

# Methodology for Choosing the Optimal Tolerance Parameters for Variogram Calculation

S. Hadi Derakhshan, Clayton V. Deutsch and Oy Leuangthong

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

*The variogram model used in estimation and/or simulation is important. Calculating and fitting a stable more representative variogram model leads to more reasonable models. The variogram is a two-point statistic that spatially relates two random variables (RV),  $Z(\mathbf{u})$  and  $Z(\mathbf{u}+\mathbf{h})$ . Data are used to infer the variogram for all  $\mathbf{h}$  lag vectors. Calculating experimental variograms in the case of regularly gridded data is much easier than the case of irregular data. Tolerance parameters must be chosen for irregularly located data to permit sufficient data pairs for reliable calculation. Defining these tolerance parameters is a challenge for newcomers to geostatistics. An optimization procedure is developed to establish good automatically chosen variogram parameters. These tolerance parameters are optimized by using a penalty function that accounts for the difference in the fitted variogram and the assumed to be true variogram. The optimal tolerance parameters are the ones that minimize the penalty function. This is applied for omnidirectional and directional variograms in 2 and 3-D. This provides a practical tool to establish an experimental variogram with less uncertainty.*

## Introduction

The GSLIB (Deutsch and Journel, 1998) program, gamv, can be used to calculate the experimental variogram points for a given irregular data set. The user should define some tolerance parameters (they will be defined later in this paper) based on the data locations to calculate the experimental points. The parameter file for gamv is written below:

```
Parameters for GAMV
*****

START OF PARAMETERS:
data.dat          -file with data
 1  2  0          - columns for X, Y, Z coordinates
 1  3            - number of variables,col numbers
-1.0e21  1.0e21  - trimming limits
gamv.out         -file for variogram output
10              -number of lags
10              -lag separation distance
5              -lag tolerance
1              -number of directions
0.0  90.0  400.0  0.0  90.0  400.0 -azm,atol,bandh,dip,dtol,bandv
0              -standardize sills? (0=no, 1=yes)
1              -number of variograms
1  1  1         -tail var., head var.,variogram type
```

The bolded text shows the tolerance parameters. They should be set reasonably in the case of 2D or 3D data set to get a meaningful variogram. Each of these parameters affects the calculated variogram. The “best” parameters would lead to a final fit that is close to the unknown true variogram. Choosing the tolerance parameters in the case of irregularly gridded data is complicated. The tolerance parameters should be chosen in such a way that there is enough number of pairs for a reliable variogram value. There are more data aligned in the vertical direction than horizontal directions; therefore it is better to separate these directions. The tolerance parameters are different in these two cases.

### *Tolerance in vertical directions*

There are four tolerance parameters in this case (Deutsch, 2002):

- Unit lag separation distance,  $h$  ( $h$  is a scalar quantity); it is equal to data spacing in the case of regularly spaced data. In petroleum applications, the vertical data are regularly spaced, therefore choosing the unit lag separation distance is not a problem.
- The distance tolerance,  $h_{tol}$  (similar to  $h$ ,  $h_{tol}$  is a scalar quantity); in real case this parameter usually takes the values of  $0.25h$ ,  $0.5h$  and  $0.75h$ . The value of  $0.5h$  is used in almost all of the cases. The value of  $0.25h$  is used when there are many data on a nearly regular grid and in the case of small number of pairs for each lag, it is recommended to use  $0.75h$ .
- The angle tolerance,  $a_{tol}$ ; this parameter is used when there is some deviation from vertical direction.
- A bandwidth parameter,  $b$ ; this parameter is defined to limit the region for finding data pairs for variogram calculation in vertical direction. After a certain lag distances bandwidth parameter is applied to limit the tolerance region.

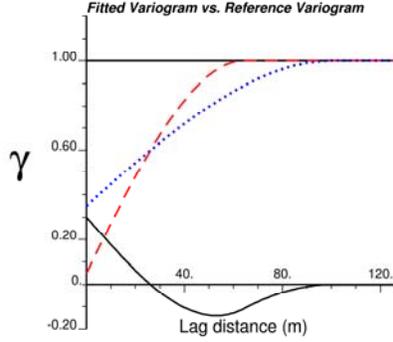
#### *Tolerance in horizontal directions*

There are six tolerance parameters in this case; they are (Deutsch, 2002):

- Unit lag separation distance,  $h$ ; the properties of this parameter are the same as vertical case.
- The distance tolerance,  $h_{tol}$ ; the properties of this parameter are the same as vertical case.
- The horizontal angle tolerance,  $a_{tol}^h$ ; this parameter can be applied to limit the direction which we are interested to calculate the associated variogram. In the case of omnidirectionality, it can be set to 90 degree or a greater angle to have all of the points in horizontal plane. The omnidirectional experimental variogram averages the variability over all directions.
- The horizontal bandwidth,  $b_{hor}$ ; similar to the vertical bandwidth, it is used to define a limited tolerance region in horizontal plane. In the case of omnidirectionality, the horizontal bandwidth should be set to a large number to have all of the points in all directions to calculate the omnidirectional variogram value.
- The vertical angle tolerance,  $a_{tol}^v$ , used to define an angle tolerance about the horizontal plane. It should be set to a small value because of the large variability in vertical direction which can affect the true variogram calculation for a specific stratigraphic layer.
- The vertical bandwidth,  $b_{ver}$ ; this bandwidth relates to the vertical angle tolerance and is defined to limit the stratigraphic layer. It should be set to a small value to have a good approximation of the variogram.

#### **Methodology and Procedures**

To optimize the tolerance parameters a penalty function is defined and the optimal tolerance parameters are obtained by minimizing this penalty function. This penalty function gives a penalty for each set of parameters, this can be obtained by adding all of the errors (between the fitted variogram and the true or assumed to be true variogram) associated to each lag distance (and for each set of parameters). Figure 1 shows schematically the difference function between the fitted variogram and the reference variogram which is used in the definition of the penalty function.



**Figure 1** Fitted variogram (dotted line) versus the reference variogram (dashed line) and their difference (fitted minus reference; solid line)

So the error can be written in integral form as below:

$$p_{i_{realz}}(\theta) = \frac{1}{2a} \int_0^{2a} [\gamma_{i_{realz}}^{fit}(\mathbf{h}; \theta) - \gamma^{ref}(\mathbf{h})]^2 \cdot wt(\mathbf{h}) \cdot d\mathbf{h} \quad (1)$$

Where:

- $p_{i_{realz}}(\theta)$  ; Penalty function for the  $i_{realz}$  realization and set of tolerance parameters  $\theta$
- $i_{realz}$  ; Realization number
- $\theta$  ; A set of tolerance parameters
- $a$  ; Reference variogram range
- $\gamma_{i_{realz}}^{fit}(\mathbf{h}; \theta)$  ; Fitted variogram of the experimental variogram points based on the tolerance parameters  $\theta$ .
- $\gamma^{ref}(\mathbf{h})$  ; Reference variogram
- $wt(\mathbf{h})$  ; Inverse distance weight associated to each lag distance

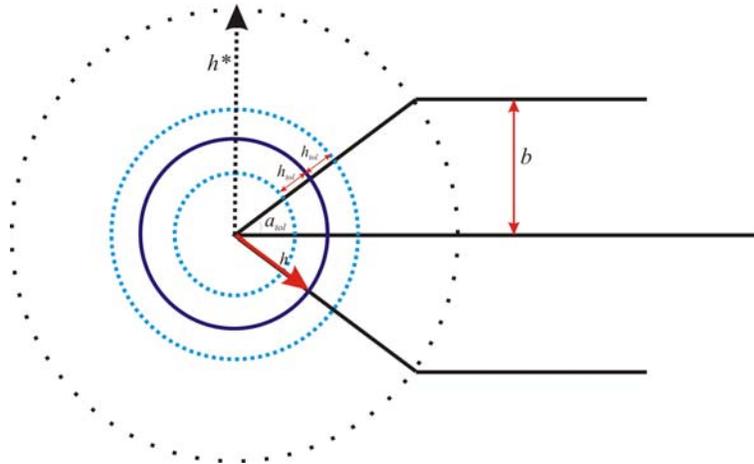
The inverse distance weighting associated to each lag is calculated from below formula:

$$wt(\mathbf{h}) = \begin{cases} \frac{1}{\mathbf{h} + \varepsilon} & ; \quad \text{for } \mathbf{h} \text{ such that } \gamma^{fit}(\mathbf{h}; \theta) > \gamma^{ref}(\mathbf{h}) \\ \frac{0.5}{\mathbf{h} + \varepsilon} & ; \quad \text{for } \mathbf{h} \text{ such that } \gamma^{fit}(\mathbf{h}; \theta) < \gamma^{ref}(\mathbf{h}) \end{cases} \quad (2)$$

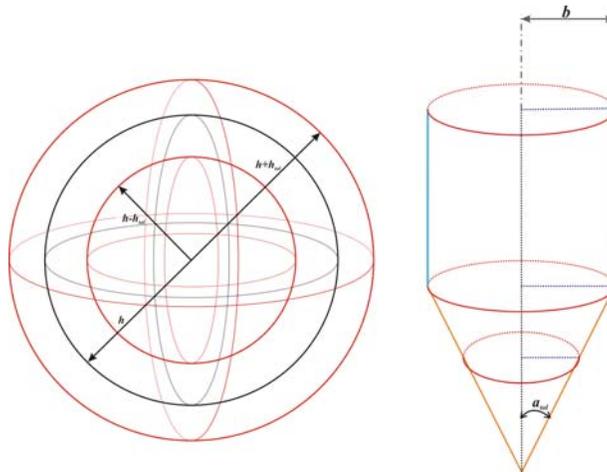
The reason for inverse weighting of the difference between the true variogram and the reference variogram is that the small lag distances which show the short variability get more weight and the large lag distances get less weight. The small number,  $\varepsilon$ , is used in the denominator because in this case the weight at  $\mathbf{h}=0$  can be defined (which is a large number).  $\theta$  has different number of components depending on 2D or 3D cases. In 2D case  $\theta$  has at most 4 components: lag separation distance,  $h$ , lag tolerance,  $h_{tol}$ , angle tolerance,  $a_{tol}$ , and the bandwidth,  $b$ . In the case of omnidirectionality where the angle tolerance is 90 degrees and the bandwidth is a large value,  $\theta$  has 2 components of lag separation distance and the lag tolerance only. These four parameters are shown in Figure 2.

In 3D case since we are dealing with vertical and horizontal (minor and major) directions, therefore two different sets of parameters can be defined. For the vertical lag distance (Figure 3)  $\theta$  has at most 4

components these are the unit lag separation distance,  $h$ , lag tolerance,  $h_{tol}$ , angle tolerance,  $a_{tol}$ , and the bandwidth,  $b$ .

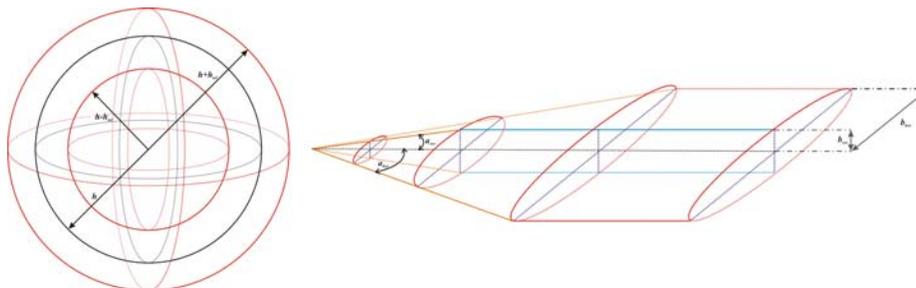


**Figure 2** Lag separation distance,  $h$ , lag tolerance,  $h_{tol}$ , azimuth tolerance,  $a_{tol}$  and bandwidth,  $b$



**Figure 3** Lag separation distance,  $h$ , lag tolerance,  $h_{tol}$ , angle tolerance,  $a_{tol}$  and bandwidth,  $b$  in three dimensional case-vertical lag distance

For the horizontal lag distance (Figure 4)  $\theta$  has at most 6 components: unit lag separation lag distance,  $h$ , the lag tolerance,  $h_{tol}$ , the vertical angle tolerance,  $a_{tol}^v$ , the horizontal angle tolerance,  $a_{tol}^h$ , the vertical bandwidth,  $b_{ver}$  and the horizontal bandwidth,  $b_{hor}$ . The number of components can be decreased in case of omnidirectionality or in the case with no bandwidth.



**Figure 4** Lag separation distance,  $h$ ; lag tolerance,  $h_{tol}$ ; vertical angle tolerance,  $a_{tol}^v$ ; horizontal angle tolerance,  $a_{tol}^h$ ; vertical bandwidth,  $b_{ver}$  and horizontal bandwidth,  $b_{hor}$  in three dimensional case-horizontal lag distance

Besides the tolerance parameters another parameter can be defined in order to capture the relationship between the tolerance parameters and the penalty function. This parameter which is named tolerance ratio is a function of all of the tolerance parameters which are used to calculate the experimental variogram. A tolerance ratio is equal to the tolerance area (in 2D case) or volume (in 3D case) divided by the area or the volume of the semi-circle or hemi-sphere with a radius of  $2h$  ( $h$  is the unit lag distance). For example this tolerance ratio is 100 % in the case of the calculation of the omnidirectional variogram (the only tolerance parameters are unit lag distance and the lag tolerance) with the lag tolerance,  $h_{tol}$ , which is equal to unit separation lag distance;  $h$ . Figure 5 schematically explains the tolerance ratio in 2D case. The tolerance ratio will be used as one of the parameters for optimizing the tolerance parameters. In general there are two different cases for calculating the tolerance ratio; 2D (omnidirectional and general case) and 3D (omnidirectional, horizontal and vertical lag distances).

In next sections the formulas for tolerance ratio are given for different cases.

#### Tolerance ratio in 2D

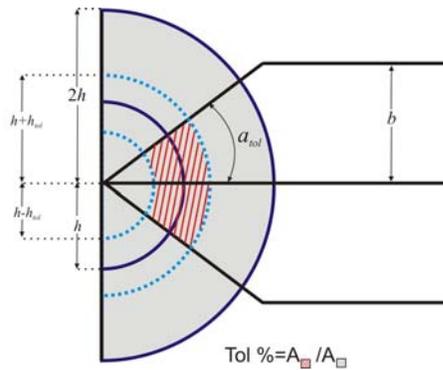
##### a) Omnidirectional case

In this case tolerance ratio depends only on unit lag distance,  $h$  and the lag tolerance,  $h_{tol}$ . The bandwidth and the angle tolerance should be set to a large number and 90 degree in order to calculate the omnidirectional experimental variogram by using `gamv` program in GSLIB. The area function  $A(r)$  is the area of the semi-circle with the radius of  $r$  (see Figure 5).

$$A(r) = \frac{\pi}{2} r^2 \quad (3)$$

Therefore the tolerance ratio is:

$$Tol = \frac{A(h + h_{tol}) - A(h - h_{tol})}{A(2h)} = \frac{h_{tol}}{h} \quad (4)$$



**Figure 5** Tolerance ratio for two dimensional case

##### b) General case

In this case the tolerance area is function of unit lag distance,  $h$ , the lag tolerance,  $h_{tol}$ , angle tolerance,  $a_{tol}$  and the bandwidth,  $b$ . Depending on the magnitude of the unit lag distance,  $h$ , three cases would be happen. To specify these three cases another parameter,  $h^*$ , can be defined which is function of the angle tolerance,  $a_{tol}$  and the bandwidth,  $b$ .  $h^*$  can be calculated by using below formula (see Figure 2):

$$h^* = \frac{b}{\sin a_{tol}} \quad (5)$$

It can be seen from Figure 2 that when  $h \leq (h^* - h_{tol})$ , the tolerance area is made by using the unit lag distance, lag tolerance and angle tolerance (note that the tolerance area in this case is independent of the magnitude of bandwidth). If  $h > (h^* + h_{tol})$  the tolerance area is function of unit lag distance, lag tolerance and bandwidth (in this case the tolerance area is independent of the magnitude of angle tolerance) and if  $(h^* - h_{tol}) < h \leq (h^* + h_{tol})$  then the tolerance area is function of all four tolerance parameters.

The three cases for different lag distances can be defined as:

1.  $h \leq (h^* - h_{tol})$ ;

$$A_1(r, \alpha) = \alpha \cdot r^2 \quad (6)$$

$$Tol = \frac{A_1(h + h_{tol}, a_{tol}) - A_1(h - h_{tol}, a_{tol})}{A(2h)} = \left(\frac{2a_{tol}}{\pi}\right) \cdot \left(\frac{h_{tol}}{h}\right) \quad (7)$$

$a_{tol}$  is in radians.

2.  $h > (h^* + h_{tol})$ ;

$$A_2(r, b) = \left[ \left(\frac{b}{r}\right) \cdot \sqrt{1 - \left(\frac{b}{r}\right)^2} + \arcsin\left(\frac{b}{r}\right) \right] \cdot r^2 \quad (8)$$

$$Tol = \frac{A_2(h + h_{tol}, b) - A_2(h - h_{tol}, b)}{A(2h)} \quad (9)$$

3.  $(h^* - h_{tol}) < h \leq (h^* + h_{tol})$ ;

$$Tol = \frac{A_1(h^*, a_{tol}) - A_1(h - h_{tol}, a_{tol}) + A_2(h + h_{tol}, b) - A_2(h^*, b)}{A(2h)} \quad (10)$$

### Tolerance ratio in 3D

#### a) Omnidirectional case

Similar to the 2D case, the tolerance ratio only depends on the unit lag distance,  $h$ , and the lag tolerance,  $h_{tol}$ , in this case instead of area function; the volume function should be defined which is equal to the volume of the hemisphere with a radius of  $r$ .

$$V(r) = \frac{2\pi}{3} \cdot r^3 \quad (11)$$

Therefore from equation(11), we will have:

$$Tol = \frac{V(h + h_{tol}) - V(h - h_{tol})}{V(2h)} = \frac{1}{4} \cdot \left(\frac{h_{tol}}{h}\right) \cdot \left[ 3 + \left(\frac{h_{tol}}{h}\right)^2 \right] \quad (12)$$

#### b) General case

For the general 3D case, two scenarios can be considered; the first one is for the vertical lag distance (Figure 3) and the second one is for horizontal lag distance (Figure 4).

*b-1) Vertical lag distance*

In this case the tolerance volume is a function of unit lag distance,  $h$ , the lag tolerance,  $h_{tol}$ , angle tolerance,  $a_{tol}$  and the bandwidth,  $b$  (see Figure 3). Based on the magnitude of the unit lag distance three different cases exist.

1.  $h \leq (h^* - h_{tol})$ ;

$$V_1(r, \alpha) = \frac{2\pi}{3} \cdot (1 - \cos \alpha) \cdot r^3 \quad (13)$$

$$Tol = \frac{V_1(h + h_{tol}, a_{tol}) - V_1(h - h_{tol}, a_{tol})}{V(2h)} \quad (14)$$

Therefore from equations(11), (13) and(14), we will have:

$$Tol = \frac{1}{4} \cdot (1 - \cos a_{tol}) \cdot \left(\frac{h_{tol}}{h}\right) \cdot \left[3 + \left(\frac{h_{tol}}{h}\right)^2\right] \quad (15)$$

2.  $h > (h^* + h_{tol})$ ;

$$V_2(r, b) = \frac{2\pi}{3} \cdot \left[1 - \left(1 - \left(\frac{b}{r}\right)^2\right)^{\frac{3}{2}}\right] \cdot r^3 \quad (16)$$

$$Tol = \frac{V_2(h + h_{tol}, b) - V_2(h - h_{tol}, b)}{V(2h)} \quad (17)$$

3.  $(h^* - h_{tol}) < h \leq (h^* + h_{tol})$ ;

$$Tol = \frac{V_1(h^*, a_{tol}) - V_1(h - h_{tol}, a_{tol}) + V_2(h + h_{tol}, b) - V_2(h^*, b)}{V(2h)} \quad (18)$$

*b-2) Horizontal lag distance*

In this case the tolerance volume is a function of the unit lag distance,  $h$ , the lag tolerance,  $h_{tol}$ , the vertical angle tolerance,  $a_{tol}^v$ , the horizontal angle tolerance,  $a_{tol}^h$ , the vertical bandwidth,  $b_{ver}$  and the horizontal bandwidth,  $b_{hor}$  (see Figure 4):

$$Tol = f(h, h_{tol}, a_{tol}^v, a_{tol}^h, b_{ver}, b_{hor}) \quad (19)$$

As before the magnitude of the unit lag distance results in different scenarios:

1.  $h \leq (h_{ver}^* - h_{tol})$ ;  $(h_{ver}^* = \frac{b_{ver}}{\sin a_{tol}^v})$ ;

$$V_3(r, \alpha, \beta) = \frac{2\pi}{3} \cdot [1 - I_1(\alpha, \beta)] \cdot r^3 \quad (20)$$

Where

$$I_1(\alpha, \beta) = \frac{1}{2\pi} \cdot \int_0^{2\pi} \left( \frac{\cot^2 \alpha \cdot \cos^2 \theta + \cot^2 \beta \cdot \sin^2 \theta}{1 + \cot^2 \alpha \cdot \cos^2 \theta + \cot^2 \beta \cdot \sin^2 \theta} \right)^{\frac{1}{2}} \cdot d\theta \quad (21)$$

$$Tol = \frac{V_3(h + h_{tol}, a_{tol}^h, a_{tol}^v) - V_3(h - h_{tol}, a_{tol}^h, a_{tol}^v)}{V(2h)} \quad (22)$$

Therefore from equations(20), (21) and(22), we will have:

$$Tol = \frac{1}{4} \cdot \left[ 1 - I_1(a_{tol}^h, a_{tol}^v) \right] \cdot \left( \frac{h_{tol}}{h} \right) \cdot \left[ 3 + \left( \frac{h_{tol}}{h} \right)^2 \right] \quad (23)$$

$$2. \quad (h_{ver}^* + h_{tol}) < h \leq (h_{hor}^* - h_{tol});$$

$$V_4(r, b, \alpha) = \frac{2\pi}{3} \cdot \left[ 1 - I_2(r, b, \alpha) - \frac{3 \sin(2\alpha)}{8} \cdot \left( \frac{b}{r} \right) \right] r^3 \quad (24)$$

Where

$$I_2(r, b, \alpha) = \frac{1}{2\pi} \cdot \int_0^{2\pi} \left[ \frac{\csc^2 \alpha \cdot \cos^2 \theta + \left( \frac{r}{b} \right)^2 \cdot \sin^2 \theta - 1}{\csc^2 \alpha \cdot \cos^2 \theta + \left( \frac{r}{b} \right)^2 \cdot \sin^2 \theta} \right]^{\frac{3}{2}} \cdot d\theta \quad (25)$$

$$Tol = \frac{V_4(h + h_{tol}, b_{ver}, a_{tol}^h) - V_4(h - h_{tol}, b_{ver}, a_{tol}^h)}{V(2h)} \quad (26)$$

$$3. \quad (h_{ver}^* - h_{tol}) < h \leq (h_{ver}^* + h_{tol});$$

$$Tol = \frac{V_3(h_{ver}^*, a_{tol}^h, a_{tol}^v) - V_3(h - h_{tol}, a_{tol}^h, a_{tol}^v) + V_4(h + h_{tol}, b_{ver}, a_{tol}^h) - V_4(h_{ver}^*, b_{ver}, a_{tol}^h)}{V(2h)} \quad (27)$$

$$4. \quad h > (h_{hor}^* + h_{tol});$$

$$V_5(r, a, b) = \frac{2\pi}{3} \cdot \left[ 1 - I_3(r, a, b) \right] \cdot r^3 \quad (28)$$

Where

$$I_3(r, a, b) = \frac{1}{2\pi} \cdot \int_0^{2\pi} \left[ \frac{\left( \frac{r}{a} \right)^2 \cdot \cos^2 \theta + \left( \frac{r}{b} \right)^2 \cdot \sin^2 \theta - 1}{\left( \frac{r}{a} \right)^2 \cdot \cos^2 \theta + \left( \frac{r}{b} \right)^2 \cdot \sin^2 \theta} \right]^{\frac{3}{2}} d\theta \quad (29)$$

$$Tol = \frac{V_5(h + h_{tol}, b_{hor}, b_{ver}) - V_5(h - h_{tol}, b_{hor}, b_{ver})}{V(2h)} \quad (30)$$

$$5. \quad (h_{hor}^* - h_{tol}) < h \leq (h_{hor}^* + h_{tol});$$

$$Tol = \frac{V_4(h_{hor}^*, b_{ver}, a_{tol}^h) - V_4(h - h_{tol}, b_{ver}, a_{tol}^h) + V_5(h + h_{tol}, b_{hor}, b_{ver}) - V_5(h_{hor}^*, b_{hor}, b_{ver})}{V(2h)} \quad (31)$$

### Methodology

The procedure for optimizing the tolerance parameters is as below:

1. Use spatial bootstrap (Deutsch, 2004) to obtain  $L$  values (realizations) for each of the data locations.
2. Determine different reasonable sets of tolerance parameters for variogram calculation.
3. For each set of tolerance parameters and each realization:
  - a. Calculate the experimental variogram.
  - b. Fit the experimental variogram with a variogram model,  $\gamma_{i_{realz}}^{fit}(\mathbf{h}; \theta)$ , by using `varfit` program (Neufeld and Deutsch, 2004).
  - c. Calculate the penalty function,  $p_{i_{realz}}(\theta)$ , by using equation (1).
4. For each set of tolerance parameters, calculate final penalty by averaging over different calculated penalty for different spatial bootstrap realizations.
5. Map penalty function as a function of tolerance parameters.
6. Find the optimum tolerance parameters by minimizing the penalty function.

### Spatial bootstrap

Spatial bootstrap (Deutsch, 2004) is used to assess and quantify the uncertainty in the variogram at each lag. The spatial bootstrap procedure is as follow:

1. Preliminary analysis:
  - a. Assemble the representative distribution of the random variable  $Z$ ,  $F(z)$ ;
  - b. Define a 3-D variogram model  $\gamma(\mathbf{h})$  of the normal scores of the random variable  $Z$ ;
  - c. Decompose the  $n$  by  $n$  covariance matrix by the Cholesky decomposition into product of upper ( $U$ ) and lower ( $L$ ) triangular matrices:  $C = LU$
2. Generate a new set of data,  $z$ , as

$$z = F_z^{-1}(G(Lw)) \quad (32)$$

where  $w$  is a  $n$  by 1 vector of independent Gaussian values and  $G(\cdot)$  denotes the standard Gaussian cumulative distribution function.

3. Calculate the statistic of interest from the resampled dataset.
4. Repeat Steps 2 and 3 many times, say,  $L=100$ .
5. Establish the distribution of uncertainty in the calculated statistic.

### Choosing the reference variogram model

Reference variogram model (or the assumed to be true variogram),  $\gamma^{ref}(\mathbf{h})$ , is unknown at the first stage of the optimization. A reasonable guess for the reference variogram model can be a single structure with an isotropic variogram with the range of one-third of the field size and the nugget effect of zero, a sensitivity study on choosing the range of correlation and the nugget effect can be performed to see the impact of these two parameters on tolerance parameters (see Derakhshan *et. al*, 2007-2). After performing the proposed optimization technique and getting the optimal set of tolerance parameters, the experimental variogram can be calculated. The tolerance correction factor should be applied to these calculated points (see Derakhshan *et. al*, 2007-3). By using the corrected experimental variogram points, the variogram model can be fitted. This model can be fed into the first stage of the procedure; it can be iterated to get a reasonable reference variogram model.

This uncertainty has effect on the selection of the optimal tolerance parameters. The penalty function,  $p_{i_{realz}}(\theta)$ , can be calculated for different realizations which are output of spatial bootstrap. Spatial bootstrap needs the original data and also the reference variogram to create different realizations. This reference variogram is the same as  $\gamma^{ref}(\mathbf{h})$  which is used to calculate the penalty function,  $p_{i_{realz}}(\theta)$ . Basically to obtain the tolerance parameters for a given data set first the spatial bootstrap is performed to get different realization at each data location. Two assumptions are made before performing spatial bootstrap; the first one is that the initial data distribution is representative of the entire population and the second one is that the data are spatially correlated; this spatial correlation is represented by the reference variogram which is unknown and also is not modeled at the first step of the tolerance parameters calculations. In chapter 4, a sensitivity analysis will be shown for a real data set on how to choose reasonable variogram range and nugget effect for a single spherical structure for the reference variogram model. The reasonable assumptions for the variogram model could be single spherical structure with range of one-third of the variogram range. Declustering techniques might be performed to get the weights for the data points. Spatial bootstrap needs the distribution of the original data. The cumulative distribution function of the data is required for randomly drawing the data values for the bootstrap. After determining the reference distribution and variogram the spatial bootstrap can be applied to get  $L$  realizations for each of the data locations.

### Experimental variogram calculations for different tolerance parameters

The GSLIB program, `gamv`, can be used to calculate the experimental variogram for different tolerance parameters. Reasonable range for tolerance parameters should be defined for calculations. This range should contain the optimal tolerance parameters. The range for unit lag distance or lag separation distance can be obtained by building the cumulative distribution function of the minimum distance between the data locations. There is just one value for the minimum distance which corresponds to a data location. The minimum distance between the data points is calculated as:

$$d_{i_{data}} = \min \left\{ \left| \mathbf{u}_{i_{data}} - \mathbf{u}_{j_{data}} \right| \right\} \quad (33)$$

where  $d_{i_{data}}$  is the minimum distance between the data points which corresponds to each data location,  $\left| \mathbf{u}_{i_{data}} - \mathbf{u}_{j_{data}} \right|$  is the norm of the separation vector between data locations  $i_{data}$  and  $j_{data}$  ( $i_{data} \neq j_{data}$  and  $i_{data}, j_{data} = 1, 2, 3, \dots, n_{data}$ ). The cumulative distribution function for  $d_{i_{data}}$  can be built and the reasonable range of values for unit lag separation distance can be obtained by using this distribution. The values between  $P10$  and  $P90$  of the  $d_{i_{data}}$  values are used for different lag separation distances. Because of the CPU cost, the number of unit lag distances for different variogram calculations,  $n_h$ , should not be large. To cover the whole range and test different lag separation distance, the following relationship is considered for calculating the unit lag distance for different cases.

$$h = \left( \frac{P90 - P10}{n_h} \right) \cdot i_h + P10 \quad (34)$$

Where  $h$  is the unit lag distance,  $P10$  and  $P90$  are the distances calculated from distribution of  $d_{i_{data}}$ .  $n_h$  is the number of different unit lag distance used for calculating the optimal one and  $i_h$  is the index for  $n_h$  ( $i_h = 0, 1, 2, \dots, n_h$ ).

For the range of the lag tolerance, the values between 0 % and 100 % of the unit lag distance are used. Therefore

$$h_{tol} = \left( \frac{i_{tol}}{n_{tol}} \right) \cdot h \quad (35)$$

where  $n_{tol}$  is the number of different lag tolerances used for calculating the optimal one ( $i_{tol} = 1, 2, \dots, n_{tol}$ ). Similar to  $n_h$ ,  $n_{tol}$  should not be a large number for CPU efficiency.

For azimuth tolerance and dip tolerance, we can use values between 0 and 90 degrees. The increments can be defined to calculate different angle tolerances. Again the number of increments has impact on CPU time. The range for horizontal and vertical bandwidth can be estimated by using some geological information. In the case of lack of geological information the minimum and maximum possible values for the bandwidths can be considered. The maximum possible value for horizontal bandwidth can be the maximum areal dimension of the field and for the vertical bandwidth it is the thickness of the stratigraphic layer. After obtaining reasonable ranges for the tolerance parameters, the experimental variogram is calculated for each set of tolerance parameters and for each realization obtained from spatial bootstrap.

#### *Fitting the experimental variogram for each realization*

In this step the compatible GSLIB program, `varfit` (Neufeld and Deutsch, 2004), is used to fit the experimental variogram points with an optimized variogram model, the experimental points are different for different tolerance parameters and different realizations. The fitted model,  $\gamma_{i_{realz}}^{fit}(\mathbf{h}; \theta)$ , is used to calculate the penalty function.

#### *Calculating and minimizing the final penalty function*

First the penalty function (see equation(1)) is calculated for each realization and each set of tolerance parameters. After calculating these penalties the final penalty,  $\bar{p}(\theta)$  which is function of tolerance parameters is calculated by averaging the calculated penalties over different realizations for the fixed tolerance parameters. Therefore

$$\bar{p}(\theta) = \frac{1}{n_{realz}} \cdot \sum_{i_{realz}=1}^{n_{realz}} p_{i_{realz}}(\theta) \quad (36)$$

where  $n_{realz}$  is the number of realizations,  $\theta$  corresponds to the fixed tolerance parameters and  $p_{i_{realz}}(\theta)$  is the penalty for realization  $i_{realz}$  and is calculated by using equation(1).

After calculating the final penalty, it should be minimized to get the optimal tolerance parameters. The tolerance parameters have different number of components for different cases. For example for omnidirectional 2D and 3D data sets, there are two components for variogram calculation, they are unit lag distance and the lag tolerance (instead of lag tolerance the tolerance ratio can also be used). For other general cases the number of tolerance components is greater than 2.

For minimizing the penalty function in the presence of multiple variables a sequential type approach could be used to obtain the optimal point which minimizes the penalty function. For example if there are three variables, at first step two of them should be fixed and the penalty should be minimized with respect the variable which is not fixed. After specifying this point the next variable should be optimized and the others should be fixed. This procedure can be applied step by step to get the final optimal point and as mentioned at each step one of the variables should be optimized and the others are fixed and each of the variables is optimized once. After obtaining the optimal tolerance parameters, the experimental variogram should be calculated by using the optimal tolerance parameters and fitted with the `varfit` program.

## Conclusion

A methodology for optimizing tolerance parameters for variogram calculation has been proposed. The methodology is straightforward and the workflow was presented. The optimality is defined as how close the fitted variogram comes to the true variogram. Since the true variogram is unknown, it is assumed and a resampling scheme is followed to establish the optimal tolerance parameters. The entire procedure could be iterated with the latest best estimate of the variogram. A variogram model will be fitted with the experimental variogram points; a penalty is defined based on the difference between these two variogram functions. A spatial bootstrap is used to permit multiple sets of fitted variogram models. The penalty values for each set of tolerance parameters are averaged over different realizations.

A wide variety of optimization techniques might be used for optimizing tolerance parameters. Spatial bootstrap was presented to assess uncertainty in tolerance parameters optimization. Conditional finite domain (CFD) approach (Babak and Deutsch, 2007) can be used instead of spatial bootstrap to parameter the uncertainty in the statistics of interest (i.e. penalty function).

## Reference

- Babak, O. and Deutsch C.V., 2007, "Reserves Uncertainty Calculation Accounting for Parameter Uncertainty", *Canadian International Petroleum Conference*, paper no. 2007-099.
- Derakhshan, S. H., 2007-1, "A Decision Support System for Variogram Calculation", MSc Thesis, Department of Civil and Environmental Engineering, University of Alberta, Alberta, Canada.
- Derakhshan, S. H., Deutsch C.V. and Leuangthong, O., "Application of Optimal Tolerance Parameters for Variogram Calculation", *Centre for Computational Geostatistics Annual Report 9*, Department of Civil and Environmental Engineering, University of Alberta, 2007-2.
- Derakhshan, S. H., Deutsch C.V. and Leuangthong, O., "Variogram Range Estimation from Experimental Variogram with Significant Tolerance", *Centre for Computational Geostatistics Annual Report 9*, Department of Civil and Environmental Engineering, University of Alberta, 2007-3.
- Deutsch, C.V. and Journel, A.G., 1998: *GSLIB - Geostatistical software library and users guide*. Oxford University press, 2nd Edition
- Deutsch, C.V., 2002: *Geostatistical Reservoir Modeling*. Oxford University press
- Deutsch, C.V., "A Statistical Resampling Program for Correlated Data: `Spatial_Bootstrap`", *Centre for Computational Geostatistics Annual Report 6*, Department of Civil and Environmental Engineering, University of Alberta, 2004.
- Neufeld, C. T. and Deutsch, C. V., "Developments in Semiautomatic Variogram Fitting", *Centre for Computational Geostatistics Annual Report 6*, Department of Civil and Environmental Engineering, University of Alberta, 2004.