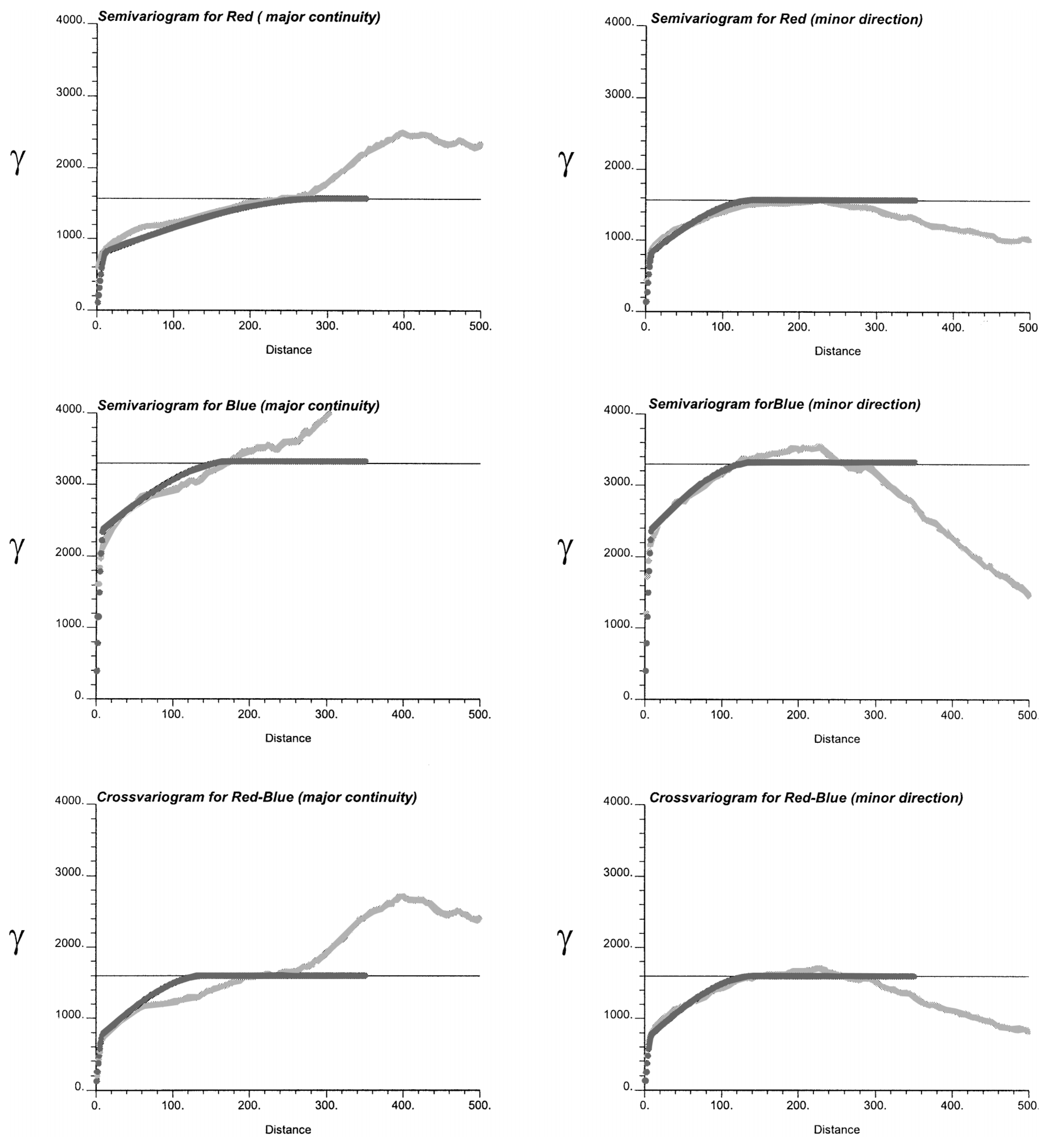

Figure 9. Experimental directional direct and cross variograms for "red" and "blue" data along with fitted LMC model.

spherical models without nugget effect were used:

$$\begin{aligned} \gamma_{red}(h) &= 770Sph_{(8,10)} + 800Sph_{(380,105)} \\ \gamma_{blue}(h) &= 2300Sph_{(8,10)} + 1027Sph_{(380,105)} \quad (16) \\ \gamma_{redblue}(h) &= 695Sph_{(8,10)} + 900Sph_{(380,105)} \end{aligned}$$

Because the sill (contribution of each nested structure) values of the direct variograms (red and blue) are greater than zero and $770 \cdot 2300 > 695^2$ and $800 \cdot 1027 > 900^2$, the LMC given in Equation (16) is positive definite.

A 2D linear averaging was applied using 2 by 2, 5 by 5, 10 by 10, 20 by 20, 25 by 25, 50 by 50, 100 by 100, and 250 by 250 block dimensions. A new cross-correlation coefficient was calculated for each upscaled set of images. Using the LMC model in Equation (16), and the definition of volume-dependent cross-correlation coefficient given by Equation (13), cross correlations were calculated numerically from theory by using the `VarScale` program (Oz, Deutsch, and Frykman, 2000). The comparison of the experimental and theoretical volume-dependent cross-correlation coefficients is presented in Figure 10. The correlation coefficient increases and approachs a steady-state value gradually after the averaging volume of 50. This increase is the result of the sill contribution of the large-scale nested structure component of the cross-variogram model (see earlier discussion). For the small blocks, the experimental and theoretical results are not in a good agreement. The difference for the small blocks might be explained by the existence of spatial correlation at small scales that the variogram cannot fully capture. Also, there is a clear lack of stationarity which is identified from the variograms given by Figure 9. Notwithstanding this small mismatch for the small blocks, the general conformity in the trends of two curves (see Fig. 10) encourages us to seek for analytical relations between the averaging volume and the volume-dependent cross-correlation coefficient.

ANALYTICAL ANALYSIS

The characteristics of the volume-dependent cross-correlation coefficient would be understood better by analytical relations. The theoretical equation and numerical solution are brute force with little recourse for understanding except through repeated numerical experiments. The terms controlling the volume-dependent cross-correlation coefficient will be investigated more completely here. We start

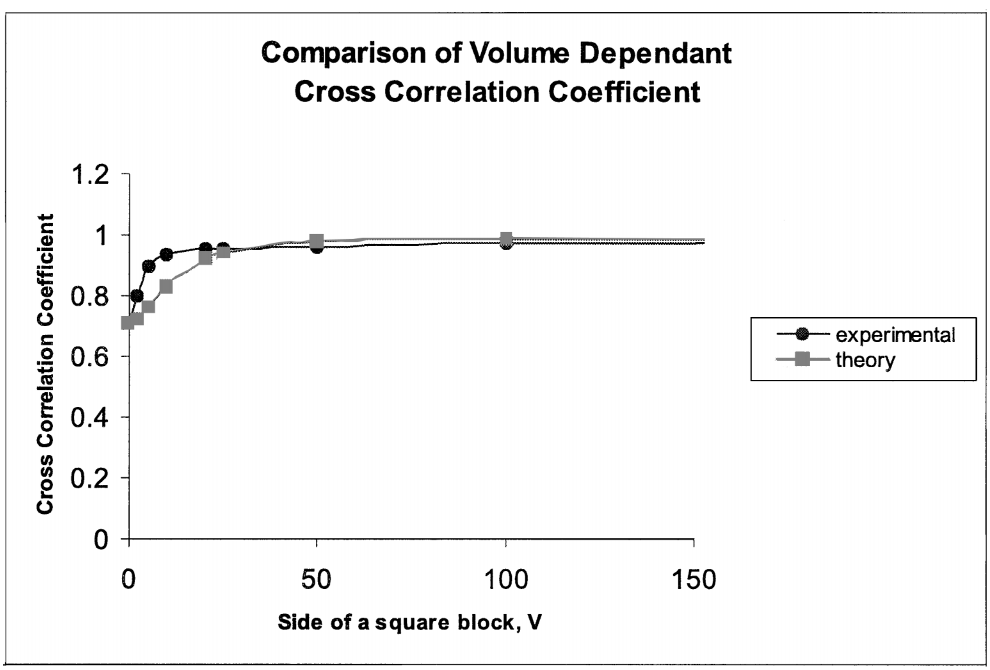

Figure 10. Comparison of cross correlation coefficient.

with differentiating Equation (7):

$$\frac{\partial\rho(V)}{\partial V} = \left(\frac{\partial\bar{C}_{ZY}(V)}{\partial V}\right)(\bar{C}_{ZZ}^{-1/2}(V))(\bar{C}_{YY}^{-1/2}(V)) - \frac{1}{2}\left(\frac{\partial\bar{C}_{ZZ}(V)}{\partial V}\right)(\bar{C}_{ZZ}^{-3/2}(V))(\bar{C}_{ZY}(V)) \times(\bar{C}_{YY}^{-1/2}(V)) - \frac{1}{2}\left(\frac{\partial\bar{C}_{YY}(V)}{\partial V}\right) \times(\bar{C}_{YY}^{-3/2}(V))(\bar{C}_{ZY}(V))(\bar{C}_{ZZ}^{-1/2}(V)) \quad (17)$$

The correlation coefficient at large scale can be written as:

$$\rho(V) = \rho(v_0 + \Delta V) = \rho(v_0) + \frac{\partial\rho(V)}{\partial V}\Delta V \quad (18)$$

The right-hand side of the Equation (18) is the volume-dependent cross-correlation coefficient where, $V = v_0 + \Delta V$ represents the volume at larger scales. Moreover, v_0 is the point-scale, $\rho(v_0)$ is the point-scale cross-correlation coefficient and $\Delta V = V - v_0$ is the volume difference.

Equation (18) is the general equation that can be used to calculate volume-dependent cross-correlation coefficient. The $\frac{\partial\rho(V)}{\partial V}$ term, given in Equation (17), is important to determine in terms of the characteristic path of volume-dependent cross-correlation coefficient. Depending on its rate of change, the cross-correlation coefficient at larger scale also may increase or decrease. Calculation of $\frac{\partial\rho(V)}{\partial V}$ term is mainly controlled by:

$$\left(\frac{\partial\bar{C}(V)}{\partial V}\right) = \frac{\partial}{\partial V}\int_0^V\int_0^V C(x - x')\,dx\,dx' \quad (19)$$

The differentation of dispersion variances is the inverse of computing auxiliary fuctions (Journel and Huijbregts, 1978; Vargas-Guzman, Myers, and Warrick, 2000) for the average variograms or cross variograms. We can readily calculate Equation (19) numerically or simplify it by applying Leibniz's Theorem twice. The terms in Equation (19) control the $\frac{\partial\rho(V)}{\partial V}$ term [Equation (17)] and $\rho(v_0 + \Delta V)$ term [Equation (18)]. They are highly nonlinear functions and their further simplifications depend on the LMC model.

Although it will not be a general solution, we want to go one more step and present a simple approximate solution to the Equation (18). Let us assume that we have same direct variograms (i.e. $\gamma_{ZZ}(\mathbf{h}) = \gamma_{YY}(\mathbf{h})$). Then Equation (17) reduces to:

$$\frac{\partial\rho(V)}{\partial V} = \frac{\left(\frac{\partial\bar{C}_{ZY}(V)}{\partial V}\right)}{\bar{C}_{ZZ}(V)} - \frac{\left(\frac{\partial\bar{C}_{ZZ}(V)}{\partial V}\right)}{\bar{C}_{ZZ}^2(V)}\bar{C}_{ZY}(V) \quad (20)$$

If we assume one-dimensional upscaling with length l is not greater than the variogram range, a then, we can approximate the average variogram as (Journel and Huijbregts, 1978):

$$\bar{\gamma}(l) = C\left[\frac{l}{2a} - \frac{l^3}{20a^3}\right], \quad \forall\mathbf{a} \geq l \quad (21)$$

where C is the sill contribution or variance. Equation (21) can be written in terms of average covariance as:

$$\bar{C}(l) = C\left[1 - \frac{l}{2a} - \frac{l^3}{20a^3}\right], \quad \forall\mathbf{a} \geq l \quad (22)$$

By taking the derivative of Equation (22), we can estimate the value of $\frac{\partial\bar{C}(l)}{\partial l}$ as:

$$\frac{\partial\bar{C}(l)}{\partial l} = C\left[-\frac{0.5}{a} + \frac{0.15l^2}{a^3}\right], \quad \forall\mathbf{a} \geq l \quad (23)$$

By using the Equations (22) and (23), we can estimate easily the value of volume-dependent cross-correlation coefficient when $\mathbf{a} \geq l$. This type of approximation may be used when we are upscaling along the wellbore from core-scale to log-scale.

Actually, when we analyze the values calculated from Equation (23), $\frac{0.5}{a}$ is the dominant term and after some larger averaging values of l, the second term $\frac{0.15l^2}{a^3}$ also contributes. Therefore we can assume that:

$$\frac{\partial\bar{C}(l)}{\partial l} \cong -C\frac{0.5}{a} \quad (24)$$

We can rewrite the Equation (21) assuming:

$$\bar{\gamma}(l) \cong C\frac{l}{2a} \quad (25)$$

We can relate the Equations (24) and (25):

$$\frac{\partial\bar{C}(l)}{\partial l} \cong -\frac{\bar{\gamma}(l)}{l} \quad (26)$$

Although it has some limititations, Equation (26) is a straightforward relationship depending on the average variogram and averaging length.

In order to test the efficiency of this approximation for 1D averaging, we used the same LMC model as Equation (15) to estimate the correlation coefficients for different length scales. Our results show an error of about 3 to 5% for small averaging lengths, when we apply 1D scaling. When we go to averaging lengths larger than the variogram range, we can use another form of auxiliary function (Vargas-Guzman, Myers, and Warrick, 2000) for the spherical variogram:

$$\bar{C}(l) = C\left[0.75 * \frac{a}{l} - 0.2 * \frac{a^2}{l^2}\right], \quad \forall \mathbf{l} \geq a \qquad (27)$$

Then the derivative is given as:

$$\frac{\partial \bar{C}(l)}{\partial l} = C\left[-0.75 * \frac{a}{l^2} + 0.4 * \frac{a^2}{l^3}\right], \quad \forall \mathbf{l} \geq a \qquad (28)$$

By using the appropriate forms of Equations (27) and (28) in Equation (17), we can estimate the volume-dependent cross-correlation coefficient via Equation (18) for large averaging volumes.

Our aim here is to explore the governing equations of volume-dependent cross-correlation coefficient and hightlight some critical terms. It is intractable to present a general analytical equation for complex coregionalization models. In general, we resort to numerical techniques.

LIMIT VALUE OF VOLUME-DEPENDENT CROSS-CORRELATION COEFFICIENT

As shown in all our case studies, the volume-dependent cross-correlation coefficient converges to a specific limit or "plateau" value for large averaging volumes. In order to estimate this limit value, we need to seek for a solution to Equation (29):

$$\lim_{V\to\infty} \rho(V) \qquad (29)$$

Inserting the definition of $\rho(V)$ we get:

$$\lim_{V\to\infty} \times \frac{\int_0^V\int_0^V C_{ZY}(x-x')\,dx\,dx'}{\sqrt{\int_0^V\int_0^V C_{ZZ}(x-x')\,dx\,dx'} \cdot \sqrt{\int_0^V\int_0^V C_{YY}(x-x')\,dx\,dx'}} \qquad (30)$$

Equation (30) is a general equation for the limit value of volume-dependent cross-correlation coefficient. Without going through intermediate steps, we are giving directly a solution to Equation (30) by assuming variograms are isotropic spherical models:

$$\lim_{V\to\infty} \rho(V) = \frac{\sum_{i=1}^{nst} C_{ZY}^i a^i}{\sqrt{\left[\sum_{j=1}^{nst}\sum_{i=1}^{nst} C_{ZZ}^i C_{YY}^j a^i a^j\right]}} \qquad (31)$$

where (C^i) values are the sill contribution of either direct variograms or cross variograms and the (a^i) are the range values of the isotropic variograms or cross variograms for the corresponding nested structure component.

Assuming that direct variograms are same and LMC model is composed of two nested structures, then we can rewrite Equation (31):

$$\lim_{V\to\infty} \rho(V) = \frac{C_{ZY}^1 a^1 + C_{ZY}^2 a^2}{\sqrt{\left[C_{ZZ}^1 a^1 + C_{ZZ}^2 a^2\right]^2}} \qquad (32)$$

Now, let us calculate the asymptotic value for *Case 3* in Exploratory Research section:

$$\lim_{V\to\infty} \rho(V) = \frac{0.5 * 1 + 0.2 * 5}{\sqrt{[0.5 * 1 + 0.5 * 5]^2}} = 0.5 \qquad (33)$$

From the numerical calculations, this limit value is expected to be between 0.5 and 0.51, which is close to our analytical limit value of 0.50. When we look for the limit value for the *Case 5*, we get 0.90, which is close to the numerically estimated one.

INTERPRETATION OF RESULTS AND CONCLUSIONS

- The theory of volume-dependent cross-correlation coefficient is explained and a general definition is provided by Equation (13). The dependence of volume-dependent cross-correlation coefficient on dispersion variance and dispersion covariance has been discussed. The concept and the calculation precedures for dispersion variances and dispersion covariances are presented. A numerical example is given to illustrate a solution to the volume-dependent cross-correlation coefficient given by Equation (13).
- The cross-correlation exhibits a functional relationship to averaging volumes. It can increase or decrease with as volume support increases depending on the relative importance of long- and short-scale variogram structures. If the direct and cross variograms are proportional, there is no change in the cross correlation as

the averaging volume changes. After some averaging volume, the volume-dependent cross-correlation coefficient reaches a stabilized-value. This plateau value is controlled mainly by the large-scale nested structure component of cross-variogram and direct variograms. Our study also shows that volume-dependent cross-correlation coefficient is sensitive to the shape and sill contribution structure of cross-variogram and the asymmetry of the two direct variograms

- Increasing the contribution of long-scale variogram structures in the cross-variogram increases the correlation coefficient; increasing the contribution of short-scale decreases the correlation coefficient. Increasing the asymmetry of the direct variograms increases the correlation coefficient. The volume-dependent correlation coefficient stabilizes and reaches a plateau-value for large averaging volumes.
- The trend for the numerically calculated volume-dependent cross correlations and the ones obtained from a real field example are promising. Besides, they both approach to the same limit or plateau value. This prompts us to seek for analytical realations to estimate cross-correlation coefficient as a function of averaging volume.
- The general equations explaining the dependency of the cross correlation on averaging volume have been presented and explained. Understanding the nature of this scaling for different coregionalization models is practically important for the usual situation of data integration where the data are at different volume support, for example, large-scale seismic data and small-scale well-log data. The controlling factors and limit values for large averaging volumes have been derived.
- Additional work is warranted to extend the analytical results to make them applicable to the complexity of real problems; however, the numerical solution is fast, accurate, and adequate in all situations.
- Because cross correlation is the key element for data-integretion techniques, the LMC model of coregionalization should be selected carefully. A wrong LMC model may cause cross correlation to decrease instead of increasing and vice versa. A significant conclusion of this paper is that the volume-dependent cross-correlation should be determined from the available data instead of assuming that it is independent of scale.

REFERENCES

Almeida, A., and Journel, A. G., 1994, Joint simulation of multiple variables with Markov-type coregionalization model: Math. Geology, v. 26, no. 5, p. 565–568.

Behrens, R. A., and Tran, T. T., 1998, Incorporating seismic data of intermediate vertical resolution into 3D Reservoir: SPE Ann. Tech. Conf. Exhibition, Soc. of Petroleum Engineers (New Orleans, LA), SPE Paper Number 49143, p. 1– 11.

Behrens, R. A., MacLeod, M. K., Tran, T. T., and Alimi, A. O., 1996, Incorporating seismic attribute maps in 3D reservoir models: 1996 SPE Ann. Tech. Conf. and Exhibition, Soc. Petroleum Engineers (Denver, CO), SPE Paper Number 36499, p. 122–126.

David, M., 1977, Geostatistical ore reserve estimation: Elsevier Publ., Amsterdam, 213 p.

Deutsch, C. V., and Journel, A. G., 1997, GSLIB: Geostatistical Software Library and user's guide (2nd edn.): Oxford Univ. Press, New York, 369 p.

Deutsch, C. V., Srinivasan, S., and Mo, Y., 1996, Geostatistical reservoir modeling accounting for the scale and precision of seismic data: 1996 SPE Ann. Tech. Conf. and Exhibition, Soc. Petroleum Engineers (Denver, CO), SPE Paper Number 36497, p. 9–19.

Doyen, P. M., 1988, Porosity from seismic data: a geostatistical approach: Geophysics, v. 53, no. 10, p. 1263–1275.

Doyen, P. M., den Boer, L. D., and Pillet, W. R., 1996, Incorporating seismic attribute maps in 3D reservoir models: 1996 SPE Ann. Tech. Conf. Exhibition, Soc. Petroleum Engineers (Denver, CO), SPE Paper Number 36498, p. 21–30.

Doyen, P. M., Psaila, D. E., and Strandenes, S., 1994, Bayesian sequential indicator simulation of channel sands from 3D seismic data in the Oseberg field, Norwegian North Sea: 69th Ann. Tech. Conf. Exhibition, Soc. Petroleum Engineers (New Orleans, LA), SPE Paper Number 28382, p. 197–212.

Goovaerts, P., 1997, Geostatistics for natural resources evaluation: Oxford Univ. Press, New York, 483 p.

Journel, A. G., 1999a, Conditioning geostatistical operation to non-linear volume averages: Math. Geology, v. 31, no. 8, p. 931–953.

Journel, A. G., 1999b, Markov models for cross covariances: Math. Geology, v. 31, no. 8, p. 955–963.

Journel, A. G., and Huijbregts, Ch. J., 1978, Mining geostatistics: Academic Press, New York, 600 p.

Journel, A. G., and Zhu, H., 1990, Integrating soft seismic data: Markov-Bayes updating, an alternative to cokriging and traditional regression: Stanford Center for Reservoir Forecasting, Rept. 3, 25 p.

Kupfersberger, H., Deutsch, C. V., and Journel, A. G., 1998, Deriving constraints on small-scale variograms due to variograms of large-scale data: Math. Geology, v. 30, no. 7, p. 837–851.

Oz, B., Deutsch, C. V., and Frykman, P., 2000, A visual Basic program for histogram and variogram scaling: Computers & Geosciences, accepted.

Shmaryan, L. E., and Journel, A. G., 1999, Two Markow models and their applications: Math. Geology, v. 31, no. 8, p. 965–988.

Tran, T. T., Wen, X.-H., and Behrens, R. A., 1999, Efficient conditioning of 3D fine-scale reservoir model to multiphase production data using streamline-based course-scale inversion and geostatistical downscaling: SPE Ann. Tech. Conf. C (Houston, TX), SPE Paper Number 56518, p. 1–13.

Vargas-Guzman, J. A., Myers, D. E., and Warrick, A. W., 2000, Derivatives of spatial variances of growing windows and the variogram: Math. Geology, v. 32, no. 7, p. 851–871.

Vargas-Guzman, J. A., Warrick, A. W., and Myers, D. E., 1999a, Multivariate correlation in the framework of support and spatial scales of variability: Math. Geology, v. 31, no. 1, p. 85–103.

Vargas-Guzman, J. A., Warrick, A. W., and Myers, D. E., 1999b, Scale effect on principal component analysis for the vector random functions: Math. Geology, v. 31, no. 6, p. 701–722.

Xu, W., Tran, T. T., Srivastava, R. M., and Journel, A. G., 1992, Integrating seismic data in reservoir modeling: the collocated cokriging alternative: 67th Ann. Tech. Conf. Exhibition, Soc. Petroleum Engineers (Washington, DC), SPE Paper Number 24742, p. 833–842.