Direct Upscaling of Variograms and Cross Variograms for Scale Consistent Geomodeling

Olena Babak and Oy Leuangthong

Integration of data from multiple sources and/or multiple scales is a common, yet challenging aspect of geostatistical modeling. Common approaches to data integration are based on a cokriging framework, that often assumes the input variogram/covariance models of coregionalization are at a scale consistent with the data and the model grid. The scaling laws for the variogram have often been applied to ensure consistency of the input variogram model; however, these laws are based on a strict assumption of invariance of the variogram shape. We propose a direct upscaling approach to the variogram that is theoretically derived. A numerical integration approach to determine the upscaled variogram is presented with an example showing the difference between this theoretical proxy approach and the scaling laws approach. The results show that the shape of the variogram does indeed change with scale. Further, the extension to cross variograms is straightforward and an upscaled consistent linear model of coregionalization is presented.

Introduction

Data from well cores are among the finest scale of information available for geostatistical inference, and as a result it is often considered to be point scale data with an infinitesimally small volume support. Data from log traces represent an incremental increase in volume support relative to core data; however, these are also considered to be relatively fine scale information. Seismic surveys, if available, generally cover much larger lateral extents with shallow depth. These different data at different support sizes must then be reconciled to some intermediate modeling volume that is determined based on the resolution required and the computational resources available. Integration of these data from various sources and at different volume supports is a longstanding challenge in geostatistical modeling (Kupfersberger et al 1998).

In most cases, the smallest size that the fine scale model can be is limited by computer storage and professional time. Consequently, even the finest scale model is larger than the data support. In the context of estimation, several geostatistical tools exist to facilitate data integration. One set of tools considers that secondary data inform a trend about the primary data; these techniques include external drift and locally varying mean (Marechal 1984; Deutsch and Journel 1998). Another group of methods uses the secondary data as additional conditioning information for estimation of the primary variable; these methods include collocated cokriging (Xu et al 1992; Almeida and Journel 1994), Bayesian updating (Doyen 1996; Ren et al 2007) and block cokriging (Goovaerts 1997). With the exception of block cokriging, most methods assume the primary and all other secondary information are at a consistent scale.

Consistency of scale in the input data and the intended grid can easily be an oversight. In the context of geostatistical simulation, the model is constructed based on a point-support simulation, the input data is considered at a point support and the input variogram corresponds to this same data. This is an acceptably consistent scale model at the point support (Journel and Kyriakidis 2004). However, the model is commonly required at an intermediate block scale and high resolution fine scale models are practically infeasible; populating block simulated values is desired.

Direct sequential simulation (DSS) is one possible approach to simulate at a block scale; however, issues with histogram inference and reproduction (Bourgault 1997; Caers 2000; Oz et al 2003) have limited many applications to univariate modeling. Soares (2001) proposed a direct cosimulation approach to handle multivariate problems based on a collocated cokriging approach. If the correlation coefficient between primary and secondary data are properly scaled to account for any scale differentials between the data, then this approach can and will account for different supports.

An alternative to DSS is to consider a ‘point’ simulation at a block-consistent grid specification, using block averaged data and the block-scale variogram. Once again, the premise for this approach is to ensure consistency in the input data and the required model. Data compositing is a common practice to upscale

available hard data. Using this composited data, we can calculate the corresponding composite-support variogram; however, compositing reduces the number of available samples for reliable variogram inference. In such cases, modelers may calculate and fit the variogram using the original support data. This variogram model must then be upscaled for consistency with the model grid and composited data.

Scaling of the variogram model to reflect different volume supports is not new. The scaling laws associated with the variogram are well known (Journel and Huijbregts 1978; Frykman and Deutsch 1999, 2002). They predict how the variogram changes from one volume to another, with specific laws linked to the nugget effect, range and the variance contributions of each constituent structure. A number of simplifying assumptions are made to determine these scaling laws, including volume averaging is performed for non-overlapping volumes, the underlying variable averages linearly, and the shape of the variogram does not change with scale. In particular, the latter assumption of shape invariance of the variogram is a strong and often unrealistic assumption.

This paper proposes to directly upscale the variogram through a numerical integration approach. This numerical approach is used as a proxy to a theoretically developed expression to analytically upscale the variogram. A review of the variogram scaling laws is provided. The proposed approach to explicitly upscale the variogram model is then described. To facilitate data integration, the extension to cross variograms is straightforward; this is presented in a section that considers the linear model of coregionalization at the required block support. A small example is then used to compare the results of the scaling laws to the direct upscaling approach.

Regularization

Very seldom, in practice, point data $z(\mathbf{u})$ is available. Most often, the data at hand is defined on a certain support $V = V(\mathbf{u})$ centered on a point $\mathbf{u}$, that is, $z^V(\mathbf{u})$. The value of $z^V(\mathbf{u})$ is the average of the point data $z(\mathbf{u})$ in the volume V, that is,

$$z^V(\mathbf{u}) = \frac{1}{V}\int_V z(\mathbf{w})d\mathbf{w}. \tag{1}$$

The value $z^V(\mathbf{u})$ is called the regularization of the point variable $z(\mathbf{w})$ over the volume V (Journel and Huijbregts 1978). If the point-regionalized variable $z(\mathbf{w})$ is a realization of a second-order stationary random function $Z(\mathbf{w})$, then the regularization of the point random function $Z(\mathbf{w})$ over the volume V is also a second-order stationary random function given by:

$$Z^V(\mathbf{u}) = \frac{1}{V}\int_V Z(\mathbf{w})d\mathbf{w}. \tag{2}$$

Variogram Scaling Laws

The variogram model is linked to the volume support of the data. To represent the volume support that we are interested in, the variogram models need necessarily be scaled. To represent the change in the variogram with the change in the volumetric scale the variogram scaling laws are commonly applied (Frykman and Deutsch 1999, 2002). A short description of these laws follows.

Let us consider a semivariogram model $\gamma_v(\mathbf{h})$ at arbitrary scale v (v usually the small core scale):

$$\gamma_v(\mathbf{h}) = C_v^0 + \sum_{i=1}^{k} C_v^i \Gamma_v^i(\mathbf{h}), \tag{3}$$

where C_v^0 is the nugget effect, k is the number of nested variogram structures used to fit the experimental variogram of the data, $C_v^i,\ i = 1,\ldots,k,$ is the variance contribution of each nested structure, and $\Gamma_v^i(\mathbf{h}),\ i = 1,\ldots,k,$ are individual nested variogram structures with sill of one. Each nested structure is given by analytical function (e.g., spherical, exponential, etc.).

Then a semivariogram model $\gamma_V(\mathbf{h})$ at a larger volume *V* is given by

$$\gamma_V(\mathbf{h}) = C_V^0 + \sum_{i=1}^{k} C_V^i \Gamma_V^i(\mathbf{h}), \tag{4}$$

where C_V^0 is the nugget effect, $C_V^i, \quad i = 1, \ldots, k,$ is the variance contribution of each nested structure, and $\Gamma_V^i(\mathbf{h}), \quad i = 1, \ldots, k,$ are individual nested variogram structures (all for the scale *V*). And the following is assumed:

1. The shape of each individual nested structure (i.e., exponential, spherical) remains invariant when the scale changes. The variogram range of each rested structure increases as the size of the volume *V* increases. In particular, if a_v^i is the range of $\Gamma_v^i(\mathbf{h}), \quad i = 1, \ldots, k,$ for the small scale, then the range a_V^i of $\Gamma_V^i(\mathbf{h}), \quad i = 1, \ldots, k,$ for the large scale is given by

$$a_V^i = a_v^i + (|V| - |v|), \tag{5}$$

where $|v|$ and $|V|$ relate to the size of the volume in a particular direction.

2. The variance contributions $C_v^i,$ of each nested structure $\Gamma_V^i(\mathbf{h}), \quad i = 1, \ldots, k,$ change from the small scale to the large scale as follows

$$C_V^i = C_v^i \frac{1 - \overline{\Gamma}^i(V,V)}{1 - \overline{\Gamma}^i(v,v)}, \tag{6}$$

where $\overline{\Gamma}^i(V,V)$ and $\overline{\Gamma}^i(v,v)$ are the average variogram or "gamma-bar" values. In particular, $\overline{\Gamma}^i(V,V)$ represents the mean variogram $\Gamma^i(\mathbf{h})$ when one extremity of the vector **h** describes the domain *V* and the other extremity of this vector independently describes the same domain. The values of gamma-bar's are usually estimated numerically by volume discretization, that is,

$$\overline{\Gamma}(V,V) = \frac{1}{V}\frac{1}{V}\int_V \int_V \Gamma(\mathbf{y} - \mathbf{y}')d\mathbf{y}d\mathbf{y}' \approx \frac{1}{n}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\Gamma(\mathbf{u}_i - \mathbf{u}_j), \tag{7}$$

where n is the number of regular spaced points discretizing the volume *V*.

3. The variance of the purely random component, called nugget effect, is inversely related to the volume, that is,

$$C_V^0 = C_v^0 \frac{|v|}{|V|}. \tag{8}$$

It should be noted that as scale increases, the range of correlation increases, and the variogram sill decreases due to high and low values being averaged out. Moreover, it is worth noting also that the scaling laws are established under the following additional assumptions (Journel and Huijbregts 1978):

1. The averaging is performed with non-overlapping volumes.
2. The variables scale in a linear manner.

In general, the assumptions underlying scaling laws appear to be very strong and limiting. In particular, the assumption of no shape change for the variogram nested structures is very unrealistic. It has been observed in many examples that the shape of the experimental variograms at volume *V* is different from the ones predicted from scaling laws especially for short lag distances (e.g., Frykman and Deutsch 2002). Therefore, a direct approach for exact calculation of the upscaled variograms is of great practical interest.

Direct Variogram Upscaling

If $Z(\mathbf{w})$ is assumed to be a second-order stationary random function with mean *m*, covariance $C(\mathbf{h})$ and variogram $2\gamma(\mathbf{h})$. Then $Z^V(\mathbf{u})$ given by

$$Z^V(\mathbf{u}) = \frac{1}{V}\int_V Z(\mathbf{w})d\mathbf{w}$$

is a second-order stationary random function representing the upscaled point random function $Z(\mathbf{w})$ or the regularization of the point random function $Z(\mathbf{w})$ over the volume V.

Then, the mean $E(Z^V(\mathbf{u}))$, variance $Var(Z^V(\mathbf{u}))$ semivariogram $\gamma^V(\mathbf{h})$ and covariance $C^V(\mathbf{h})$ of $Z^V(\mathbf{u})$ for the scale *V* can be calculated based on the mean *m*, covariance $C(\mathbf{h})$ and semivariogram $\gamma(\mathbf{h})$ of $Z(u)$ for the point scale as follows (Journel and Huijbregts, 1978):

$$\begin{aligned} &E(Z^V(\mathbf{u})) = m; \\ &Var(Z^V(\mathbf{u})) = C^V(\mathbf{0}) = \overline{C}(V,V); \\ &\gamma^V(\mathbf{h}) = \bar{\gamma}(V,V_{\mathbf{h}}) - \bar{\gamma}(V,V); \\ &C^V(\mathbf{h}) = C^V(\mathbf{0}) - \gamma^V(\mathbf{h}) = \overline{C}(V,V_{\mathbf{h}}); \end{aligned} \tag{9}$$

where $V_{\mathbf{h}}$ denote the support *V* translated from *V* by the vector h; $\bar{\gamma}(V,V_{\mathbf{h}})$ represents the average of the point semivariogram $\gamma(\mathbf{h})$ when one extremity of the vector h describes the support *V* and the other extremity independently describes the translated support $V_{\mathbf{h}}$, that is,

$$\bar{\gamma}(V,V_{\mathbf{h}}) = \frac{1}{V}\frac{1}{V_{\mathbf{h}}}\int_V\int_{V_{\mathbf{h}}}\gamma(\mathbf{w}-\mathbf{x})d\mathbf{w}\,d\mathbf{x} = \frac{1}{V}\frac{1}{V}\int_V\int_V\gamma(\mathbf{w}-(\mathbf{x}+\mathbf{h}))d\mathbf{w}\,d\mathbf{x}; \tag{10}$$

$\overline{C}(V,V_{\mathbf{h}})$ represents the average of covariance $C(\mathbf{h})$ and is given by

$$\overline{C}(V,V_{\mathbf{h}}) = \frac{1}{V}\frac{1}{V_{\mathbf{h}}}\int_V\int_{V_{\mathbf{h}}}C(\mathbf{w}-\mathbf{x})d\mathbf{w}\,d\mathbf{x} = \frac{1}{V}\frac{1}{V}\int_V\int_V C(\mathbf{w}-(\mathbf{x}+\mathbf{h}))d\mathbf{w}\,d\mathbf{x}. \tag{11}$$

Note that if the semivariogram $\gamma(\mathbf{h})$ of the point-regularized random variable is made up of several nested structures, that is,

$$\gamma(\mathbf{h}) = C^0 + \sum_{i=1}^{k} C^i\gamma^i(\mathbf{h}), \tag{12}$$

where C^0 is the nugget effect, *k* is the number of nested variogram structures, C^i, $i=1,\ldots,k,$ is the variance contribution of each nested structure, and $\gamma^i(\mathbf{h})$, $i=1,\ldots,k,$ are individual nested variogram structures given by analytical function (e.g., spherical, exponential, etc.) with a sill of one. Then, the average of the point semivariogram $\gamma(\mathbf{h})$ can be calculated as

$$\begin{aligned} \bar{\gamma}(V,V_{\mathbf{h}}) &= \frac{1}{V}\frac{1}{V_{\mathbf{h}}}\int_V\int_{V_{\mathbf{h}}}\left[C^0 + \sum_{i=1}^{k} C^i\gamma^i((\mathbf{w}-\mathbf{x}))\right]d\mathbf{w}\,d\mathbf{x} \\ &= C^0 + \sum_{i=1}^{k} C^i\left[\frac{1}{V}\frac{1}{V}\int_V\int_V\gamma(\mathbf{w}-(\mathbf{x}+\mathbf{h}))d\mathbf{w}\,d\mathbf{x}\right] = C^0 + \sum_{i=1}^{k} C^i\bar{\gamma}^i(V,V_{\mathbf{h}}). \end{aligned} \tag{13}$$

And, thus, the semivariogram $\gamma^V(\mathbf{h})$ for the scale *V* is given by

$$\begin{aligned} \gamma^V(\mathbf{h}) = \bar{\gamma}(V,V_{\mathbf{h}}) - \bar{\gamma}(V,V) &= C^0 + \sum_{i=1}^{k} C^i\bar{\gamma}^i(V,V_{\mathbf{h}}) - \left[C^0 + \sum_{i=1}^{k} C^i\bar{\gamma}^i(V,V)\right] \\ &= \sum_{i=1}^{k} C^i\left[\bar{\gamma}^i(V,V_{\mathbf{h}}) - \bar{\gamma}^i(V,V)\right] \end{aligned} \tag{14}$$

It is worth noting that semivariogram $\gamma^V(\mathbf{h})$ does not contain any nugget effect. The nugget effect vanishes when upscaling to a larger volume *V*, it does matter how large the nugget effect was in the point scale variogram. Similarly, it can be shown that in this case the covariance for the scale *V* is given by

$$C^V(\mathbf{h}) = \overline{C}(V,V_{\mathbf{h}}) = C(0) - \left[C^0 + \sum_{i=1}^{k} C^i\bar{\gamma}^i(V,V_{\mathbf{h}})\right]. \tag{15}$$

Since the variance $C(0)$ at the point scale for semivariogram $\gamma(\mathbf{h})$ is given by

$$C(0) = C^0 + \sum_{i=1}^{k} C^i, \tag{16}$$

then it follows from Equation (15) that covariance for the scale *V* can be calculated as

$$C^V(\mathbf{h}) = \sum_{i=1}^{k} C^i[1 - \bar{\gamma}^i(V, V_{\mathbf{h}})]. \tag{17}$$

Linear Model of Coregionalization (LMC) at a Block Support

In the case when multiple interdependent random variables are available, the spatial relationship between them must be described in a feasible manner. Let us now derive a linear model of coregionalization for *N* stationary regularizations $\{Z_1^V, \ldots, Z_N^V\}$ of the point random functions $\{Z_1, \ldots, Z_N\}$ over the volume V based on a linear model of coregionalization for $\{Z_1, \ldots, Z_N\}$ at a point support.

Let us consider *N* stationary point random functions $\{Z_1, \ldots, Z_N\}$. Further let us assume that each point support random function $Z_i, \quad i = 1, \ldots, N,$ can be expressed as a linear combination of *K* independent zero mean second-order stationary random functions $Y_k, \quad k = 1, \ldots, K,$ each with covariance function $C_k(\mathbf{h})$ as follows

$$Z_i(\mathbf{u}) = \sum_{k=1}^{K} a_{ik} Y_k(\mathbf{u}) + \mu_i. \tag{18}$$

The random functions $Y_k, \quad k = 1, \ldots, K,$ are assumed to be unknown. If we group the random functions $Y_k, \quad k = 1, \ldots, K,$ according to distinct direct covariances $C_k(\mathbf{h})$, then Equation (18) can be rewritten:

$$Z_i(\mathbf{u}) = \sum_{l=1}^{L} \sum_{p=1}^{n_l} a_{ip}^l Y_p^l(\mathbf{u}) + \mu_i \tag{19}$$

with

$$C(Y_p^l(\mathbf{u}), Y_{p'}^{l'}(\mathbf{u}+\mathbf{h})) = \begin{cases} C^l(\mathbf{h}), & \text{if} \quad p = p' \text{ and } l = l'; \\ 0, & \text{otherwise;} \end{cases} \quad .$$

where $L+1$ is the number of groups with distinct direct covariances; and n_l is the number of random functions with the same covariance $C^l(\mathbf{h}), \ l = 0, \ldots, L.$

Then direct and cross covariances between two random variables $Z_i(\mathbf{u})$ and $Z_j(\mathbf{u}+\mathbf{h})$, $i, j = 1, \ldots, N,$ can be calculated as

$$\begin{aligned} C_{ij}(\mathbf{h}) &= Cov(Z_i(\mathbf{u}), Z_j(\mathbf{u}+\mathbf{h})) = Cov\left(\sum_{l=1}^{L} \sum_{p=1}^{n_l} a_{ip}^l Y_p^l(\mathbf{u}) + \mu_i, \sum_{l'=1}^{L} \sum_{p'=1}^{n_{l'}} a_{jp'}^{l'} Y_{p'}^{l'}(\mathbf{u}+\mathbf{h}) + \mu_j \right) \\ &= Cov\left(\sum_{l=1}^{L} \sum_{p=1}^{n_l} a_{ip}^l Y_p^l(\mathbf{u}), \sum_{l'=1}^{L} \sum_{p'=1}^{n_{l'}} a_{jp'}^{l'} Y_{p'}^{l'}(\mathbf{u}+\mathbf{h}) \right) = \sum_{l=1}^{L} \sum_{p=1}^{n_l} \sum_{l'=1}^{L} \sum_{p'=1}^{n_{l'}} a_{ip}^l a_{jp'}^{l'} Cov(Y_p^l(\mathbf{u}), Y_{p'}^{l'}(\mathbf{u}+\mathbf{h})) \\ &= \sum_{l=1}^{L} \sum_{p=1}^{n_l} a_{ip}^l a_{jp}^l C^l(\mathbf{h}). \end{aligned} \tag{20}$$

If we set

$$b_{ij}^l = \sum_{p=1}^{n_l} a_{ip}^l a_{jp}^l, \quad i = 1, \ldots, N, \quad j = 1, \ldots, N, \tag{21}$$

then it follows from Equation (20):

$$C_{ij}(\mathbf{h}) = \sum_{l=1}^{L} b_{ij}^{l} C^{l}(\mathbf{h}), \quad i = 1,\ldots,N, \quad j = 1,\ldots,N. \tag{22}$$

Therefore to determine the direct and cross covariances between any two random variables $Z_i(\mathbf{u})$ and $Z_j(\mathbf{u}+\mathbf{h})$, $i, j = 1,\ldots,N,$ we need only to determine covariances $C^{l}(\mathbf{h})$, $l = 0,\ldots,L,$ and the $(L+1)P^2$ coefficients b_{ij}^{l}. Because a joint matrix of covariance functions $C_{ij}(\mathbf{h})$, $i, j = 1,\ldots,N,$ must be positive semi-definite; this requires the covariance models $C^{l}(\mathbf{h})$, $l = 0,\ldots,L,$ and *L*+1 matrices of b_{ij}^{l} coefficients to be positive semi-definite. In practice the covariance models $C^{l}(\mathbf{h})$, $l = 0,\ldots,L,$ are chosen to be known positive semi-definite models such as Spherical, Exponential, Gaussian, etc. Model (22) is the liner model of coregionalization at a point support. Moreover, since

$$Z_i^{V}(\mathbf{u}) = \frac{1}{V}\int_{V} Z_i(\mathbf{w})d\mathbf{w}, \tag{23}$$

it follows from Equation (19)

$$\begin{aligned} Z_i^{V}(\mathbf{u}) &= \frac{1}{V}\int_{V}\left[\sum_{l=1}^{L}\sum_{p=1}^{n_l} a_i^{l} Y_k^{l}(\mathbf{w}) + \mu_i\right]d\mathbf{w} = \frac{1}{V}\int_{V}\sum_{l=1}^{L}\sum_{p=1}^{n_l} a_{ik}^{l} Y_k^{l}(\mathbf{w})d\mathbf{w} + \frac{1}{V}\int_{V}\mu_i d\mathbf{w} \\ &= \sum_{l=1}^{L}\sum_{p=1}^{n_l} a_{ik}^{l}\left[\frac{1}{V}\int_{V} Y_k^{l}(\mathbf{w})\right]d\mathbf{w} + \mu_i. \end{aligned} \tag{24}$$

Therefore the direct and cross covariances between two regularization random variables $Z_i^{V}(\mathbf{u})$ and $Z_j^{V}(\mathbf{u}+\mathbf{h})$, $i, j = 1,\ldots,N,$ are given by

$$\begin{aligned} C_{ij}^{V}(\mathbf{h}) &= Cov(Z_i^{V}(\mathbf{u}), Z_j^{V}(\mathbf{u}+\mathbf{h})) \\ &= Cov\left(\sum_{l=1}^{L}\sum_{p=1}^{n_l} a_{ip}^{l}\left[\frac{1}{V}\int_{V} Y_p^{l}(\mathbf{w})d\mathbf{w}\right] + \mu_i, \sum_{l'=1}^{L}\sum_{p'=1}^{n_{l'}} a_{jp'}^{l}\left[\frac{1}{V}\int_{V} Y_{p'}^{l'}(\mathbf{y}+\mathbf{h})d\mathbf{y}\right] + \mu_j\right) \\ &= Cov\left(\sum_{l=1}^{L}\sum_{p=1}^{n_l} a_{ip}^{l}\left[\frac{1}{V}\int_{V} Y_p^{l}(\mathbf{w})d\mathbf{w}\right], \sum_{l'=1}^{L}\sum_{p'=1}^{n_{l'}} a_{jp'}^{l}\left[\frac{1}{V}\int_{V} Y_{p'}^{l'}(\mathbf{y}+\mathbf{h})d\mathbf{y}\right]\right) \\ &= \sum_{l=1}^{L}\sum_{p=1}^{n_l}\sum_{l'=1}^{L}\sum_{p'=1}^{n_{l'}} a_{ip}^{l} a_{jp'}^{l'} Cov\left(\frac{1}{V}\int_{V} Y_p^{l}(\mathbf{w})d\mathbf{w}, \frac{1}{V}\int_{V} Y_{p'}^{l'}(\mathbf{y}+\mathbf{h})d\mathbf{y}\right) \\ &= \sum_{l=1}^{L}\sum_{p=1}^{n_l}\sum_{l'=1}^{L}\sum_{p'=1}^{n_{l'}} a_{ip}^{l} a_{jp'}^{l'} \frac{1}{V}\frac{1}{V}\int_{V}\int_{V} Cov(Y_p^{l}(\mathbf{w}), Y_{p'}^{l'}(\mathbf{y}+\mathbf{h}))d\mathbf{w}d\mathbf{y} \\ &= \sum_{l=1}^{L}\sum_{p=1}^{n_l} a_{ip}^{l} a_{jp}^{l} \frac{1}{V}\frac{1}{V}\int_{V}\int_{V} C^{l}(\mathbf{w}-(\mathbf{y}+\mathbf{h}))d\mathbf{w}d\mathbf{y} \\ &= \sum_{l=1}^{L}\sum_{p=1}^{n_l} a_{ip}^{l} a_{jp}^{l} \overline{C}^{l}(V, V_{\mathbf{h}}). \end{aligned} \tag{25}$$

Note that $\overline{C}^{l}(V, V_{\mathbf{h}})$ represents the average of the point covariance $C^{l}(\mathbf{h})$ when one extremity of the vector **h** describes the support *V* and the other extremity independently describes the translated (by vector **h**) support $V_{\mathbf{h}}$. If we set

$$b_{ij}^{l} = \sum_{p=1}^{n_l} a_{ip}^{l} a_{jp}^{l}, \quad i = 1,\ldots,N, \quad j = 1,\ldots,N, \tag{26}$$

then it follows from Equation (25):

$$C_{ij}^{V}(\mathbf{h}) = \sum_{l=1}^{L} b_{ij}^{l} \overline{C}^{l}(V, V_{\mathbf{h}}),, \quad i = 1, \ldots, N, \quad j = 1, \ldots, N. \tag{27}$$

Therefore to determine the direct and cross covariances between any two random variables $Z_i^V(\mathbf{u})$ and $Z_j^V(\mathbf{u}+\mathbf{h})$, $i, j = 1, \ldots, N,$ defined at the block support of size *V* we need only to determine average covariances $\overline{C}^{l}(V, V_{\mathbf{h}})$, $l = 0, \ldots, L,$ and the $(L+1)P^2$ coefficients b_{ij}^{l}. Because a joint matrix of covariance functions $C_{ij}(\mathbf{h}), \quad i, j = 1, \ldots, N,$ must be positive semi-definite; this requires the covariance models $C^{l}(\mathbf{h}), \ l = 0, \ldots, L,$ and *L*+1 matrices of b_{ij}^{l} coefficients to be positive semi-definite. Note, however, if the covariance models $C^{l}(\mathbf{h}), \ l = 0, \ldots, L,$ are chosen to be known positive semi-definite models such as Spherical, Exponential, Gaussian, etc., then the average covariances calculated based on this models will be also positive-definite (this is a know property of integration). Model (27) is the liner model of coregionalization at a block support. It is interesting to note that if

$$C_{ij}(\mathbf{h}) = \sum_{l=1}^{L} b_{ij}^{l} C^{l}(\mathbf{h}), \quad i = 1, \ldots, N, \quad j = 1, \ldots, N, \tag{28}$$

is a feasible linear model of coregionalization at a point support then

$$C_{ij}^{V}(\mathbf{h}) = \sum_{l=1}^{L} b_{ij}^{l} \overline{C}^{l}(V, V_{\mathbf{h}}),, \quad i = 1, \ldots, N, \quad j = 1, \ldots, N. \tag{29}$$

is a feasible linear model of coregionalization at a block support and vise versa. This establishes an interesting link between point and block linear model of coregionalization.

Calculating Average Covariance $\overline{C}(V, V_{\mathbf{h}})$ and Avarage Variogram $\overline{\gamma}(V, V_{\mathbf{h}})$

Now the only issue remains is to calculate the values of the average covariances $\overline{C}(V, V_{\mathbf{h}})$ and average variogram values $\overline{\gamma}(V, V_{\mathbf{h}})$. Unfortunately, except for simplistic cases (see Journel and Huijbregts 1978), there is no analytical solution for average covariances and semivariograms for any of the commonly used variogram models, that is, spherical, exponential and Gaussian. If you recall the same is the case for gamma-bar values. Interestingly enough, this fact in itself shows that the assumption of invariance of variogram shape to the scale change is unrealistic.

Therefore, to calculate the average covariances $\overline{C}(V, V_{\mathbf{h}})$, giving the covariance $C^{V}(\mathbf{h})$ for the scale *V*, numerical discretization of the volumes V and $V_{\mathbf{h}}$ is used. Specifically, the average covariances are calculated as

$$\begin{aligned} \overline{C}(V, V_{\mathbf{h}}) = C^{V}(\mathbf{h}) = \frac{1}{V}\frac{1}{V}\int_{V}\int_{V} C(\mathbf{w} - (\mathbf{y} + \mathbf{h}))d\mathbf{w}d\mathbf{y} \approx \\ \frac{1}{n}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} C(\mathbf{u}_i - (\mathbf{u}_j + \mathbf{h})) = C(0) - \frac{1}{n}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\gamma(\mathbf{u}_i - (\mathbf{u}_j + \mathbf{h})). \end{aligned} \tag{30}$$

The average semivariograms are approximated by

$$\overline{\gamma}(V, V_{\mathbf{h}}) = \frac{1}{V}\frac{1}{V}\int_{V}\int_{V}\gamma(\mathbf{w} - (\mathbf{x} + \mathbf{h}))d\mathbf{w}\, d\mathbf{x} \approx \frac{1}{n}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\gamma(\mathbf{u}_i - (\mathbf{u}_j + \mathbf{h})), \tag{31}$$

then the semivariogram $\gamma^{V}(\mathbf{h})$ for the scale *V* is calculated as

$$\gamma^{V}(\mathbf{h}) = \overline{\gamma}(V, V_{\mathbf{h}}) - \overline{\gamma}(V, V) \approx \frac{1}{n}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\gamma(\mathbf{u}_i - (\mathbf{u}_j + \mathbf{h})) - \frac{1}{n}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\gamma(\mathbf{u}_i - \mathbf{u}_j). \tag{32}$$

One of the existing programs of the extended GSLIB group called `gammabar` was modified in order to implement calculation of the semivariograms and covariances for the scale *V* for different lag distances **h** in any desired direction. The parameter file for the updated program called `gammabar_new` is given below.

```
                  Parameters for GAMMABAR_NEW
                  ***************************

START OF PARAMETERS:
160.0  160.0   50.0           -X,Y,Z size of block
 21     21     21             -X,Y,Z discretization
2   0.0                       -nst, nugget effect
1 0.6    0.0     0.0   0.0    -  it,cc,ang1,ang2,ang3
      1000.0  3000.0  12.0    -  a_hmax, a_hmin, a_vert
1 0.4    0.0     0.0   0.0    -  it,cc,ang1,ang2,ang3
     30000.0  6000.0  50.0    -  a_hmax, a_hmin, a_vert
average_vmodel.var            -file for the average variogram output
2   40                        -number of directions and lags
 0.0   0.0    0.5             -azm, dip, lag distance
90.0   0.0    0.5             -azm, dip, lag distance
```

Example: Scaling Laws vs. Direct Variogram Upscaling

This section is aimed at accentuating the difference between upscaled variograms obtained using scalling laws and the ones obtained from theory via direct variogram upscaling approach.

Let us consider a point second-order stationary random function $Z(\mathbf{w})$ with the following 3D isotropic semivariogram model characterizing its spatial continuity

$$\gamma(\mathbf{h}) = 0.7Sph_{a=5}(\mathbf{h}) + 0.3Exp_{a=30}(\mathbf{h});$$

and calculate the semivariogram for the regularization $Z^V(\mathbf{w})$ of the point random function $Z(\mathbf{w})$ over the volume V for different volume sizes *V*. In particular, we consider three different block sizes for volume V. These are cubes with length 2m, 5m, and 10m.

Figure 1 shows comparison of the upscaled variograms obtained using scaling laws and direct variogram upscaling approach for the three considered block sizes used for averaging. Figure 1 also shows point scale variogram.

It can be clearly noted from Figure 1 that the shape of the variogram changes when changing support of data. With increase in the block size the shape of the block support variograms becomes more pronouncedly Gaussian for small lag distances, see Figure 2. Moreover, the departure between upscaled variogram predicted using scaling laws and theoretical upscaled variograms obtained via direct upscaling approach also increase with increase in the modeling scale (block volume). Therefore, it is apparent that in order to correctly predict the variogram at larger scale using a small scale variogram model a direct upscaling approach should be used; not the scaling laws.

Conclusions

A fundamental assumption underlying the scaling laws for the variogram is that of shape invariance. The theoretically derived approach presented here makes no such assumption. In fact, the examples show that there is a change in the shape of the variogram, specifically a smooth Gaussian structure at short scale can be expected with upscaling to a larger volume. This is consistent with the effects of block averaging.

A practical approach to determine directly upscaled variograms is based on a numerical integration that approximates the analytical integral of these variogram models. As with average variogram or average covariance calculations, the approximation is robust given sufficient discretization. A prototype program *gammabar_new* is developed for this purpose.

Furthermore, this direct upscaling approach is easily extended to cross-variograms; this facilitates the development of a scale consistent linear model of coregionalization which is required for consistent modeling at the block scale.

This approach presents numerous exciting future research prospects. One area for further development is a method to downscale the block scale variogram, such that fine scale models can be constructed. This is the

same objective as the work of Kupfersberger et al (1998), but the goal here will be to avoid use of the scaling laws. Another possible area of research will be to develop a linear model of coregionalization that is consistent at all scales, which could then be used to truly integrate data at different supports without any prior compositing required.

Acknowledgements

This research was partially supported by Alberta Ingenuity Foundation, University of Alberta and industry sponsors of the Centre for Computational Geostatistics.

References

1. Almeida A, Journel A G (1994) Joint simulation of multiple variables with a Markov-type coregionalization model. Math Geol 26:565–588
2. Bourgault G (1997) Using non-Gaussian distributions in geostatistical simulations. Math Geol 29:315-334
3. Caers J (2000) Adding local accuracy to direct sequential simulation. Math Geol 32:815-850
4. Deutsch C V, Journel A G (1998) GSLIB: Geostatistical software library and user'sgGuide. Oxford University Press, New York
5. Doyen P M, den Boer L D, Pillet W R (1996) Seismic porosity mapping in the Ekofisk Field Using a New Form of Collocated Kriging. SPE 36498
6. Frykman P, Deutsch C V (2002) Practical application of geostatistical scaling laws for data integration. Petrophysics 43:153-171
7. Frykman P, Deutsch C V (1999) Geostatistical scaling laws applied to core and log data. SPE 56822
8. Goovaerts P (1997) Geostatistics for natural resources evaluation. Oxford University Press, New York
9. Journel A G, Kyriakidis P C (2004) Evaluation of mineral reserves: a simulation approach. Oxford University Press, New York
10. Journel A G, Huijbregts C J (1978) Mining geostatistics. Academic Press, London
11. Kupfersberger H, Deutsch C V, Journel A G (1998) Deriving constraints on small-scale variograms due to variograms of large-scale data. Math Geol 30:837-852
12. Marechal A (1984) Kriging seismic data in presence of faults. In Verly G et al (eds) Geostatistics for natural resources characterization. Reidel, Dordrecht, Holland
13. Oz B, Deutsch C V, Tran, T T, Xie Y (2003) DSSIM-HR: A Fortran 90 program for direct sequential simulation with histogram reproduction. Comp & Geosc 29:39-51
14. Ren W, Leuangthong O, Deutsch C V (2007) Global resource uncertainty using a spatial/multivariate decomposition approach. JCPT, March 2007
15. Soares A (2001) Direct sequential simulation and cosimulation. Math Geol 33:911-926
16. Xu W, Tran T T, Srivastava R M, Journel A G (1992) Integrating seismic data in reservoir modeling: the collocated cokriging alternative. SPE 24742

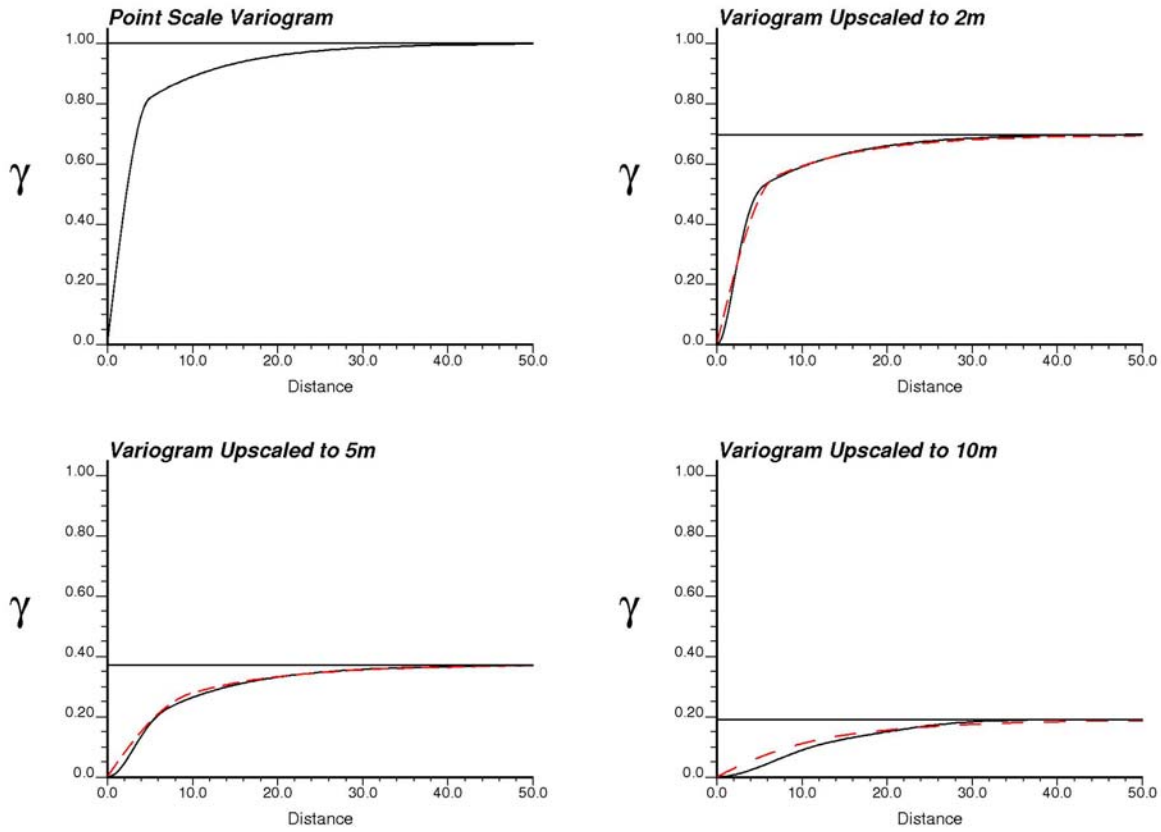

Figure 1. Comparison of the upscaled variograms obtained using scaling laws (dashed lines) and direct variogram upscaling approach (solid lines) for the block support of 2m, 5m and 10m.

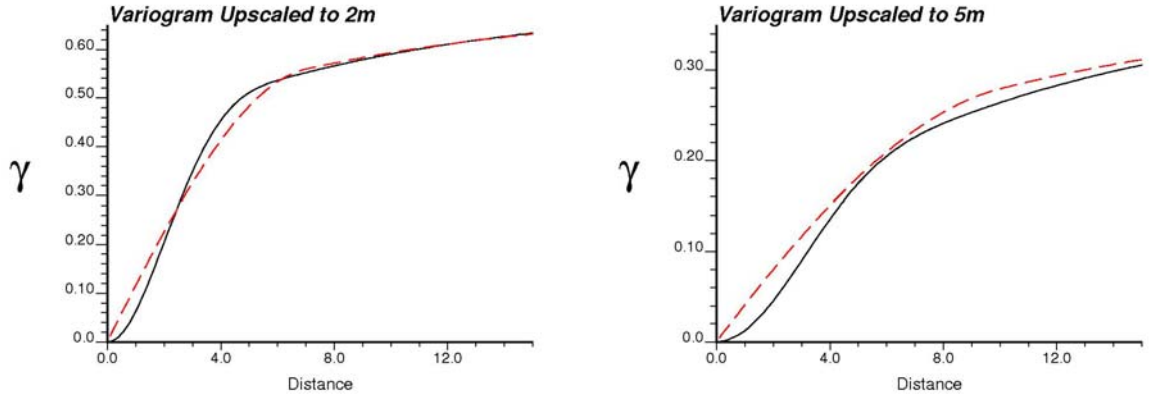

Figure 2. Comparison of the upscaled variograms obtained using scaling laws (dashed lines) and direct variogram upscaling approach (solid lines) for the block support of 2m and 5m for lag distances up to 10m.