Application of Optimal Tolerance Parameters for Variogram Calculation

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

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

For experimental variogram calculation in the case of irregular data set (in terms of spatial configurations of the data), some tolerance parameters are defined in order to get a reliable variogram with enough number of pairs. These tolerance parameters can be optimized by using the methodology presented in paper 121 of this annual report. In this paper the application of optimized tolerance parameters for experimental variogram calculation is presented. Two different example will be presented the first one is a synthetic data set and the second one is Amoco data (2D and 3D).

Introduction

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. In 2D case there are four different tolerance parameters, they are unit lag separation distance, h, distance tolerance, h_{tol}, angle tolerance, a_{tol}, bandwidth parameter, b.

Usually in 3D real cases there are more data in vertical direction than horizontal direction; therefore it is better to separate these directions. The tolerance parameters are different in these two cases. The tolerance parameters for vertical direction are the same as 2D case, for horizontal direction, there are six tolerance parameters. They are unit lag separation distance, h, distance tolerance, h_{tol}, horizontal angle tolerance, a^h_{tol}, horizontal bandwidth, b_{hor}, vertical angle tolerance, a^v_{tol} and vertical bandwidth, b_{ver}. The tolerance parameters can be optimized by using a well defined penalty function and assuming a reference variogram. It will be shown that a single structure isotropic variogram with a range of one-third of the field size and a low nugget effect would be a reasonable assumption for reference variogram. The sensitivity study is shown in this paper for choosing the nugget effect and variogram range for a single structure isotropic variogram. The penalty function is written below:

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

Where:

$p_{i_{realz}}\left(\theta\right)$; Penalty function for the i_{realz} realization and set of tolerance parameters θ

i_{realz} ; Realization number

θ ; A set of tolerance parameters

a ; Reference variogram range

$\gamma^{fit}_{i_{realz}}\left(\mathbf{h};\theta\right)$; Fitted variogram of the experimental variogram points based on the tolerance parameters θ.

$\gamma^{ref}\left(\mathbf{h}\right)$; Reference variogram

$wt\left(\mathbf{h}\right)$; Inverse distance weight associated to each lag distance

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

$$wt(\mathbf{h}) = \begin{cases} \dfrac{1}{\mathbf{h}+\varepsilon} & ; & \textit{for } \mathbf{h} \textit{ such that } \gamma^{fit}(\mathbf{h};\theta) > \gamma^{ref}(\mathbf{h}) \\[2ex] \dfrac{0.5}{\mathbf{h}+\varepsilon} & ; & \textit{for } \mathbf{h} \textit{ such that } \gamma^{fit}(\mathbf{h};\theta) < \gamma^{ref}(\mathbf{h}) \end{cases} \tag{2}$$

Synthetic Case

In this synthetic example a 2D data set is considered and we are trying to find the optimum parameters for calculation of experimental omnidirectional variogram, the parameters which can be changed are the lag distance, h and lag tolerance, it should be noted that the angle tolerance is 90 degree and the bandwidth is set to large number to get the omnidirectional variogram.

A synthetic 2D Gaussian random field (Figure 1) is generated via an unconditional simulation for a 1024 x 1024 grid with the following isotropic variogram as a reference variogram:

$$\gamma(\mathbf{h}) = 0.05 + 0.95 Sph_{a=64}(\|\mathbf{h}\|)$$

$Sph_{a=64}(\|\mathbf{h}\|)$ is the isotropic spherical structure with a range of 64:

$$Sph_{a=64}(\|\mathbf{h}\|) = \begin{cases} 1.5 \times \left(\dfrac{\|\mathbf{h}\|}{64}\right) - 0.5 \times \left(\dfrac{\|\mathbf{h}\|}{64}\right)^3 & , \text{ if } \|\mathbf{h}\| < 64 \\[2ex] 1 & , \text{ if } \|\mathbf{h}\| \geq 64 \end{cases}$$

$\|\mathbf{h}\|$ is the norm of vector $\mathbf{h}$.

To make sure that the result has a right variogram that we have used to build it, variogram reproduction is checked. Figure 2 shows the variogram reproduction. Using this reference model, we can then sample randomly at n locations (n=200 for this example) by using the `draw` program (Deutsch and Journel, 1998). The locations of the 200 picked points are shown in Figure 3. Figure 4 shows the reference distribution which is the Gaussian distribution and also the distribution of the 200 picked points.

Now the data set is constructed, the only difference of this data set with the real data set is that we know the reference variogram model of the field for this case. Spatial bootstrap (Deutsch 2004) can be used to assess uncertainty associated with variogram calculation. It gives N realization at each location, by using N sets of data; N different experimental variograms can be calculated, so at each lag there are N values for variogram which can be used to make the variogram distribution at each lag, therefore the variogram uncertainty can be quantified. Figure 5 shows the reference variogram and also the experimental points based on the 200 randomly picked points. Our goal is to quantify the uncertainty in the variogram at the lags that the experimental variogram is calculated.

By using spatial bootstrap method, different realizations (e.g. 100 realizations) can be calculated at each of the 200 data locations. For each realization (which consists of 200 simulated values), the experimental variogram can be calculated. Therefore at each lag distance we have 100 values (realizations) for the variogram. Figure 6 shows the distribution of the experimental variogram at each lag distance and the distribution of the variogram values at $\mathbf{h}$=40. The distribution of the experimental variogram values are plotted at each lag distance in Figure 7. At each lag the variance of the variogram is calculated and it is plotted against the lag distance $\mathbf{h}$. Figure 8 shows the plot of the variance of the variogram from spatial bootstrap against lag distance.

Procedures are performed to find the optimum parameters for calculation of omnidirectional variogram:

1. Using spatial bootstrap to obtain L values (realizations) for each location; e.g. L=100.

2. Calculating experimental variogram for each realization with different tolerance parameters (lag separation distance, tolerance ratio and the number of lags).

$$h = 2, 4, 6, ..., 50$$

$$Tol \ \% = \frac{h_{tol}}{h} \times 100 = 4, 8, 12, ..., 100$$

$$n_{lag} = \text{int}\left(\frac{|A|}{2h}\right) = \text{int}\left(\frac{512}{h}\right)$$

3. For each calculated experimental variogram a variogram should be fitted by using `varfit` program (Neufeld and Deutsch 2004). Fitting variogram will reduce the uncertainty in the variogram. This can be shown by using spatial bootstrap before and after variogram fitting. At each lag distance there are 100 realizations as a result of spatial bootstrap therefore there are 100 variogram fits for them. Figure 9 shows the variogram plot before and after variogram fitting, it can be seen that the range of variation for the variogram values is reduced after fitting. Figure 10 shows the reduction in the variance of the variogram as a result of variogram fitting.

4. By using the fitted variogram in step 3 and the reference variogram, penalty function, can be calculated by using equation(1).

5. Plotting the final penalty versus the lag separation distance and the tolerance ratio. The final penalty is the average of the penalties which are calculated in previous step over different realizations. Figure 11 shows the final penalty versus the tolerance parameters. The conditional expectation of the final penalty is plotted in Figure 13.

6. Trying to find the optimum lag separation distance and the lag tolerance which minimize the final penalty. The optimum values happen at $h=18$ and $Tol=56$ %.

The synthetic case that was presented already gives us the optimal tolerance parameters for a certain example. But it is obvious that the optimal tolerance parameters should be function of the field size, number of data and the reference variogram parameters (nugget effect, range and the type of variogram model, i.e. spherical). To see the effect of number of data and field size on optimal tolerance parameters, different cases were run with field size of (512x512 and 1024x1024) and reference variogram parameters (isotropic 2D spherical variogram with the nugget effect of 0.05 and the range of 64) but with different number of randomly picked data points ($n=128$, 256, 512 and 1024). The reference 2D map for 1024x1024 is the same as Figure 2 but for 512x512 another 2D Gaussian field is generated by using `sgsim`. The reference map for this case is shown in Figure 14. The variogram reproduction is checked for this case and is plotted in Figure 15. The locations of the different randomly picked points are shown in Figure 16. To see better the effect of number of data, for each specific number of randomly picked data points and field size different realizations are considered. The realizations are considered by changing the random seed number in `draw` program. Figure 17 shows four different realizations for 256 randomly picked data points for a fixed field size of 1024.

By using the same procedure the penalty function is calculated for various tolerance parameters. The plots of penalty function versus the lag separation distance and the tolerance ratio are shown in Figure 18, these plots are for just one realization of the data locations (one realization means that a fixed value for random seed number is used in drawing the data points from reference model). The conditional expectation plots for the penalty function for different number of data points and fixed field size of 1024 for different realizations (random seed numbers) are shown in Figure 19. The solid line shows the average over different realizations. The optimum tolerance parameters which minimize the penalty function are calculated and the associated experimental variogram and fitted variogram along with the reference variogram are plotted in Figure 20 for different cases of number of randomly picked data points with a fixed field size. From these plot it can be concluded that as number of data increases (with fixed field size) the number of data per unit area (or volume) will increase, therefore the optimal tolerance ratio should be decreased (Figure 21).

Real Case

The methodology for improving variogram calculation will be applied for Amoco data. Amoco 2D and 3D data set are considered for implementing the methodology. The spatial variable is the porosity of the reservoir. The 2D map of the well locations along with the data values for porosity is shown in Figure 22.

There are 62 wells in the reservoir. The 2D field size of the reservoir is 10400x10400. The histogram and cumulative distribution function of the porosity data are shown in Figure 23.

The nominal spacing between data is 1320.80:

$$\text{nominal spacing} = \sqrt{\frac{A}{n}} = \sqrt{\frac{10400 \times 10400}{62}} = 1320.80$$

The histogram and the cumulative distribution function of the minimum distance between wells are shown in Figure 24.

The procedure that is considered to optimize the variogram tolerance parameters is as below:

1. Transforming data to normal score

2. Spatial Bootstrap; $L=100$, there are 100 values at each well (62 wells)

3. Calculating experimental variogram for each realization for different tolerance parameters:

$$h = 80 \times ilag \qquad ; \quad ilag = 1, 2, 3, \ldots, 25$$

$$Tol~\% = \frac{h_{tol}}{h} \times 100 = 10 \times itol \qquad ; \quad itol = 1, 2, 3, \ldots, 10$$

$$n_{lag} = \text{int}\left(\frac{|A|}{2h}\right) = \text{int}\left(\frac{10400}{2 \times 80 \times ilag}\right) = \text{int}\left(\frac{65}{ilag}\right) \qquad ; \quad ilag = 1, 2, 3, \ldots, 25$$

$25 \times 10 \times 100 = 25,000$ total variograms should be calculated for different tolerance

parameters.

4. Fitting variogram by using `varfit` program for each calculated experimental variogram (25,000 variograms).

5. Defining an assumed to be true variogram with spherical structure and nugget effect of 20 % and the range of one third of the field size as a reference variogram for calculation of the penalty function:

$$a \approx \frac{1}{3}|A| \quad \Rightarrow \quad a \approx 3500$$

$$\gamma(\mathbf{h}) = 0.2 + 0.8 \times Sph_{a=3500}\left(\|\mathbf{h}\|\right)$$

6. Calculating penalty function for each realization and each set of tolerance parameters. At last the penalty function should be averaged over 100 realizations.

Totally, there are 250 points (250 different set of parameters); the plot is smoothed by using inverse distance averaging method and shown in Figure 25. The minimum of the penalty function happens at $h=1098$ and $Tol=10$ %.

The angle tolerance is also optimized. For optimization of angle tolerance, the optimized unit lag separation distance and the lag tolerance ($h=1098$, $h_{tol}=109.8$) are used. In this case the average penalty function is calculated for different angle tolerance (5, 10,… ,90 degrees). The minimum of the penalty function happens at $a_{tol}=40$ degrees. Figure 26 shows the plot of penalty function versus the angle tolerance.

For choosing the variogram model of $\gamma(\mathbf{h}) = 0.2 + 0.8 \times Sph_{a=3500}\left(\|\mathbf{h}\|\right)$, a sensitivity analysis is performed on the selection of the nugget effect and the variogram range by using the Amoco 2D data. First the range of correlation of the assumed to be the reference variogram is fixed to the 30 percent of the field size and then the nugget effect is changed from 0 percent to 100 percent. The plots of the penalty function

for this case is shown in Figure 27. In second case the nugget effect is fixed to 0 percent and now the range of correlation of the variogram is changed from 10 percent to 100 percent of the field size. The plot of the penalty function for this case is shown in Figure 28. As it can be seen from Figure 27 and Figure 28, the low value region for the penalty function is approximately unchanged when the nugget effect varies from 0 percent to 30 percent and the range of correlation (reference variogram range) varies between 10 percent and 100 percent of the field size. In the case of fixed reference variogram range and the nugget effect between 40 percent to 100 percent, the locations of the low values are not the same as other plots. Therefore for reference variogram range, one-third of the field size is a reasonable value for reference variogram.

The Amoco 3D data set is also considered to obtain the optimized vertical bandwidth. The proportional stratigraphic coordinate is used for z-direction. The stratigraphic z-coordinate is not multiplied by the average thickness of the layer therefore it is between 0 and 1. The optimized vertical band width should be multiplied by the average thickness to have the actual value. Other tolerance parameters are assumed to be constant when the vertical bandwidth is optimized. They are assumed to be a reasonable value. The vertical angle tolerance of 15 degree, the unit lag distance was considered to be the average of the data spacing in stratigraphic coordinate (in original coordinate, since the Amoco data is well log data, the vertical data spacing is constant but when the coordinate is transformed stratigraphically, the data spacing varies). Figure 29 shows the plot of the penalty function versus the vertical bandwidth. It can be seen from plot that the optimized vertical bandwidth happens at 0.002 in stratigraphic unit.

Conclusion

A methodology for optimizing tolerance parameters for variogram calculation has been applied for both synthetic and real data. The idea is to minimize the error between the fitted variogram and an assumed to be true variogram therefore the optimal tolerance parameters minimize the penalty function. The assumed true variogram was considered to be one structure isotropic spherical variogram with no nugget effect. A sensitivity on this selection was shown. The sensitive parameters that were checked are the nugget effect (ranging from 0 to 100 percent while the reference variogram range is constant) and the reference variogram range (ranging from 10 percent to 100 percent of the field size while the nugget effect is fixed). This sensitivity showed that the tolerance parameters are approximately the same in these cases for a given data set. Fitting the optimized experimental variogram reduces the uncertainty in the variogram at each lag distance.

References

Derakhshan, S. H., 2007-1, "A Decision Support System for Variogram Calculation", M.Sc. Thesis, Department of Civil and Environmental Engineering, University of Alberta, Alberta, Canada.

Derakhshan, S. H., Deutsch C.V. and Leuangthong, O., "Methodology for Choosing the Optimal Tolerance Parameters for Variogram Calculation", *Centre for Computational Geostatistics Annual Report 9*, Department of Civil and Environmental Engineering, University of Alberta, 2007-2.

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_Boostrap`", *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.
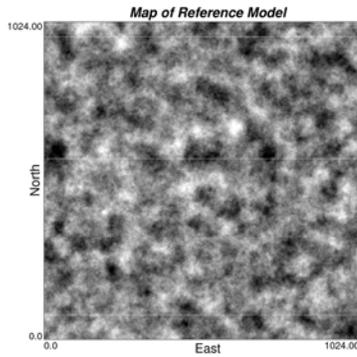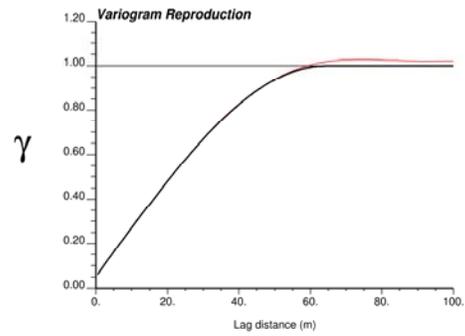
Figure 1: Map of reference model.

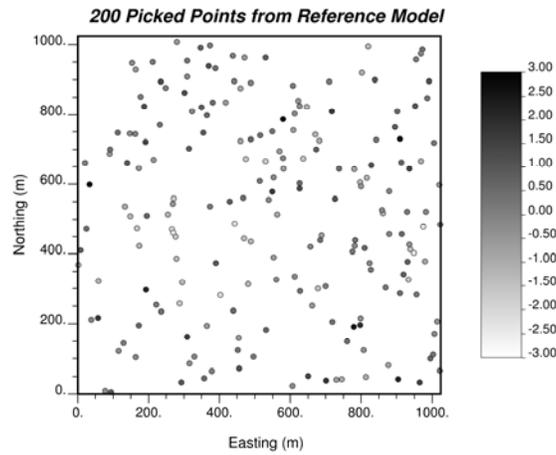**Figure 2:** Variogram reproduction.

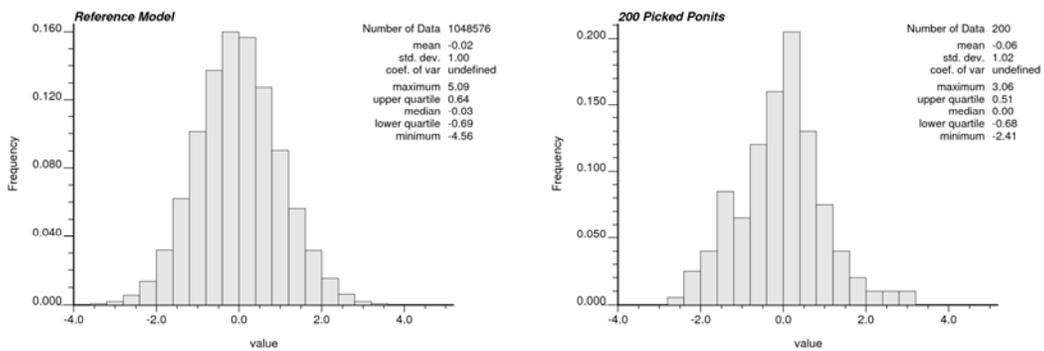**Figure 3:** Locations of 200 randomly picked points from reference model.

Figure 4: Reference Gaussian distribution (left) and the distribution of the 200 randomly picked points (right), the values are in Gaussian units.
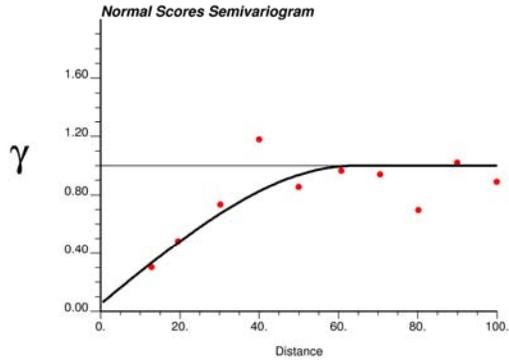
Figure 5: The Reference variogram (solid line) and the experimental variogram points based on 200 picked points (bullets), the distance is in meters (m).
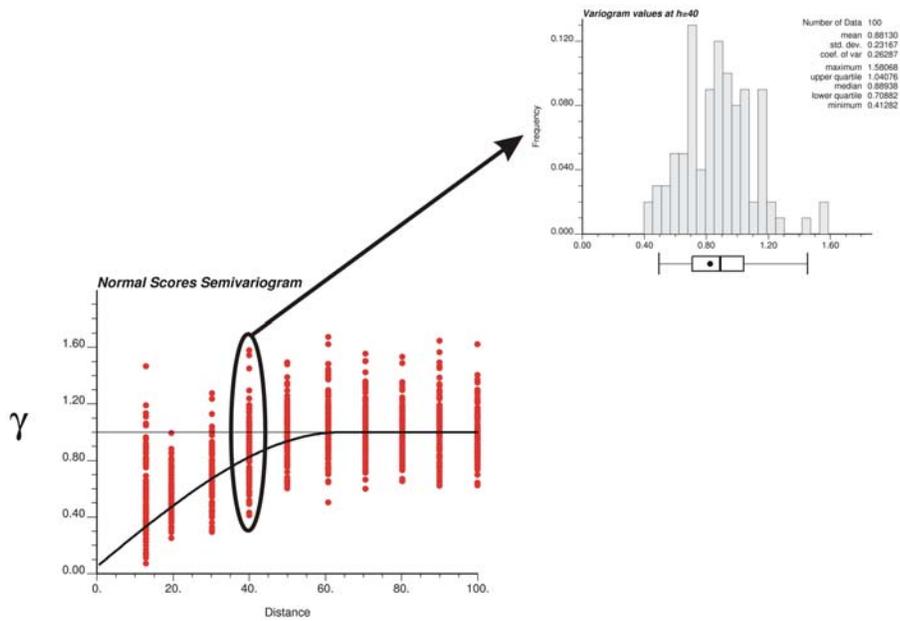

Figure 6: Different realizations for variogram at each lag accompanying by the probability density function at **h**=40.
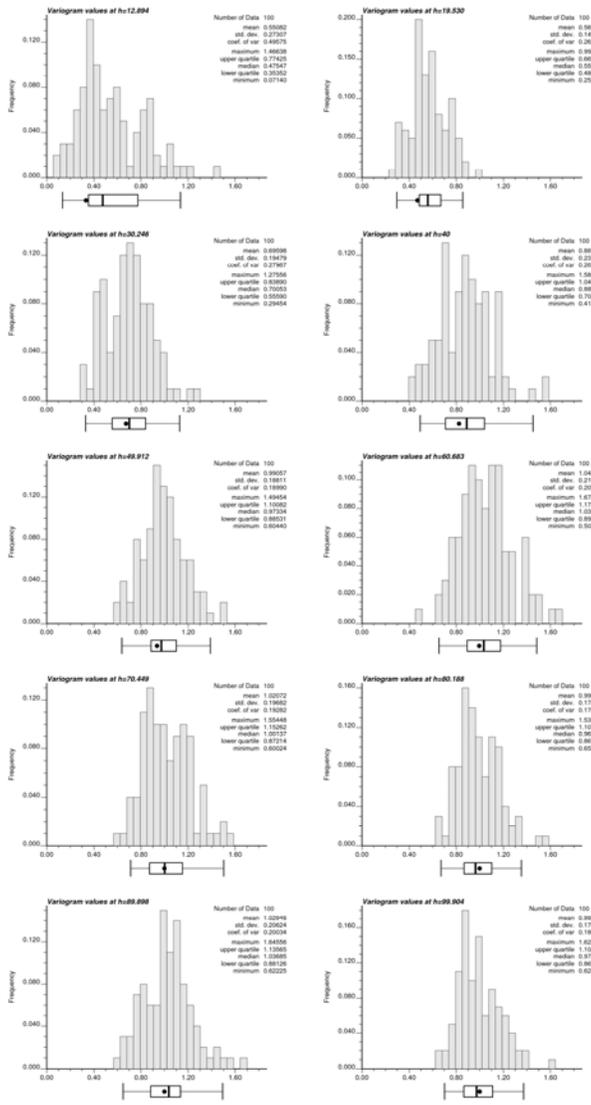
Figure 7: Probability density function of the variogram values at each lag distances (see Figure 6), for each histogram a box plot is shown, the black dot is the reference variogram value.
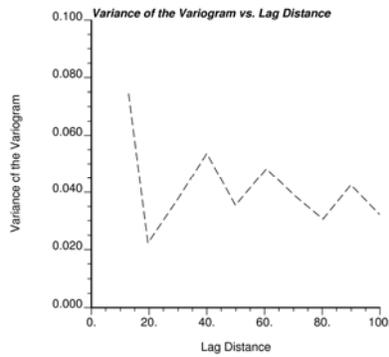
Figure 8: Variance of the variogram versus the lag distance (m).
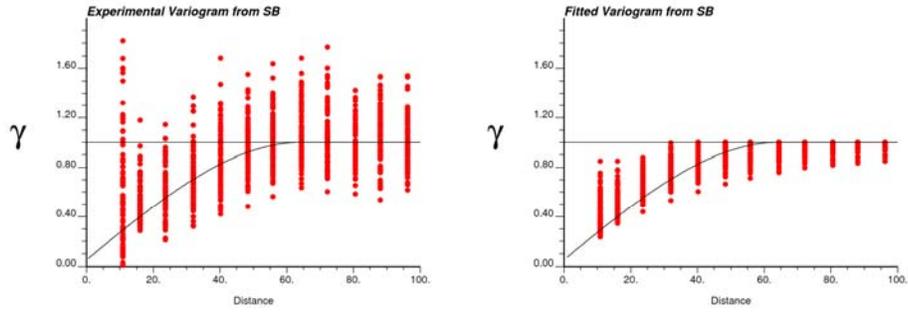
Figure 9: Calculated experimental variogram for 100 realizations from Spatial Bootstrap at *h*=8 and *h*$_{tol}$=4.48 (left) and the corresponding fitted variogram values (right).
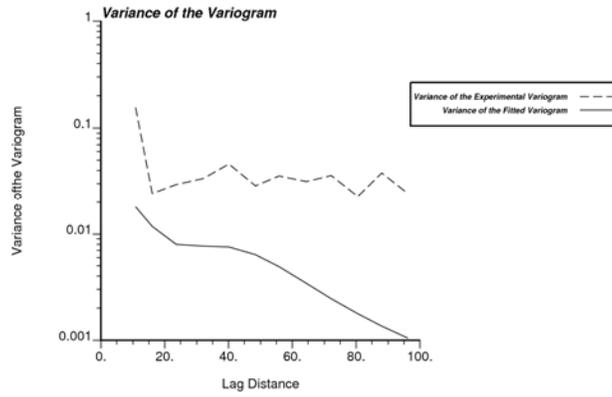
Figure 10: Variance of the experimental variogram (dash line) and the variance of the fitted variogram (solid line) versus the lag distance.
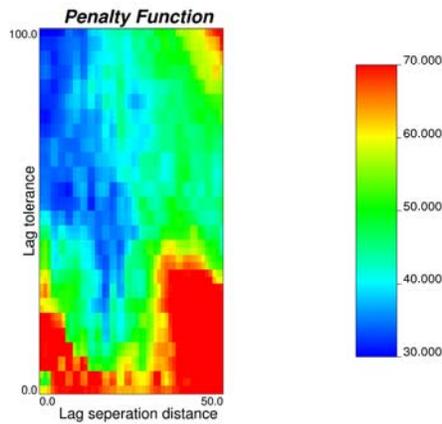
Figure 11: Final penalty as a function of both lag tolerance (m) and lag separation distance (m).
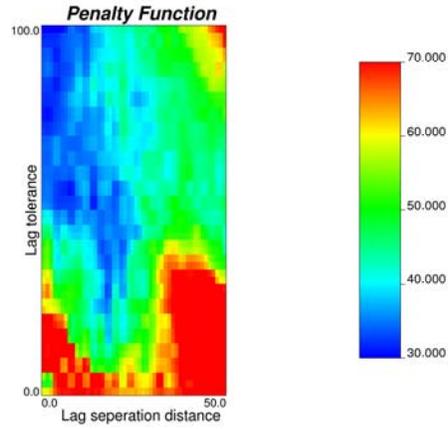
**Figure 12:** Final penalty as a function of both lag tolerance (m) and lag separation distance (m).
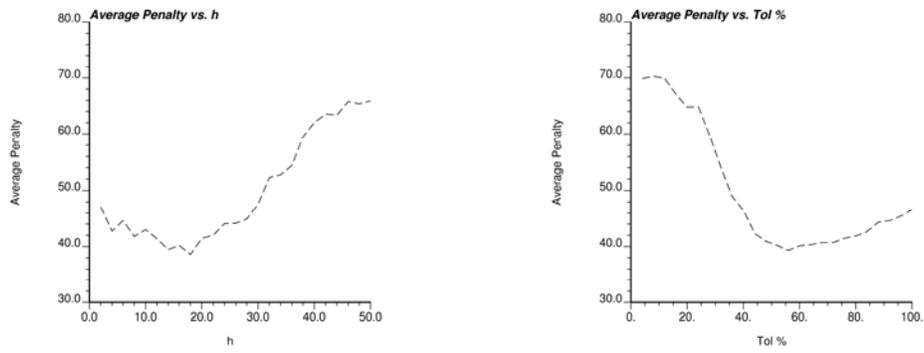
Figure 13: Conditional expectation of the average penalty versus the unit lag distance (left) and the tolerance ratio (right).
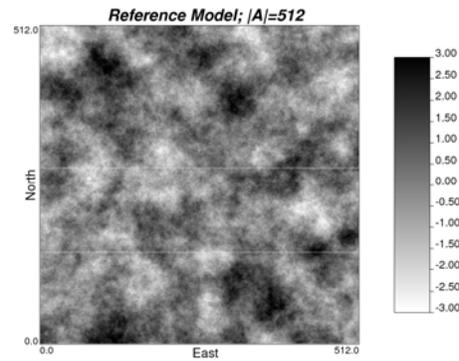
Figure 14: Map of reference model generated by sgsim for field size of 512.
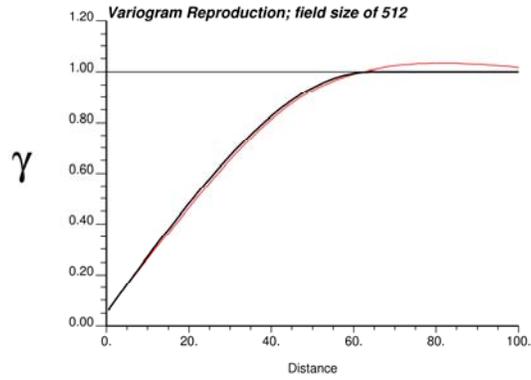
Figure 15: Variogram reproduction check for the reference model with field size of 512, distance is in meters (m).
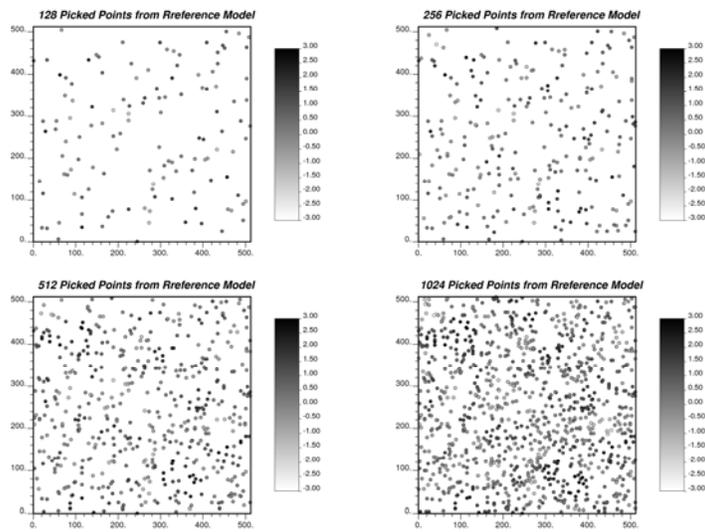
Figure 16: Locations of different number of randomly picked points from reference model with field size of 512.
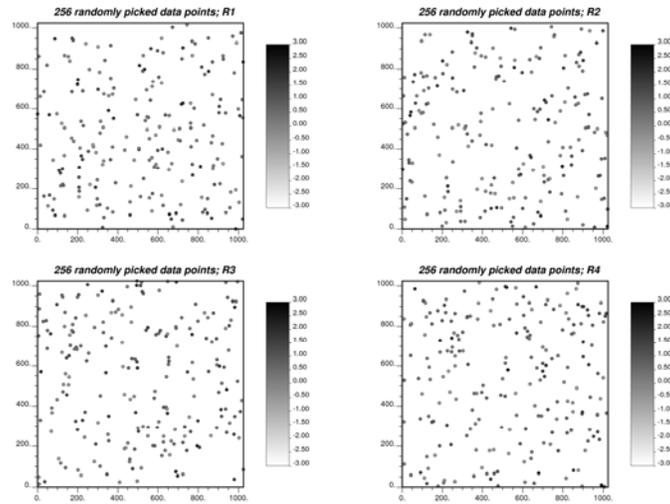
Figure 17: Locations of 256 of randomly picked points for different realizations (256 data locations are different) from the same reference model with field size of 1024.
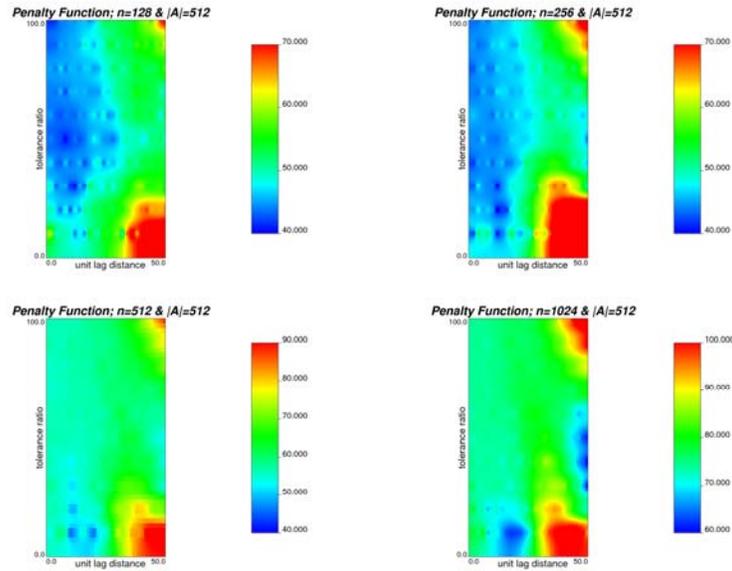
Figure 18: Penalty function in the case of different number of data and a fixed field size.
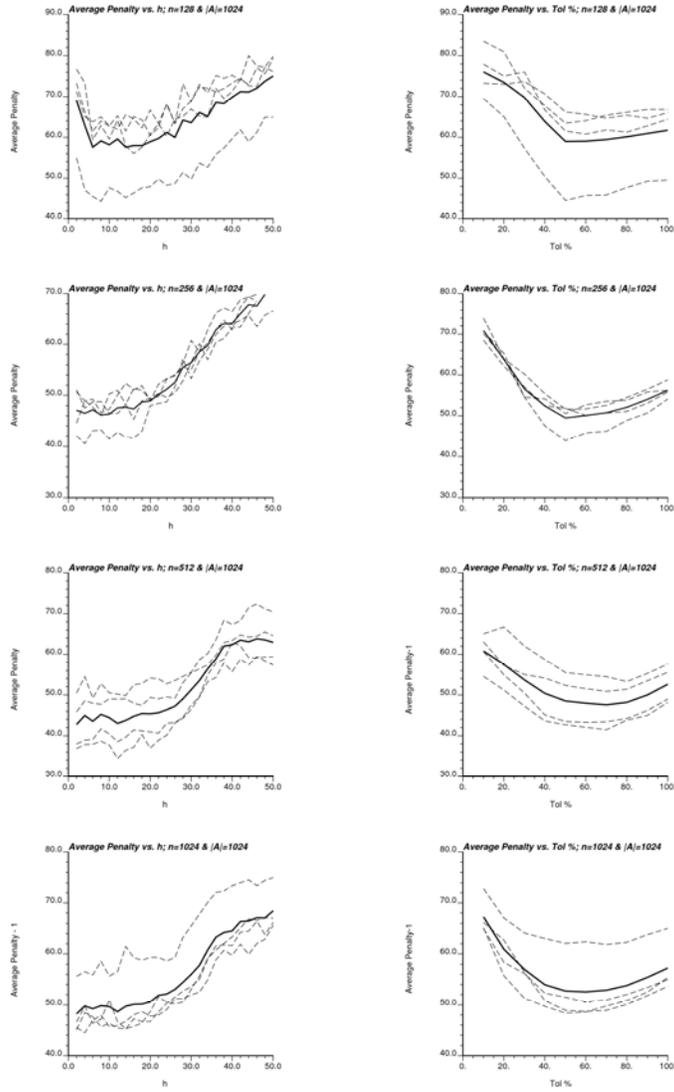
Figure 19: Conditional expectation of the average penalty versus the unit lag distance (left) and the tolerance ratio (right) for different number of picked points and a fixed field size of 1024, the solid line is the average over different realizations (the dashed lines).
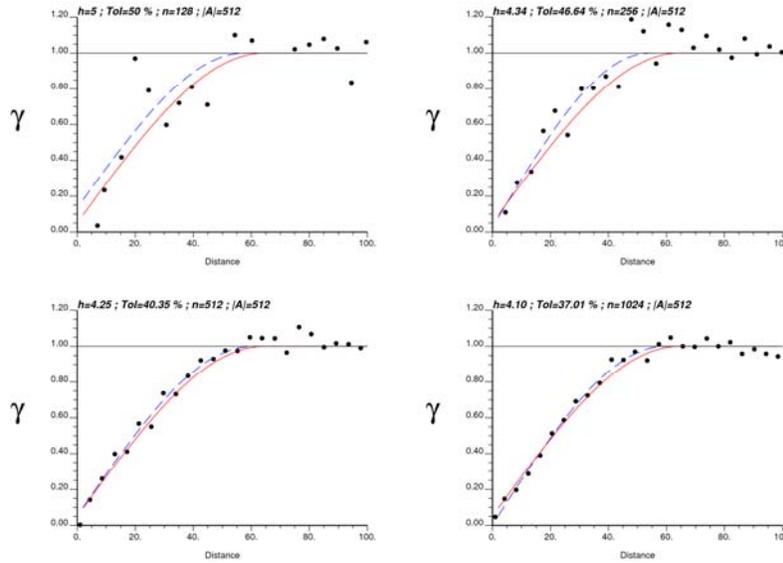
Figure 20: Fitted variogram (dashed line) versus the reference variogram (solid line) in the case of different number of data (*n*=128, 256, 512 and 1024) and a fixed field size ($|A| = 512$) using optimal tolerance parameters. Dots are the experimental variogram points.
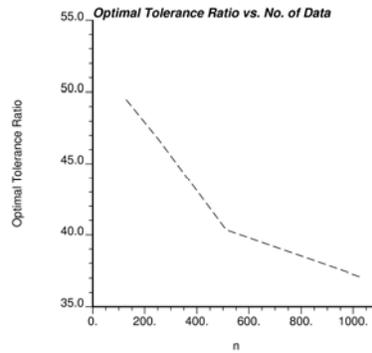
Figure 21: Optimal tolerance ratio versus the number of data.
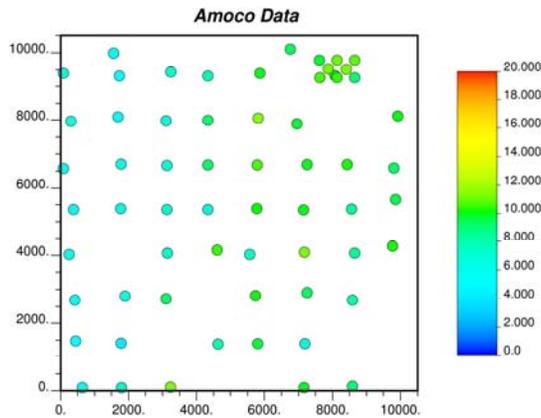
Figure 22: Location map of the 2D porosity data.
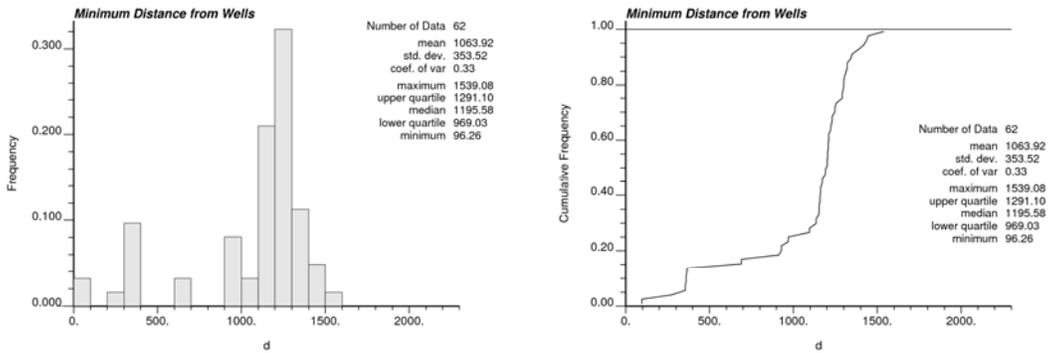
122-14
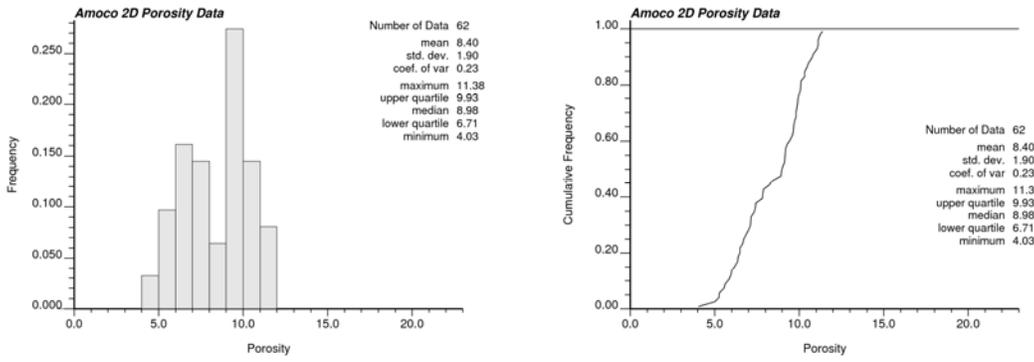
Figure 23: Location map of the 2D porosity data.

Figure 24: Distribution of the minimum distance between wells.
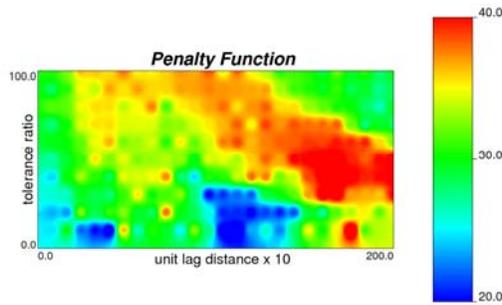

Figure 25: Plot of penalty function as a function of unit lag distance and the tolerance ratio.
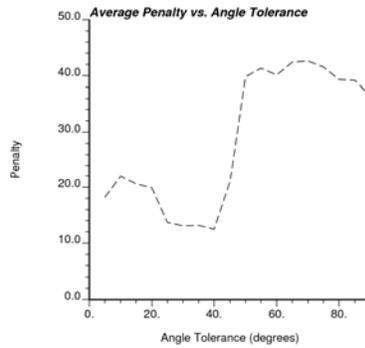

Figure 26: Plot of penalty function as a function of angle tolerance for Amoco 2D data set.
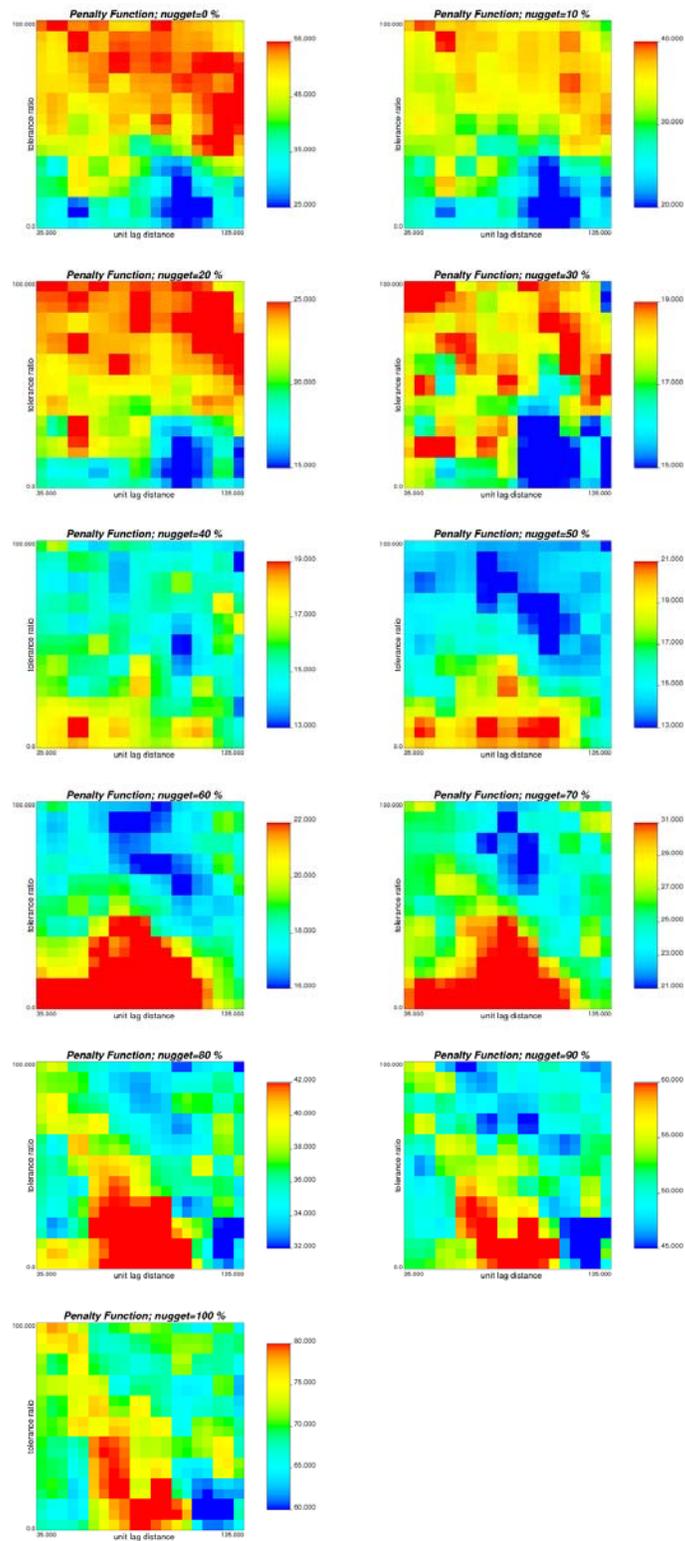
Figure 27: Plot of penalty function as a function of unit lag distance and the tolerance ratio for different nugget effect values for reference variogram model used in penalty function calculation for Amoco 2D data; the reference variogram range is set to be 30 percent of the field size.
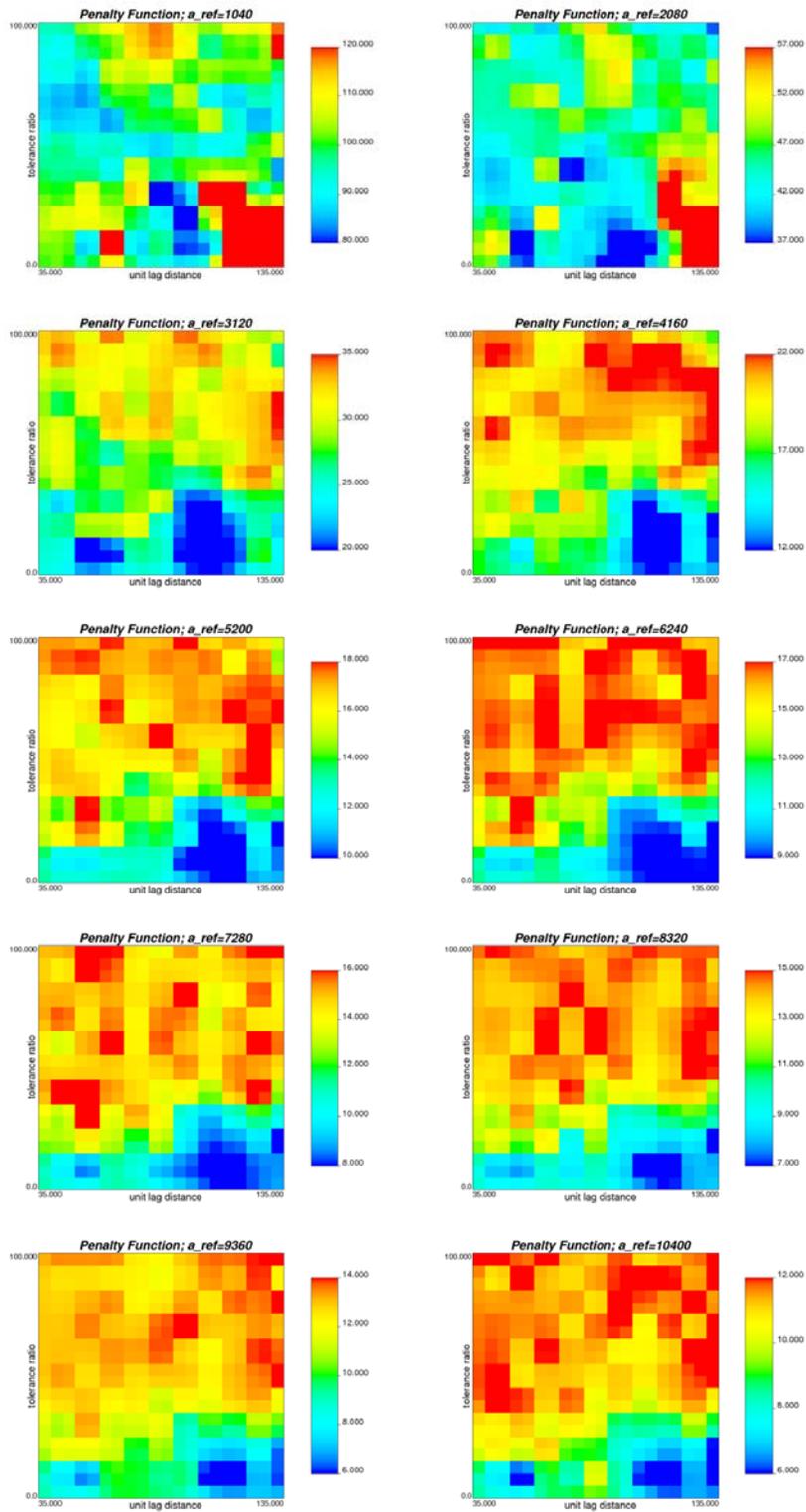
Figure 28: Plot of penalty function as a function of unit lag distance and the tolerance ratio for different reference variogram range for variogram model used in penalty function calculation for Amoco 2D data; the nugget effect is set to be 0 percent.
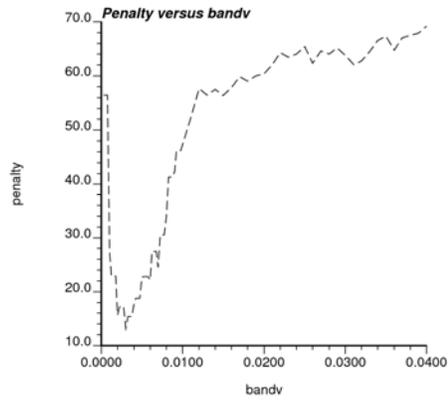
Figure 29: Plot of penalty function as a function of vertical bandwidth for Amoco 3D data set, the vertical bandwidth is a value between 0 and 1 because it is in proportional stratigraphic unit which is a fraction.