

Size Scaling of Cross-Correlation between Multiple Variables

Bora Oz and Clayton V. Deutsch¹

Department of Civil & Environmental Engineering, University of Alberta

Abstract

Reservoir models have large uncertainty due to spatial variability and limited sample data. Our 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. This seismic data is 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 has been 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 stabilized-value. This plateau value is mainly controlled by the large-scale behaviour 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 asymmetry or differences between the direct variograms of the well data and seismic data.

Keywords: data-integration, correlation coefficient, volume scaling, dispersion variance, dispersion covariance

Introduction

Reconciling different data types for spatial modeling of reservoir properties is important because different data provide complementary information about the reservoir architecture and heterogeneity. There are a variety of methods to integrate different data types including *External Drift, Locally Varying Mean, Block Kriging* [2, 3, 7, 16], *Block Cokriging* [9], *Markov-Bayes* (or *Bayesian Updating Rule*) [4, 10, 11, 12, 17, 21], *Truncated Gaussian simulation* [5] and *Collocated cokriging* [1, 10, 25]. Details and application examples of these methods are given in literature [7, 8, 16]. The main aim of this paper is to address a common requirement: the cross-covariance between multiple data types.

¹Corresponding author: Fax: +1-780-492-0249; e-mail:(cdeutsch@civil.ualberta.ca)

The problem of how to handle 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 varying mean algorithms assume the spatial variability of the secondary variable gives information on trends in the primary variable. This approach does not fully capture the spatial cross-correlation.

A better approach to data integration is cokriging, which uses a cross-covariance that explicitly measures the information content of the secondary data with regard to the primary variable [14]. Not only is the computational and inference more burdensome, 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. Inference of the cross-covariance becomes a problem.

A Markov-type assumption [1, 15, 20] simplifies inference of the cross variogram, but is not valid when the secondary variable is defined on a much larger support than the primary variable. For such cases, a different Markov hypothesis is proposed [15, 20], leading to a different cross-covariance model. The cross-covariance, for this case, is defined as the function of secondary variable (have large measurement volume) covariance and again co-located or small-scale correlation coefficient, $\rho(0)$. In both approaches, there is no explicit formalism to specify the cross-covariance; it is rescaled from either primary or secondary variable covariance by the factor of small-scale correlation coefficient, which is treated as independent of scale.

In another study, Kupfersberger and coauthors [18] propose analytical equations to infer small-scale variograms with a combination of small-scale data and large-scale data. The key idea is to downscale large-scale variograms to small-scale and complete the horizontal directions of 3-D small scale variogram with more extensive secondary data.

The better we understand the size scaling behavior of cross-correlation, the more reliable our numerical models. 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 and coauthors [23, 24] have tackled a related question. They extend the concept of dispersion variance to the multivariate case 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 with the dispersion variance and dispersion covariance terms. A general equation for volume-dependent cross-correlation coefficient is presented and the critical terms forming this equation are interpreted. Some solution techniques for the governing equation are also discussed.

- **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:** The cross-variogram, nugget effect, and asymmetry of the direct variograms are looked at in detail to better understand the characteristics of the cross-correlation with respect to “upscaling”. The functional relationship of the cross correlation to scales is complex and depends on many factors.
- **Application:** Theory and practice are compared with real satellite data. Direct variograms and cross-variogram are calculated and upscaled to estimate the cross-correlation at different scales. The experimentally-obtained results are close to the theoretical and numerical results.
- **Analytical Analysis:** The complexity and non-linearity 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 cases. The asymptotic values of the cross-correlation as the scale becomes very large are considered; a good match is seen 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 chosen 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 kinds of spatial variability in almost all natural phenomena; local *random* aspects and general *structured* aspects [16]. 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 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 occurrences of the RV. “Stationarity”, that is spatial homogeneity, removes the location-dependent nature of the expected value,

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

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

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

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:

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

The variogram is defined as:

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

The relation between the stationary semivariogram and the stationary covariance is straightforwardly derived:

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

One important remark on Equations from 1 to 5 is that they are all at “point-scale”. For example the covariance given by Equation 3 is the measure of similarity of data values, which are \mathbf{h} distance away, at the point-scale. A challenge is to be able to calculate them at different scales. This is addressed in the next Section. Another implementation detail, that will not be addressed in this paper, is the estimation of the expected values in practical settings with limited data. This is an important and critical subject that is discussed in geostatistical texts and papers.

The elementary statistics described above 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:

$$\begin{aligned} \rho &= \frac{E\{Z(\mathbf{u}) \cdot Y(\mathbf{u})\} - m_Z m_Y}{\sigma_Z \sigma_Y} \\ &= \frac{C_{Z,Y}}{C_{Z,Z} C_{Y,Y}} \end{aligned} \quad (6)$$

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

The correlation coefficient is a standardized covariance that measures the linear dependence of the two variables. The value of ρ is always between -1 and +1. The positive values represent “direct’ relations; whereas, negative values represent “inverse” relations between two data types. It is important to note that ρ provides a measure of the linear relationship between two variables. If the relationship between two variables is not linear, the correlation coefficient will not adequately reflect the relationship between the two variables. Furthermore, the correlation coefficient is strongly affected by outlier data or extreme data pairs. The correlation coefficient can be artificially enhanced by an outlier pair that falls along the general trend of the data or artificially damaged by an outlier pair that does not fall along the lines of the general trend of the data pairs. The correlation coefficient, ρ , given by Equation 6 is a point-scale measure of correlation.

0.0.1 Volume-Dependent Correlation Coefficient

The equation for the data-scale correlation coefficient was presented above in Equation 6. The correlation coefficient at a scale v different than the data scale is defined as:

$$\rho(v) = (\bar{C}_{Z,Y}(v)) \cdot \left(\frac{1}{\sqrt{\bar{C}_{Z,Z}(v)}} \right) \cdot \left(\frac{1}{\sqrt{\bar{C}_{Y,Y}(v)}} \right) \quad (7)$$

where, $\bar{C}_{Z,Y}(v)$ is the volume-averaged cross-variogram, $\bar{C}_{zz}(v)$ and $\bar{C}_{yy}(v)$ are the volume-averaged direct variograms. These volume averaged covariances are classically defined as $\int_v \int_v C(\mathbf{u} - \mathbf{u}') d\mathbf{u} d\mathbf{u}'$, which is closely approximated by numerical integration.

As a side note, the sill value of a direct semivariogram is the variance. The variance at an arbitrary scale v is also called the “dispersion variance” and is equal to the $\bar{C}(v)$ when the area is large. 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 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:

$$\gamma_V(h) = C_V^0 + \sum_{i=1}^{nst} C_V^i \Gamma_V^i(h) \quad (8)$$

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, $i = 1, \dots, nst$, 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 C_V^i terms describe the variance contributions of each nested structure. The sum of the variance contribution is the variance at the V -scale and is also called the dispersion variance:

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

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 [16]:

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

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:

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

Although there exist certain analytical solutions [6, 16] to $\bar{\gamma}(V(\mathbf{u}), V(\mathbf{u}))$, the value of “gamma-bar” is usually estimated numerically by discretizing the volume $V(\mathbf{u})$ and $V(\mathbf{u})$ into a number of points and simply averaging the variogram values:

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

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 is used.

The values of dispersion variances and covariances allow calculation of the volume-dependent correlation coefficient:

$$\rho(v, A) = (D_{zy}^2(V, A)) \cdot \left(\frac{1}{\sqrt{D_{zz}^2(V, A)}} \right) \cdot \left(\frac{1}{\sqrt{D_{yy}^2(V, A)}} \right) \quad (13)$$

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 [19] can be used to calculate these dispersion variances and dispersion covariances. This program implements the classical “Geostatistical Scaling Laws” [16]. Consider an example to illustrate the numerical calculation of volume-dependent cross-correlation coefficient via Equation 13.

Numerical Verification

A full co-simulation technique was used to simulate a prior-defined linear model of coregionalization (LMC) [13, 16]. 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` from `GSLIB` [7] and an isotropic variogram model. Using this generated image as secondary data in `Sgsimfc` [7] (Sequential Gaussian Full CoSimulation) 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.

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

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

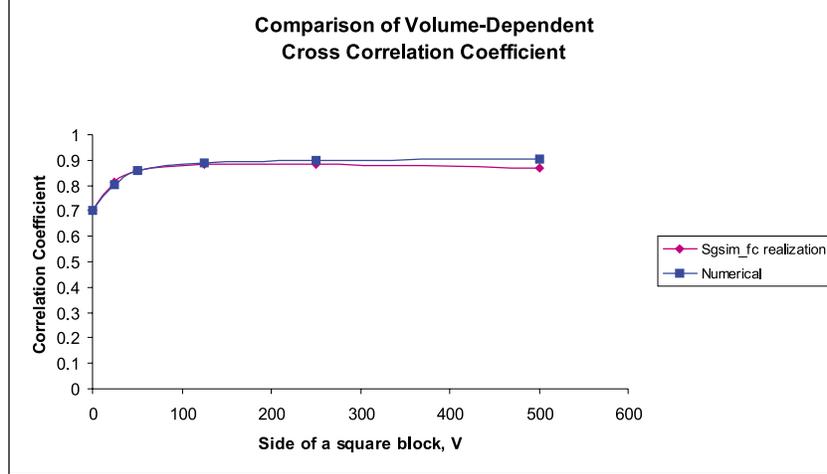


Figure 1: Illustration of cross-correlation coefficient from the results obtained by numerical calculation and upscaling the images generated by `Sgsimfc`

$$\gamma_Y(h) = 0.5Sph_{15} + 0.5Sph_{75} \quad (15)$$

$$\gamma_{ZY}(h) = 0.2Sph_{15} + 0.5Sph_{75}$$

The two images are upscaled and the corresponding volume-dependent cross-correlation coefficients 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 [19]. 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 cases 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 were considered, 0.7 and 0.3. There are three scenarios 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. 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 were again fixed and different cases of nugget effects of the cross variograms were considered. 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 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.

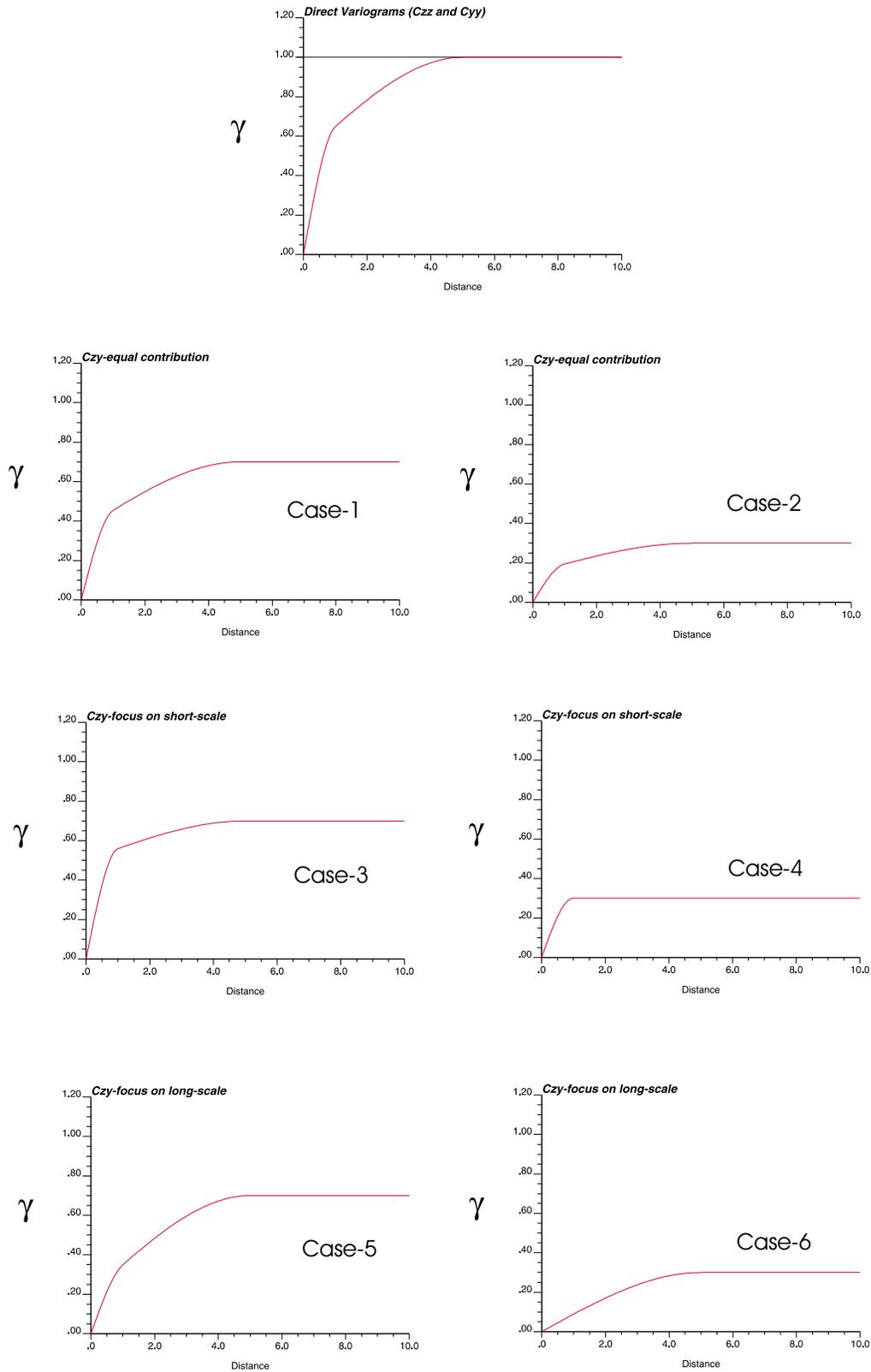
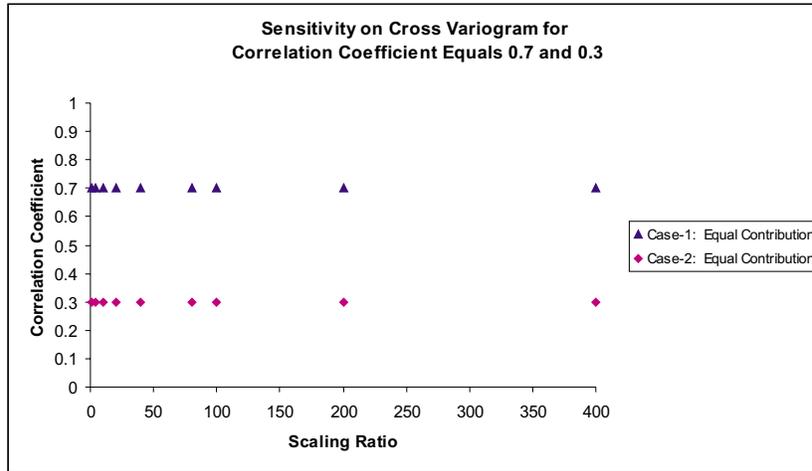
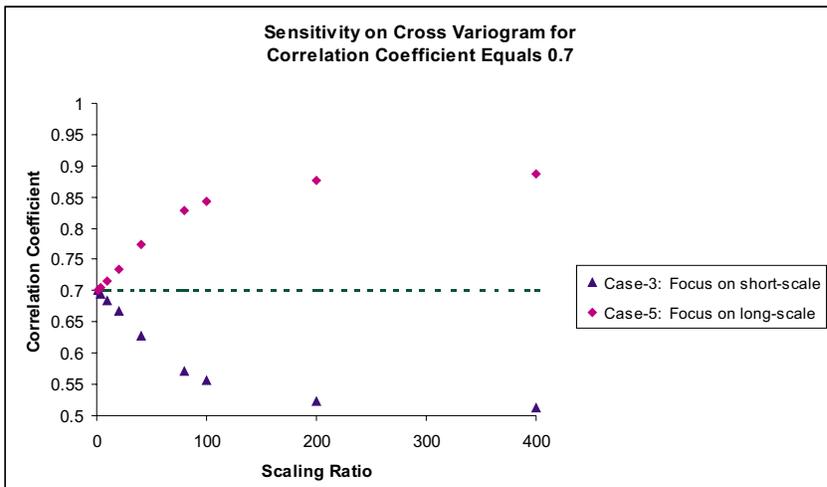


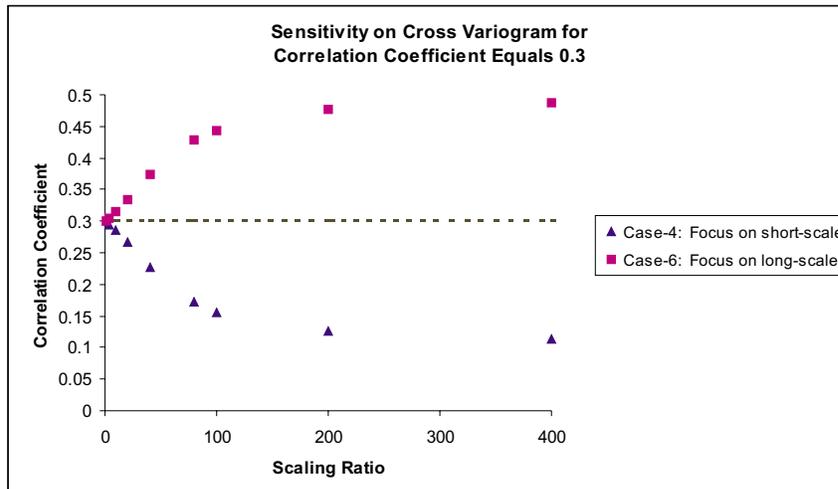
Figure 2: Fixed Direct variogram and different cross variograms for the cases of Cross Variogram Sensitivity



a) Sensitivity on Cross Variogram to illustrate the 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 the 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 the effect of focusing on long-scale and short-scale. Correlation coefficient fixed to 0.3 for cross variogram.

Figure 3: Sensitivity runs for cross variogram to illustrate the effects of Equal contribution, focusing on short-scale and long-scale

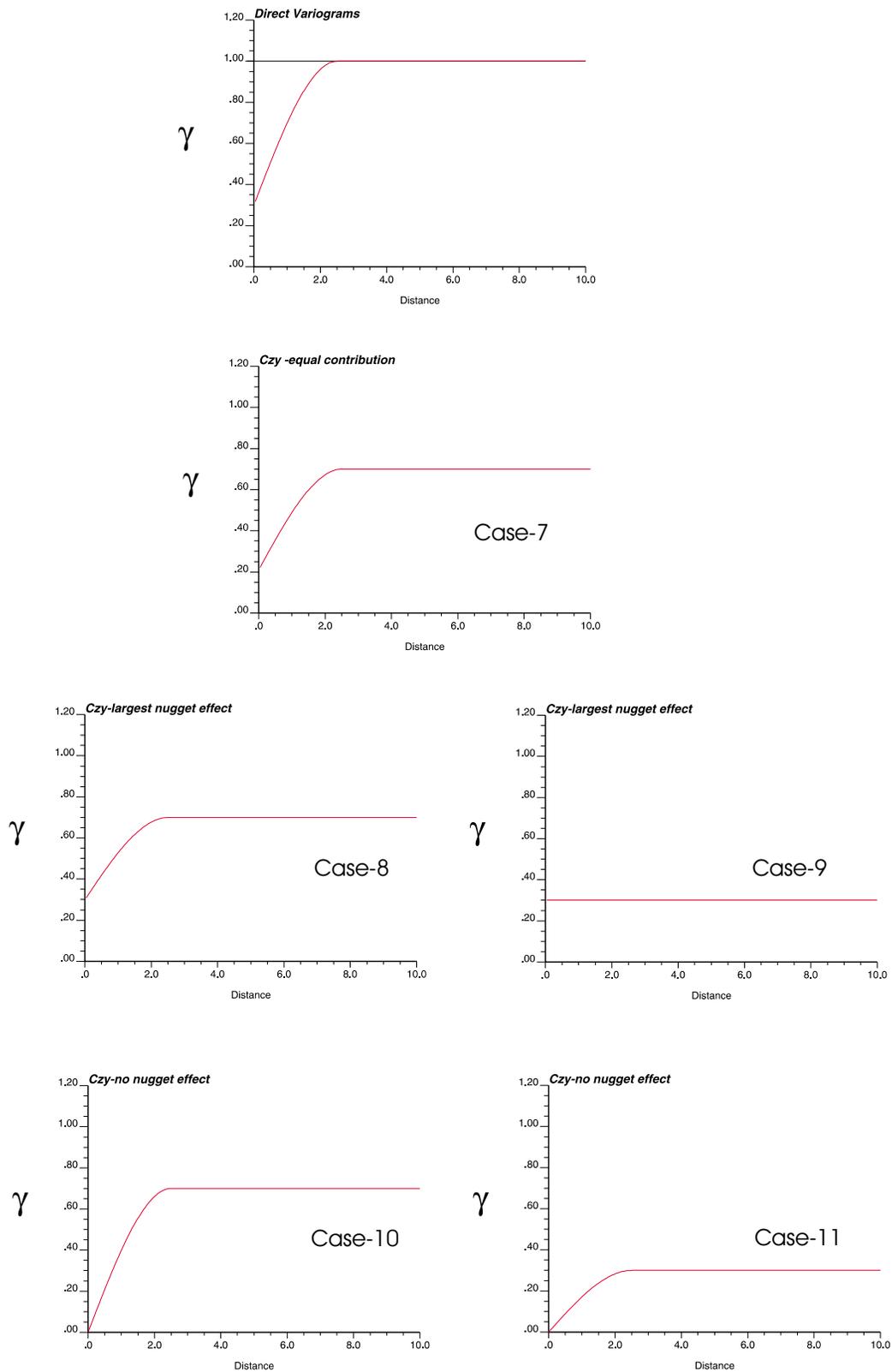
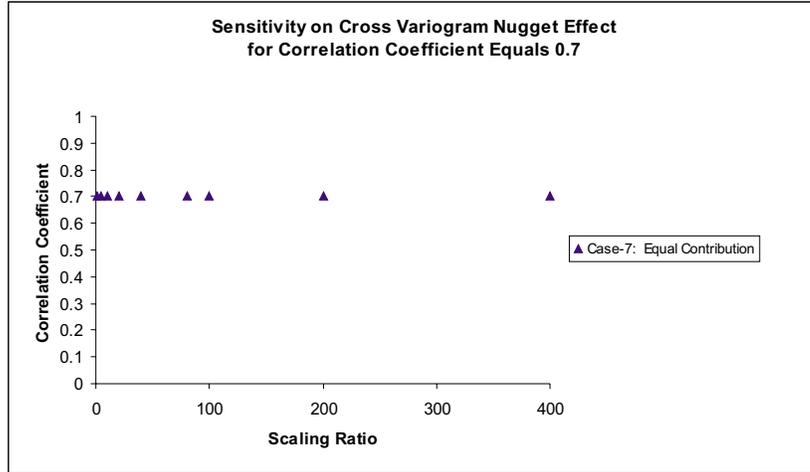
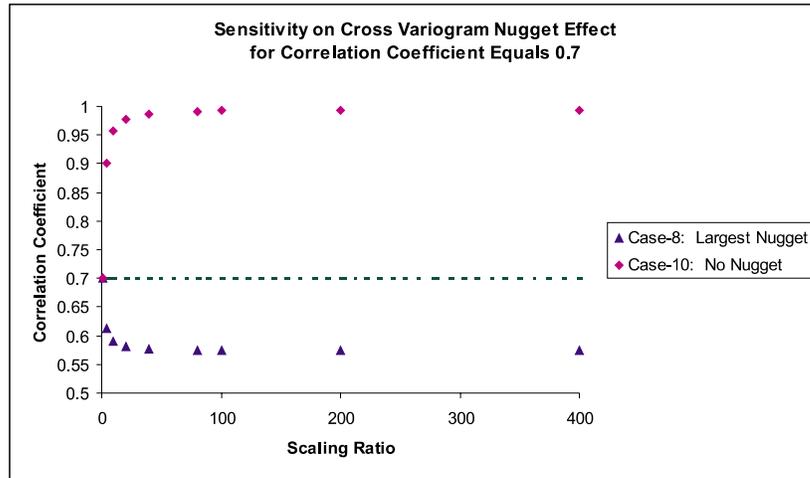


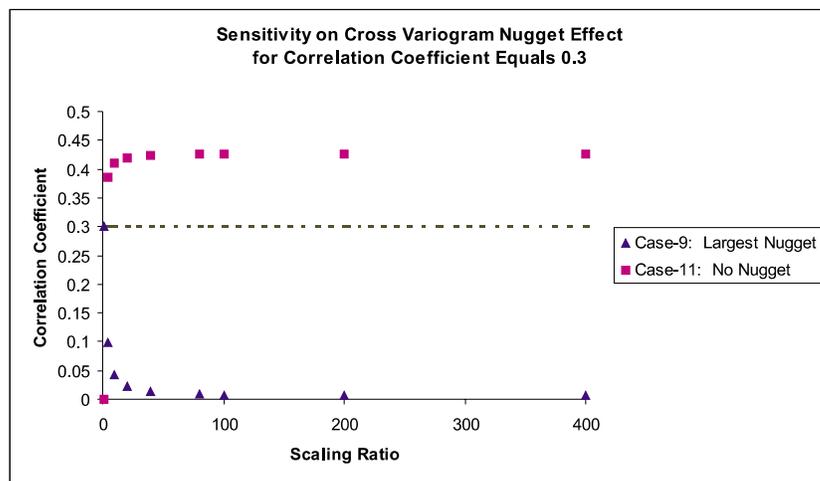
Figure 4: Fixed Direct variograms and different cross variograms for the cases of cross variogram nuggeteffect sensitivity



a) Sensitivity on Cross Variogram nugget effect to illustrate the effect of equal contribution of each structure in cross-variogram. Correlation coefficient fixed to 0.7 for cross variogram.



b) Sensitivity on cross variogram nugget effect to illustrate the effect of largest and no nugget effect. Correlation coefficient fixed to 0.7 for cross variogram.



c) Sensitivity on cross variogram nugget effect to illustrate the effect of largest and no nugget effect. Correlation coefficient fixed to 0.3 for cross variogram.

Figure 5: Sensitivity runs for cross variogram nugget effect to illustrate the effects of equal contribution, largest and no nugget

Sensitivity on the Asymmetry of Direct Variogram Structures:

In this sensitivity 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 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 directly proportional to the magnitude of the considered asymmetry ratio. Once more, the volume- dependent correlation coefficient reaches a stabilized value.

An Application to Real Data

A real field example is investigated to see if the theoretical results are validated by real data. 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 (top of Figure 8) was used. The “red” and “blue” color values of each pixel were considered, see the bottom two images of Figure 8. These two values are colocated and correlated. The histogram and the scatter plot of both red and blue data are given in Figures 9 and 10. It is interesting 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.

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 and the modelled direct and cross variograms are given in Figure 11. Two nested spherical models without nugget effect were used:

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

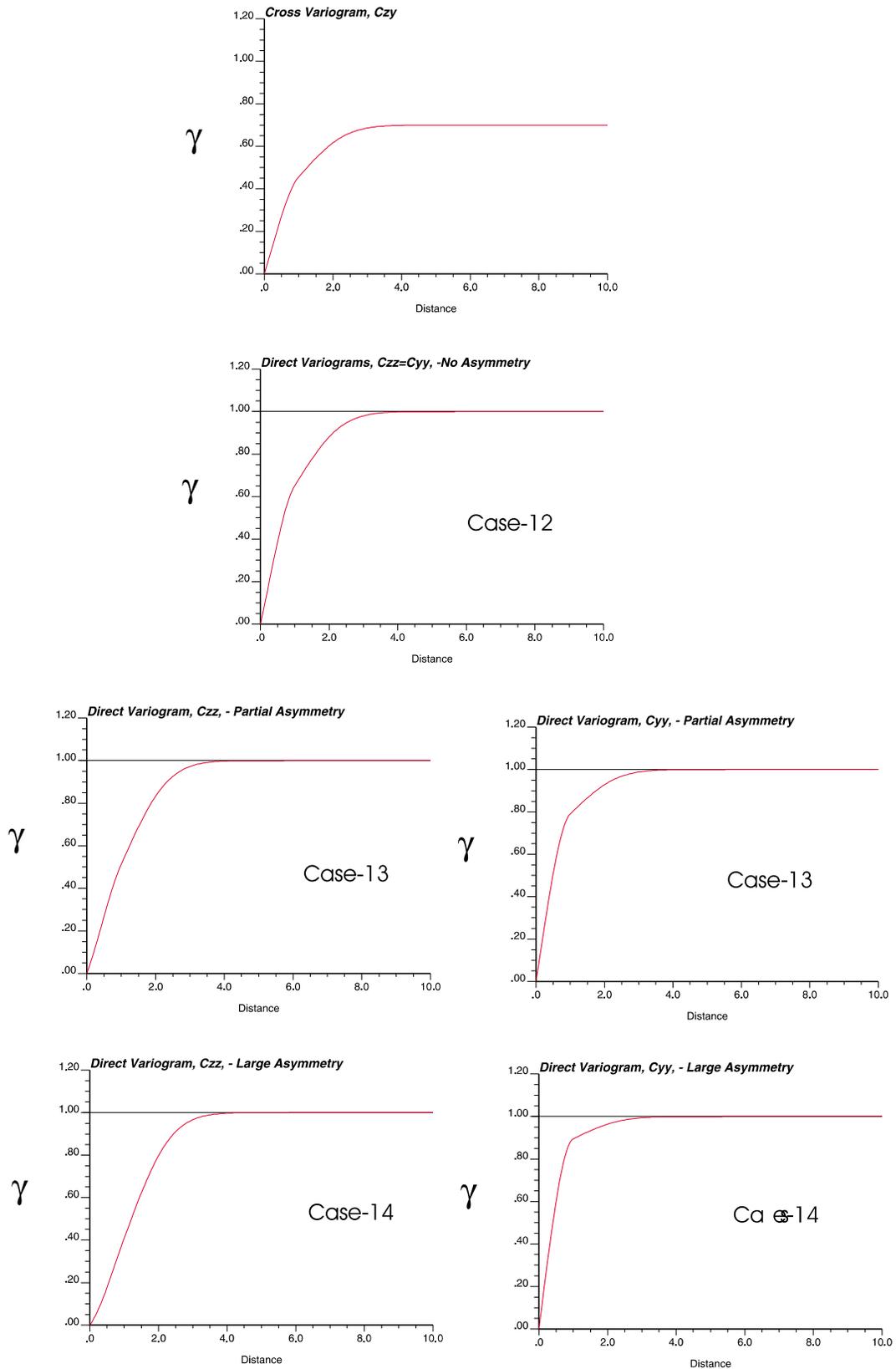
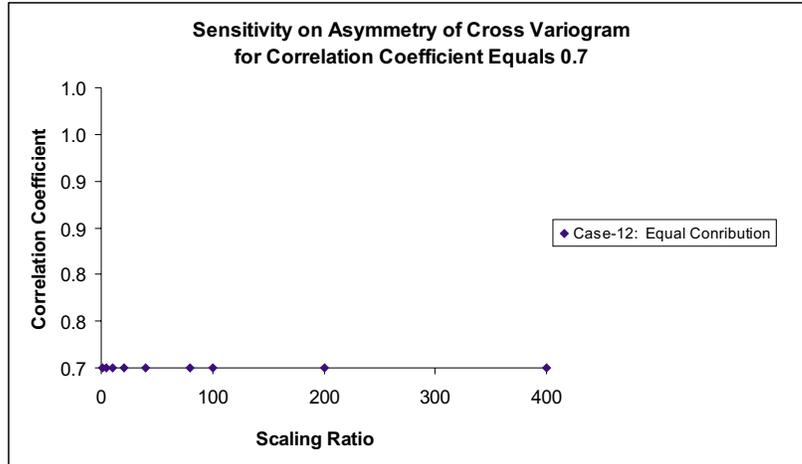
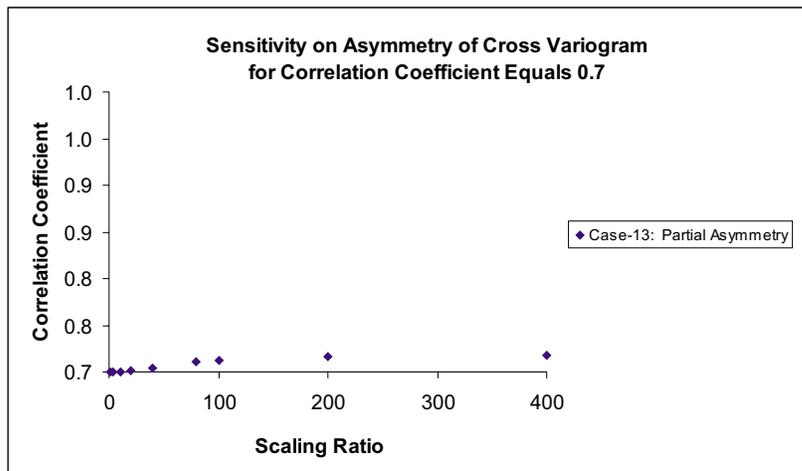


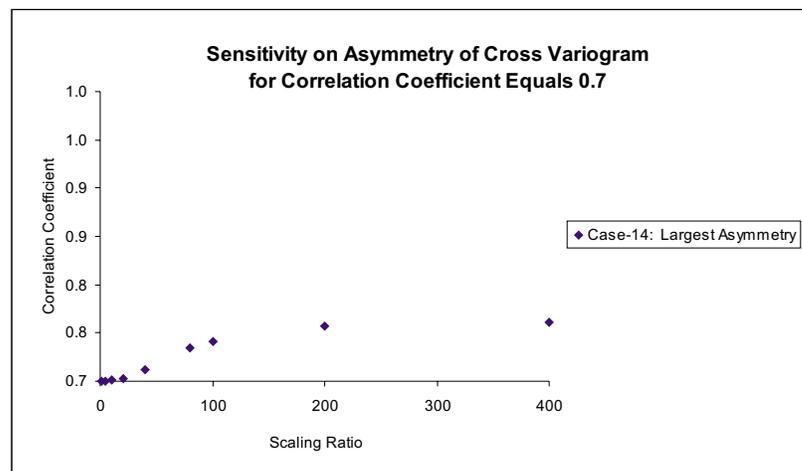
Figure 6: Fixed direct variogram and different cross variograms for the cases of cross variogram asymmetry sensitivity



a) Sensitivity on asymmetry of Cross Variogram to illustrate the effect of equal contribution of each structure in cross-variogram. Correlation coefficient fixed to 0.7 for cross variogram.



b) Sensitivity on asymmetry of cross variogram to illustrate the effect of partial asymmetry. Correlation coefficient fixed to 0.7 for cross variogram.



c) Sensitivity on asymmetry of cross variogram to illustrate the effect of largest asymmetry. Correlation coefficient fixed to 0.7 for cross variogram.

Figure 7: Sensitivity runs for cross variogram asymmetry to illustrate the effects of equal contribution, partial and large asymmetry

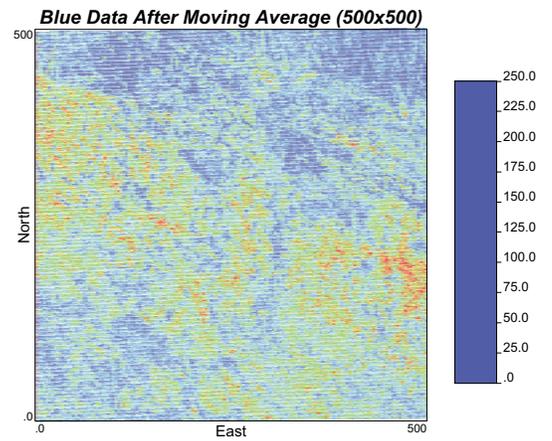
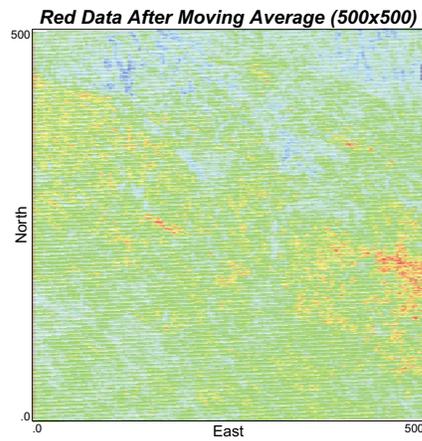


Figure 8: Top: An image from Wadi Kufra, Libya; Bottom: The the images representing the red and blue values of each pixel.

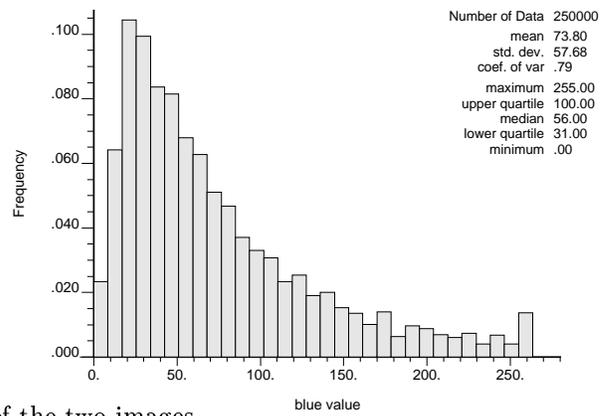
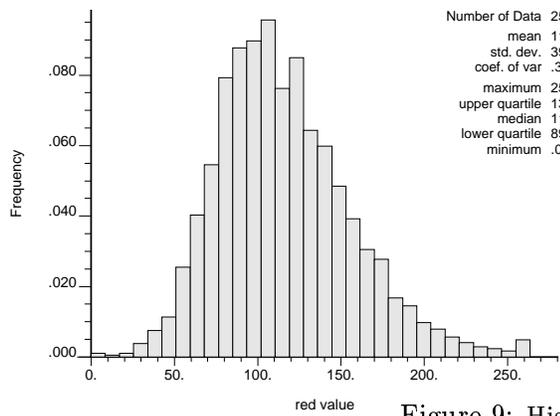


Figure 9: Histogram of the two images

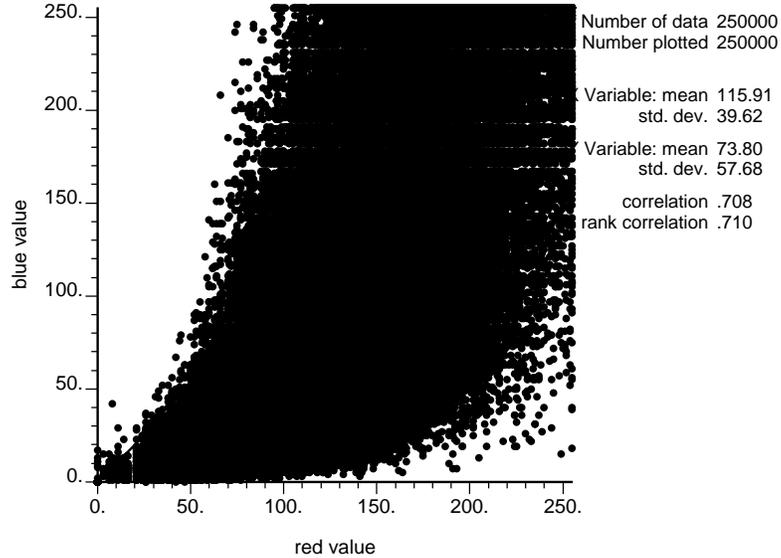


Figure 10: Scatter Plot for red and blue data

$$\gamma_{redblue}(h) = 695Sph_{(8,10)} + 900Sph_{(380,105)}$$

The sill values of the direct variograms are greater than zero, $770 \cdot 2300 > 695^2$ and $800 \cdot 1027 > 900^2$; therefore this LMC is positive definite.

A 2D linear upscaling 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 up-scaled set of images. The distribution of the scaled or “volume-dependent” cross- correlation coefficient is shown on Figure 12.

Using the LMC model in Equation 16, and the definition of volume- dependent cross- correlation coefficient in Equation 13, cross- correlations were calculated numerically from theory by using the `VarScale` program [19]. The comparison of the experimental and theoretical volume-dependent cross-correlation coefficients is presented in Figure 13. The correlation coefficient increases and approaches a steady-state value gradually after the averaging volume of 50. This increase is due to the sill contribution of the large-scale nested structure component of the cross variogram model (see earlier discussion).

The experimental and theoretical results are in general agreement; there is a particularly good match for larger blocks. The difference for the small blocks might be explained by the existence of spatial correlation at small scales that the variogram cannot capture. Notwithstanding this small mismatch, the general agreement between experimental and theoretical trends 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

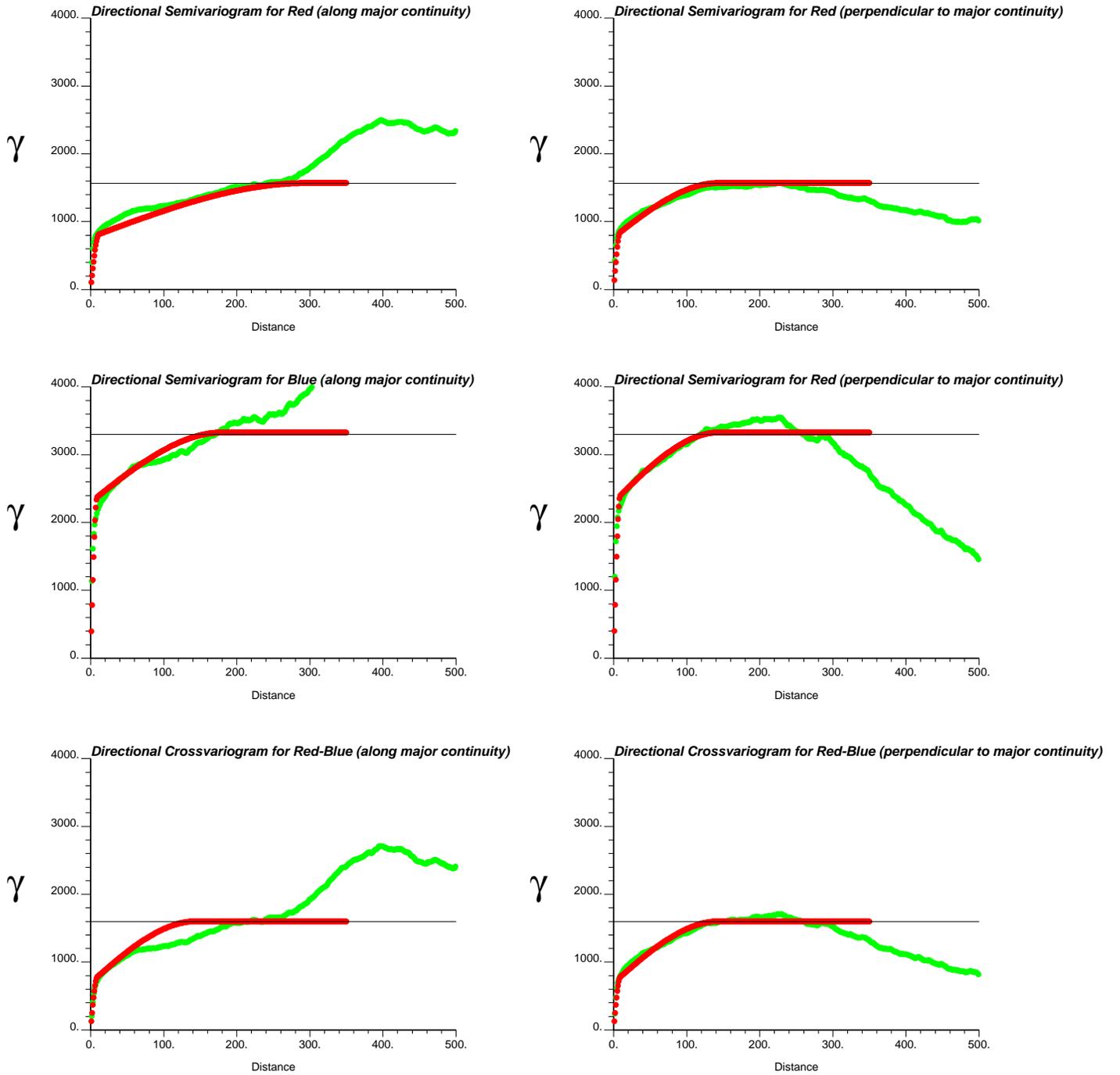


Figure 11: Directional Direct and Cros variograms for red and blue data along with the LMC model

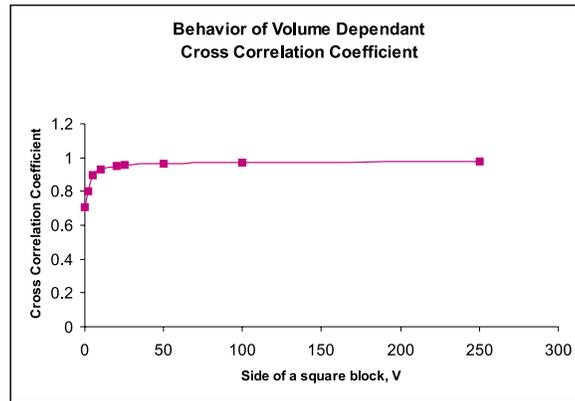


Figure 12: Behavior of cross correlation coefficient for different averaging volumes

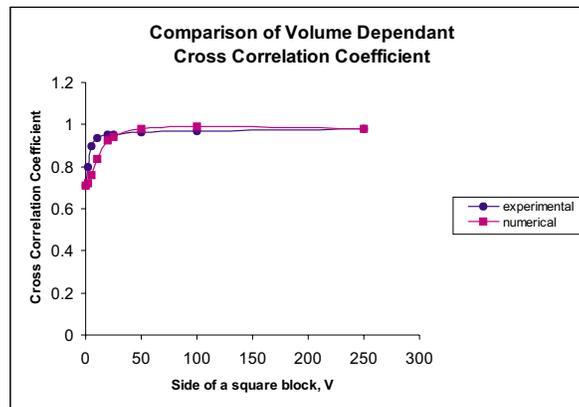


Figure 13: Comparison of Cross Correlation Coefficient

investigated more completely in this Section. Let's start with differentiating Equation 7,

$$\begin{aligned} \frac{\partial \rho(v)}{\partial v} = & \left(\frac{\partial \bar{C}_{zy}(v)}{\partial v} \right) (\bar{C}_{zz}^{-1/2}(v)) (\bar{C}_{yy}^{-1/2}(v)) - \frac{1}{2} \left(\frac{\partial \bar{C}_{zz}(v)}{\partial v} \right) (\bar{C}_{zz}^{-3/2}(v)) (\bar{C}_{zy}(v)) (\bar{C}_{yy}^{-1/2}(v)) \\ & - \frac{1}{2} \left(\frac{\partial \bar{C}_{yy}(v)}{\partial v} \right) (\bar{C}_{yy}^{-3/2}(v)) (\bar{C}_{zy}(v)) (\bar{C}_{zz}^{-1/2}(v)) \end{aligned} \quad (17)$$

The correlation coefficient at large scale can be written as:

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

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 Δv 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 the most critical determining the characteristic path of volume-dependent cross- correlation coefficient. Depending on its rate of change, the cross- correlation coefficient at larger scale may also 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 differentiation of dispersion variances is the inverse of computing auxiliary functions [22, 16] for the average variograms or cross-variograms. By defining a "growing window" concept, Vargas- Guzman and coworkers [22] linked the dispersion variances and covariances to the total variance function of $G(w)$, where w is the size of the growing window. Then, the derivative of $G(w)$ with respect to w gives the variogram itself. Alternatively, we can readily calculate Equation 19 numerically or simplify it by applying Leibniz's Theorem twice. The terms that appear in Equation 19 control the $\frac{\partial \rho(v)}{\partial v}$ (Equation 17) and $\rho(v_0 + \Delta v)$ term (Equation 18). They are highly non-linear 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's assume that we have same direct variograms, 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 [16]:

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

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

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

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 easily estimate the value of volume-dependent cross-correlation coefficient when $\mathbf{a} \geq l$. This kind 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 limitations, Equation 26 is a straightforward relationship depending on the average variogram and averaging length.

In order to test the efficiency of this approximation, we used the same LMC model as above (Equation 15) to estimate the correlation coefficients for different length scales. Our results show an error of around 3 to 5 percent for small averaging lengths. When we go to averaging lengths larger than the variogram range, we can use another form of auxiliary function [22] for the spherical variogram:

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

Then the derivative is given as:

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

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 highlight some critical terms. It is intractable to present a general analytical equation for complex coregionalization models. In general we resort to numerical techniques.

0.0.2 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:

$$\lim_{v \rightarrow \infty} \rho(v) \quad (29)$$

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

$$\lim_{v \rightarrow \infty} \frac{\int_0^v \int_0^v C_{zy}(x-x') dx dx'}{\sqrt{\int_0^v \int_0^v C_{zz}(x-x') dx dx'} \cdot \sqrt{\int_0^v \int_0^v C_{yy}(x-x') dx dx'}} \quad (30)$$

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

$$\lim_{v \rightarrow \infty} \rho(v) = \frac{\sum_{i=1}^{nst} C_{zy}^i a^i}{\sqrt{[\sum_{j=1}^{nst} \sum_{i=1}^{nst} C_{zz}^i C_{yy}^j a^i a^j]}} \quad (31)$$

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:

$$\lim_{v \rightarrow \infty} \rho(v) = \frac{C_{zy}^1 a^1 + C_{zy}^2 a^2}{\sqrt{[C_{zz}^1 a^1 + C_{zz}^2 a^2]^2}} \quad (32)$$

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

$$\lim_{v \rightarrow \infty} \rho(v) = \frac{0.5 * 1 + 0.2 * 5}{\sqrt{[0.5 * 1 + 0.5 * 5]^2}} = 0.5 \quad (33)$$

From the numerical calculations, this limit value is expected to be between 0.5 and 0.51, which is very 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 very 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 procedures 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 stablized-value. This plateau value is mainly controlled by the large-scale nested structure component of cross- variogram and direct variograms. Our study also shows that volume- dependent cross-correlation coefficient is very 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.
- There is a good match between the numerically calculated volume- dependent cross-correlations and ones obtained from a real field example (see Figure 13). This prompts us to seek for analytical relations 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. 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 cases.
- Since cross-correlation is the key element for data-integration techniques, the LMC model of coregionalization should be chosen 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

- [1] A. Almeida. Joint simulation of petrophysical properties using a Markov-type coregionalization model. In *Report 6*, Stanford, CA, May 1993. Stanford Center for Reservoir Forecasting.
- [2] R. A. Behrens, M. K. MacLeod, T. T. Tran, and A. O. Alimi. Incorporating seismic attribute maps in 3D reservoir models. In *1996 SPE Annual Technical Conference and Exhibition*, Denver, CO, October 1996. Society of Petroleum Engineers. SPE Paper Number 36499.
- [3] R. A. Behrens and T. T. Tran. Incorporating seismic data of intermediate vertical resolution into 3D models. In *SPE Annual Technical Conference and Conference*, New Orleans, LA, September 1998. SPE Paper Number 49143.
- [4] R. A. Behrens and T. T. Tran. Incorporating seismic data of intermediate vertical resolution into 3D reservoir. In *SPE Annual Technical Conference and Exhibition*, New Orleans, LA, September 1998. Society of Petroleum Engineers. SPE Paper Number 49143.

- [5] H. Beucher, F. Fournire, B. Doligez, and J. Rozanski. Using 3D seismic-derived information in lithofacies simulation: A case study. In *SPE Annual Technical Conference and Exhibition*, Houston, TX, October 1999. Society of Petroleum Engineers. SPE Paper Number 56736.
- [6] M. David. *Geostatistical Ore Reserve Estimation*. Elsevier, Amsterdam, 1977.
- [7] C. V. Deutsch and A. G. Journel. *GSLIB: Geostatistical Software Library and User's Guide*. Oxford University Press, New York, 2nd edition, 1997.
- [8] C. V. Deutsch, S. Srinivasan, and Y. Mo. Geostatistical reservoir modeling accounting for the scale and precision of seismic data. In *1996 SPE Annual Technical Conference and Exhibition*, pages 9–19, Denver, CO, October 1996. Society of Petroleum Engineers. SPE Paper Number 36497.
- [9] P. M. Doyen. Porosity from seismic data: A geostatistical approach. *Geophysics*, 53(10):1263–1275, 1988.
- [10] P. M. Doyen, L. D. den Boer, and W. R. Pillet. Incorporating seismic attribute maps in 3D reservoir models. In *1996 SPE Annual Technical Conference and Exhibition*, pages 21–30, Denver, CO, October 1996. Society of Petroleum Engineers. SPE Paper Number 36498.
- [11] P. M. Doyen, D. E. Psaila, L. D. den Boer, and D. Jans. Reconciling data at seismic and well log scales in 3d earth modelling. In *1997 SPE Annual Technical Conference and Exhibition*, pages 465–473, San Antonio, TX, October 1997. Society of Petroleum Engineers. SPE Paper Number 38698.
- [12] P. M. Doyen, D. E. Psaila, and S. Strandenes. Bayesian sequential indicator simulation of channel sands from 3D seismic data in the Oseberg field, Norwegian North Sea. In *69th Annual Technical Conference and Exhibition*, pages 197–211, New Orleans, LA, September 1994. Society of Petroleum Engineers. SPE Paper Number 28382.
- [13] P. Goovaerts. *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York, 1997.
- [14] A. G. Journel. Conditioning geostatistical operation to non-linear volume averages. *Mathematical Geology*, 31(8):931–953, 1999.
- [15] A. G. Journel. Markov models for cross covariances. *Mathematical Geology*, 31(8):955–963, 1999.
- [16] A. G. Journel and C. J. Huijbregts. *Mining Geostatistics*. Academic Press, New York, 1978.
- [17] A. G. Journel and H. Zhu. Integrating soft seismic data: Markov-Bayes updating, an alternative to cokriging and traditional regression. In *Report 3, Stanford Center for Reservoir Forecasting*, Stanford, CA, May 1990.

- [18] H. Kupfersberger, C. V. Deutsch, and A. Journel. Deriving constraints on small-scale variograms due to variograms of large-scale data. *Math. Geology*, 30(7):837–851, 1998.
- [19] B. Oz, C. V. Deutsch, and P. Frykman. A visual basic program for histogram and variogram scaling. *Computers & Geosciences*. accepted for publication in May 2000.
- [20] L. E. Shmaryan and A. G. Journel. Two markov models and their applications. *Mathematical Geology*, 31(8):965–988, 1999.
- [21] T. T. Tran, X.-H. Wen, and R. A. Behrens. Efficient conditioning of 3D fine-scale reservoir model to multiphase production data using streamline-based coarse-scale inversion and geostatistical downscaling. In *SPE Annual Technical Conference and Conference*, Houston, TX, October 1999. SPE Paper Number 56518.
- [22] J. A. Vargas-Guzman, D. E. Myers, and A. W. Warrick. Derivatives of spatial variances of growing windows and the variogram. *Mathematical Geology*, 32(7):851–871, 2000.
- [23] J. A. Vargas-Guzman, A. W. Warrick, and D. E. Myers. Multivariate correlation in the framework of support and spatial scales of variability. *Mathematical Geology*, 31(1):85–103, 1999.
- [24] J. A. Vargas-Guzman, A. W. Warrick, and D. E. Myers. Scale effect on principal component analysis for the vector random functions. *Mathematical Geology*, 31(6):701–722, 1999.
- [25] W. Xu, T. T. Tran, R. M. Srivastava, and A. G. Journel. Integrating seismic data in reservoir modeling: the collocated cokriging alternative. In *67th Annual Technical Conference and Exhibition*, pages 833–842, Washington, DC, October 1992. Society of Petroleum Engineers. SPE Paper Number 24742.