Non Stationary Variograms Based on Continuously Varying Weights

David F. Machuca-Mory and Clayton V. Deutsch

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

Location-dependent variograms are calculated for improved estimation. The weighting of the sample pairs in variogram calculation is inversely proportional to their distances to pre-defined points within the study area. The local experimental variograms or correlograms are modeled to yield locally varying parameters. These non stationary variograms are used in non stationary estimation.

Introduction

Conventional application of geostatistics assumes that the variogram depends only of the separation between sample pairs. The delimitation of geologically homogeneous domains and removal of large scale trends facilitates this decision of stationarity. Often, we cannot divide the domain into as many subdivisions as we would like; there are too few data for reliable inference. Choosing a single location invariant variogram model may not be suitable, particularly when the continuity varies smoothly.

A reasonable approach is to calculate and model variograms within sub-regions of the domain. Choosing a window of data for variogram calculation often leads to instability because of limited data. The approach may introduce unwarranted discontinuities. As an alternative, we propose to use all data, but weight the data pairs in variogram calculation. Close data pairs will receive more weight. The exponent of the inverse distance function controls the weights function. The exponent value chosen must be high enough to allow the stable identification of local spatial features; a balance between smoothness and local precision must be found. These weighted variograms are calculated for different locations within the domain. These so-called "anchor points" could be on a regular grid distributed over the study area.

Although each location-dependent variogram could be fit by hand, the variogram fitting is performed automatically in this paper. The resultant set of values of the variogram model parameters, such as nugget effect, ranges and anisotropy orientations can be fitted by polynomial functions in order to provide a continuous model for the parameters at every point of the domain.

The theory for locally weighted variograms, their implementation in the experimental variogram calculation, the modelling of local variograms and the fitting of polynomial functions for describing the variation of variogram parameters all over the domain are developed and presented in this work using synthetic and real data examples.

Distance Weighted Measures of Spatial Continuity

Two alternatives were considered for weighting the pairs involved in the experimental variogram calculation. The first approach consists in calculating the weights inversely proportional to the distance between the anchor point and the middle point of the segment formed by the head and tail samples of a pair separated by a lag distance (see Figure 1a). The weight assigned to each pair:

$$w_{\alpha o} = \frac{1}{\left(\left\|\overline{m_\alpha , o}\right\| + c\right)^p} \tag{1}$$

Were $\left\|\overline{m_\alpha , o}\right\|$ is the distance between the midpoint of the segment formed by the sample pair, α , and an anchor point, o. c is a positive offset value included to smooth the results and avoid dividing by zero, and p is an exponent to control the relative change in weight as distance increases.

The second approach for variogram pairs weighting is to use the weights of the head and tail sample points:

$$w_{\alpha o} = \frac{1}{\left(\left\|\overline{\mathrm{u}_\alpha, o}\right\| + c\right)^p + \left(\left\|\overline{\mathrm{u}_\alpha + \mathrm{h}, o}\right\| + c\right)^p} \tag{2}$$

This approach is similar to using the average of the distances from the pair's endpoints to the anchor point.

Using either weighting scheme, the locally weighted variogram is calculated as:

$$\gamma(o,\mathrm{h}) = \frac{1}{2\sum_{\alpha=1}^{N(\mathrm{h})} w_{\alpha o}} \sum_{\alpha=1}^{N(\mathrm{h})} w_{\alpha o}\left(z(\mathrm{u}_\alpha) - z(\mathrm{u}_\alpha + \mathrm{h})\right)^2 \tag{3}$$

It is clear that if the power *p* is zero, this expression becomes the traditional semivariogram measure, where all sample pairs have the same importance in the calculation regardless of their position. Under the form above, the sum of weights is constrainted to one, thus the locally weighted experimental variogram and the other measures presented below are valid (Chiles and Delfiner, 1999).

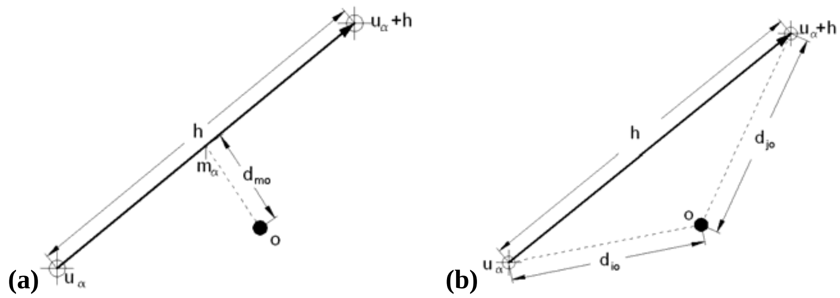


Figure 1: (a) Pair distance to the anchor point relative to the lag midpoint and (b) Pair distance to the anchor point relative to the lag endpoint.

Similarly, the variance of distance weighted pairs can be calculated as:

$$\begin{aligned}\sigma^2(o) &= \frac{1}{2\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ijo}} \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ijo}\left(z_i - z_j\right)^2 \\ &= \frac{1}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ijo}}\left(\frac{1}{2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ijo} z_i^2 + \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ijo} z_j^2\right) - \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ijo} z_i z_j\right)\end{aligned} \tag{4}$$

This variance is used for calculating the standardized weighted semivariogram. Since the weights depend of the distance between the midpoint of the segment formed by a sample pair and the anchor point, a single weight cannot be assigned to individual samples, but *n* different weights are associated to each sample, they correspond to the *n* pairs that can be formed by a given sample and the other samples included itself. Thus the distance weighted mean becomes:

$$m(o) = \frac{1}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} z_i \tag{5}$$

Then, the non stationary locally weighted covariance is expressed as:

$$C(o,\mathrm{h}) = \frac{1}{\sum_{\alpha=1}^{N(\mathrm{h})} w_{\alpha o}} \sum_{\alpha=1}^{N(\mathrm{h})} w_{\alpha o} z(\mathrm{u}_\alpha) \cdot z(\mathrm{u}_\alpha + \mathrm{h}) - m_{-\mathrm{h}}(o) \cdot m_{+\mathrm{h}}(o) \tag{6}$$

With:

$$m_{-\mathrm{h}}(o) = \frac{1}{\sum_{\alpha=1}^{N(\mathrm{h})} w_{\alpha o}} \sum_{\alpha=1}^{N(\mathrm{h})} w_{\alpha o} \cdot z(\mathrm{u}_\alpha) \qquad m_{+\mathrm{h}}(o) = \frac{1}{\sum_{\alpha=1}^{N(\mathrm{h})} w_{\alpha o}} \sum_{\alpha=1}^{N(\mathrm{h})} w_{\alpha o} \cdot z(\mathrm{u}_\alpha + \mathrm{h}) \tag{7}$$

Similarly, the distance weighted correlogram can be calculated as:

$$\rho(o,\mathrm{h}) = \frac{C(o,\mathrm{h})}{\sqrt{\sigma^2_{-\mathrm{h}}(o) \cdot \sigma^2_{+\mathrm{h}}(o)}} \tag{8}$$

With

$$\sigma^2_{-h}(o) = \frac{1}{\sum_{\alpha=1}^{N(\mathrm{h})} w_{\alpha o}} \sum_{\alpha=1}^{N(\mathrm{h})} w_{\alpha o}[z(\mathrm{u}_\alpha) - m_{-\mathrm{h}}(o)]^2 \quad \sigma^2_{+h}(o) = \frac{1}{\sum_{\alpha=1}^{N(\mathrm{h})} w_{\alpha o}} \sum_{\alpha=1}^{N(\mathrm{h})} w_{\alpha o}[z(\mathrm{u}_\alpha + \mathrm{h}) - m_{+\mathrm{h}}(o)]^2$$

Other distance weighted measures of spatial continuity including relative variograms, variograms of transformed data and cross variograms are straightforward to derive.

Programs Implementation

The calculation of experimental distance weighted variograms was implemented in a modified version of the FORTRAN program `GAMV2004`, called `GAMV-Local`. The anchor points can be on a 2 or 3-D grid or with locations specified in a separate anchor point file. This new program calculates the measures of spatial variability with the weighting of the sample pairs according to the distance weighting schemes described above. The parameter file of this program is presented in Figure 2.

```
                Parameters for GAMVLOCAL
                ************************

START OF PARAMETERS:
../data/cluster.dat              - file with data
1   2   0                        - columns for X, Y, Z coordinates
2   3   4                        - number of variables,column numbers
-1.0e21     1.0e21               - trimming limits
0                                - anchor points in file = 0, in grid = 1
50   0.5     1.0                 - nx,xmn,xsiz
50   0.5     1.0                 - ny,ymn,ysiz
1    0.0     1.0                 - nz,zmn,zsiz
apoints.dat                      - file with anchor points location
1   2   3                        - columns for X, Y, Z coordinates
1   1                            - Power and offset for distance weighting
2                                - Weighting schema: 1=midpoint, 2=endpoints, 3=both
locgamv.out                      - file for variogram output
locgamv.dbg                      - file for debugging
3                                - number of directions
  0. 90. 9999.    0. 90. 9999.   - Dir 01: azm,atol,bandh,dip,dtol,bandv
10  5.0  3.0                     -          nlag,xlag,xtol
  0. 20. 9999.    0. 90. 9999.   -Dir 02: azm,atol,bandh,dip,dtol,bandv
 8  7.0  4.0                     -         nlag,xlag,xtol
 90. 20. 9999.    0. 90. 9999.   -Dir 03: azm,atol,bandh,dip,dtol,bandv
 8  7.0  4.0                     -         nlag,xlag,xtol
1                                - standardize sills? (0=no, 1=yes)
3                                - number of variograms
1   1   1                        - tail var., head var., variogram type
2   2   1                        - tail var., head var., variogram type
1   1   9   1.0                  - tail var., head var., variogram type


type 1 = traditional semivariogram
     2 = traditional cross semivariogram
     3 = covariance   (-3 calculates variance-covariance)
     4 = correlogram  (-4 calculates 1-correlation)
```

Figure 2: Parameters for GAMV-Local

The output file of `GAMV-Local` is similar to that of `GAMV2004`, with the exception that the header of the first variogram of a set of variograms calculated for a single anchor point contains the coordinates of this point (see Figure 3).

```
-Coordinates:    5.0000000       5.0000000       0.0000000
Semivariogram           tail:value
-HDIR     0.0000000       5.0000000       50.000000
-VDIR     0.0000000       10.000000       50.000000
-LAGS            30       1.0000000       0.50000000
-VARI             1      1      1
```

Figure 3: Format of the GAMV-Local output file

Since the number of anchor points used for the calculation of locally weighted experimental variograms can be considerable, the manual fitting of them is not always practical and thus, we rely mostly on automatic fitting. For such purpose, the program `VARFIT` was modified in order to recursively read the experimental variograms from the `GAMV-Local` and fit the models corresponding to each location. `VARFIT-loc` writes the parameters of the fitted variogram models, as well as the coordinates of their corresponding anchor points, in a summary file of GSLIB format that can be used for plotting the variation of the variogram model parameters across the domain and for fitting deterministic polynomial functions. The only modifications in the parameter file for `VARFIT-loc` are the inclusion of a line for the number of anchor points to be considered, and a line to specify the name of the summary file.

Synthetic Data Example 1

The synthetic image for the first example consists of two zones was created using Sequential Gaussian Simulation; the distribution of simulated values is Gaussian with the same mean (0) and variance (1) for both zones. The parameters of the variogram used for generating the two regions are also similar: nugget effect equal 0.1, and a spherical variogram model with major and minor ranges equal to 20m and 5, respectively. The only difference between the zones is the orientation of the major anisotropy axes, 0° in the west, and 90° in the east (see Figure 5). A transition zone was created between both regions by simulating the second region with data from the first.

The anchor points are located on a 10mx10m grid arrangement, the program `GAMV-Local` was used to calculate the experimental variograms in the Az. 0° and Az. 90° directions for each anchor point and for different values of the exponent p of the inverse distance weights. Then, `VARFIT-loc` was used for fitting the variograms for the 220 anchor points. The variogram model was locked to a spherical model.

Figures 6, 8 and 9 present the nugget effect and the major and minor horizontal range parameter for the fitted spherical variogram models. The nugget effect decreases slightly as the power progress from 0 to 1 approaching the true variogram nugget effect. The nugget effect becomes increasingly variable from one anchor point location to another as the power exceeds 2 (see Figure 6).

In the same way, the major and minor ranges of the fitted variogram models approach the true variogram ranges for each domain as the power factor grows from 0 to 1, but the range values become increasingly noisy as the power factor is increased beyond a value of 2 (Figure 7).

In order to assess the most suitable values for the power and the offset parameter a measure of optimality was designed. This measure compares the local experimental or modeled variogram to the reference variogram:

$$Op(\mathrm{h}) = \frac{1}{n_{ap}} \sum_{n_{ap}} \left| \gamma_{\text{local true}}(\mathrm{h}) - \gamma_{\text{distance weighted}}(\mathrm{h}) \right| \tag{9}$$

Where n_{ap} is the number of anchor points and $\gamma_{\text{local true}}(\mathrm{h})$ and $\gamma_{\text{distance weighted}}(\mathrm{h})$ can be either the experimental variograms or fitted models. Figure 8 shows the optimality measure for experimental variograms at different power values but a fixed offset value of 0.1. As the power is increased from 0 to 1, the distance weighted variograms approach the local true variograms minimising the lag optimality measure when this factor is 1. The variograms become increasingly noisy for higher values of the power.

The fitted models (Figure 9) also indicate that a power factor of around 1 is optimal. At lag distances greater than the variogram range the lag optimality measure decreases if the fitted experimental variograms were standardized. For the shortest lags, the distance weighted variograms with a power factor of 2 can be slightly more accurate than those calculated using a factor of 1 (Figures 8 and 9). The global measure of optimality (Figure 10) shows a clear improvement for the experimental variograms when a power of 1 is used in the pairs weighting function.

Synthetic Data Example 2

Three zones were created using sequential Gaussian simulation with neighbour data from the adjacent zone, and subsequently back transforming the simulated result to three different lognormal distributions. The central zone has a high mean (2.02) and variance (3.42) with N-S preferential continuity, while the West and East zones have lower means and variances, thus, the West zone has a mean of 0.98, a variance of 1.96

and an Azimuth of 135° as direction of major continuity, and the East zone has a mean of 0.55 and a variance of 0.92 with a E-W direction of major continuity (see figure 11). The nugget was different for each of the three zones: 30%, 40% and 15% for the West, Central and East zones, respectively.

The area covers 800 x 800 pixels and was sampled in a quasi random grid of 10x10 pixels, and the anchor points were arranged in a regular 50x50 pixels grid. Both weighting approaches were tested with this data set using two different configurations of the power and offset parameter for inverse distance weighting.

Figure 12 show the local means and variances calculated using inverse distance weighting with a power of 1 and an offset of 1 pixel and using three different weighting schemas: (1) distance of the individual samples to the anchor points, (2) distances from the pair's midpoints to the anchor points (see expression 1), and (3) combined distances of the pair's endpoints to the anchor points (see Expression 2). Calculations were also performed with a power of 2 and an offset of 5 pixels; the results were similar in character, but somewhat noisier. When the correlogram pairs are weighted in relation to their midpoint, the influence of the central high grade zone is extended excessively in the East and West low grade areas. Moreover, this weighting schema introduces other artifacts in the variance maps.

When the pairs are weighted by the sum of the endpoints distance to the anchor point, the local mean and variance are more consistent with the mean and variance obtained weighting by the inverse distance of individual samples and with spatial distribution of local data (see Figure 11, bottom). These results indicate that the endpoints distance weighing scheme is more appropriate. Local correlograms were also calculated. The results with a power of 1 appear smooth and reasonable. The directions of continuity are fit well. The results are similar to the local variogram results.

Real Data Example

The clustered sampling of Walker Lake (Isaaks and Srivastava, 1989) and a quasi regular 10m x 10m grid sampled from the exhaustive Walker data are considered. The variables in the data sets present different domains and changing orientations of anisotropy as can be seen in Figures 13 a and b. The results obtained using a power of 2 and an offset of 5m are presented because they appear better than the results obtained with a power of 1. The results with a power of 1 appear too smooth.

Figure 14 shows, for the clustered data set, the local mean and variance calculated by weighting individually the samples and by weighting the sample pairs. The pair endpoints weighting produces better mean and variance maps than the individual sample weighting. The high concentration of samples in the high valued area influences the surrounding regions. A proportional effect is observed for both datasets.

The experimental local weighted variograms of both datasets were fitted using a single spherical structure, Figures 15 shows the fitted variogram model parameters of the clustered data. The spatial distribution of the local variogram and correlogram parameters are similar. The correlogram and variogram parameters of the models for the quasi regular grid data show a sharper spatial structure than those of the clustered data and a better definition of the local data features, particularly in the major values continuity orientation (Figures 16).

Discussion

A distance weighting exponent between 0.5 and 2 works well, a lower value produces smooth local experimental variograms with results approaching the traditional variogram. A higher exponent increases local precision of the local variograms at the cost of increased noise. Weighting the variograms by the endpoint distances to the anchor points leads to the best results. Correlograms account for the local head and tail means and variances at different lag distances; thus, they appear more robust and easier to interpret in this locally stationary framework. A single structure leads to stable results. If two structures are used, the parameters of the long range structure appear quite variable. Although the variogram shape could change from one region to another, there could be unwarranted variation in the model type if fit automatically. The parameters of the models fitted to the local weighted experimental variograms at each anchor point could be interpolated to provide the location-dependent parameters required at every location for non stationary or quasi stationary estimation. These parameter maps should show geologically reasonable variations.

Conclusions

Locally varying variograms are able of provide a more accurate description of the spatial behaviour of a non-stationary variable than that provided by the traditional measures of spatial variability. Weighting the pairs by inverse distance with a power between 1 and 2 appears to work well. Correlograms are preferred over variograms since they are insensitive to local means and variances. Local variations in the nugget effect and variogram parameters could have a significant affect on spatial prediction. Further research is required for understanding the behaviour and interpretation of the non stationary measures of spatial continuity, as well as for testing other pairs weighting schemas. The use of the non stationary variogram parameters in estimation will be analysed, tested and discussed in another CCG paper.

References

Chilès J.P. & Delfiner, P., 1999. *Geostatistics: modeling spatial uncertainty*. Wiley-Interscience, New York.

Deutsch, C.V., and Journel, A.G., 1998). *GSLIB: Geostatistical Software Library and User's Guide*. 2nd Edition, Oxford University Press, New York.

Goovaerts, P., 1997. *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York.

Isaaks, E. H. and Srivastava R. M., 1989. *An introduction to Applied Geostatistics*. Oxford University Press, New York.

Ren, W. and Deutsch, C.V., 2006, *Bayesian Updating with local Varying Correlation*. Paper 104, Eight Annual Report of the Centre for Computational Geostatistics, September 2006.

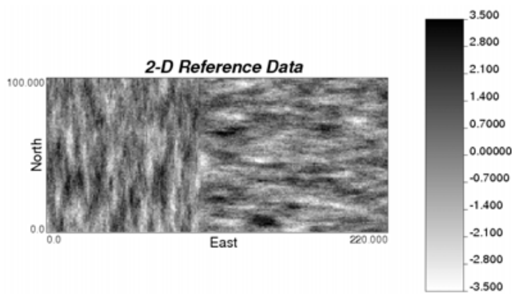


Figure 4. Synthetic example 1: Two zones with different major anisotropy axis orientation.

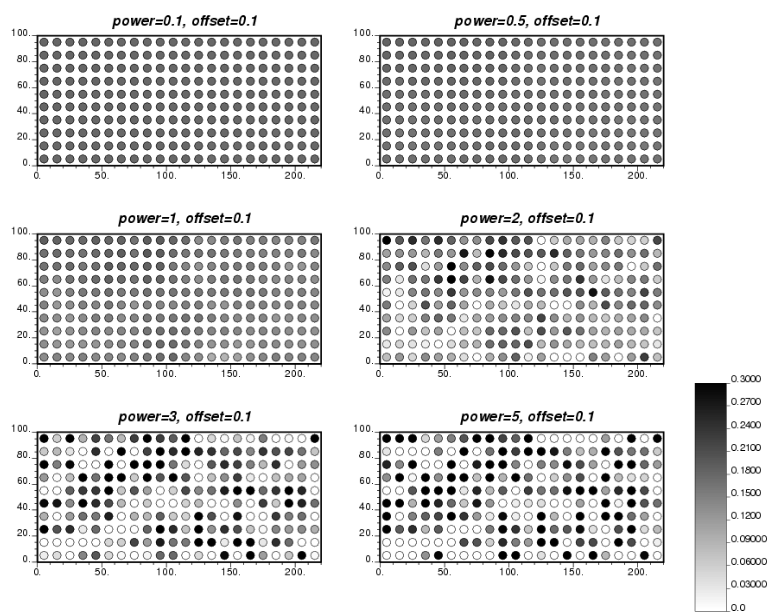


Figure 5. Synthetic example 1: Change in the local nugget effect relative to the increasing power of the weighting function.

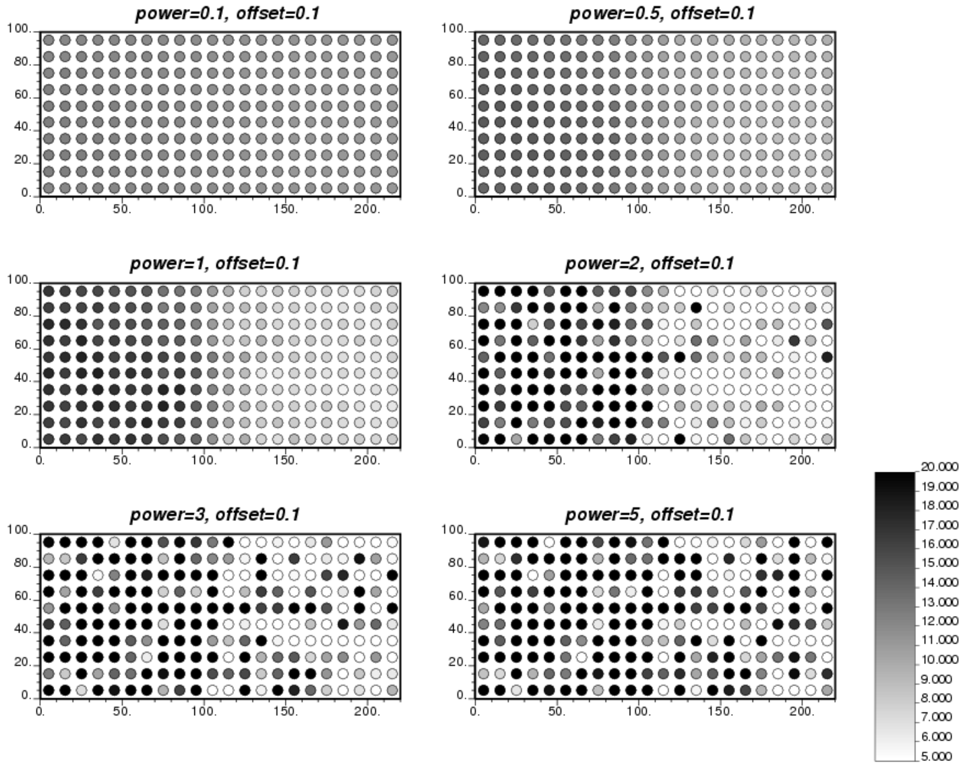


Figure 6: Synthetic example 1: Change in the N-S range relative to the increasing power of the weighting function.

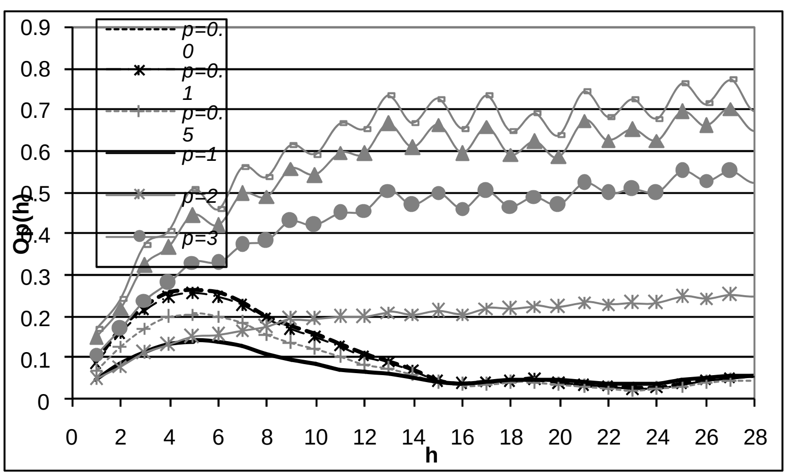


Figure 7: Absolute differences between local true and distance weighted experimental variograms at different lag distances.

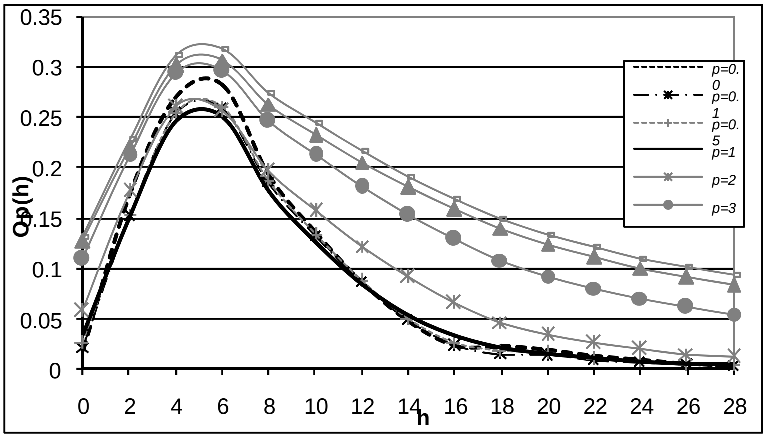


Figure 8: Absolute differences between the models of the local true and the distance weighted variograms at different lag distances.

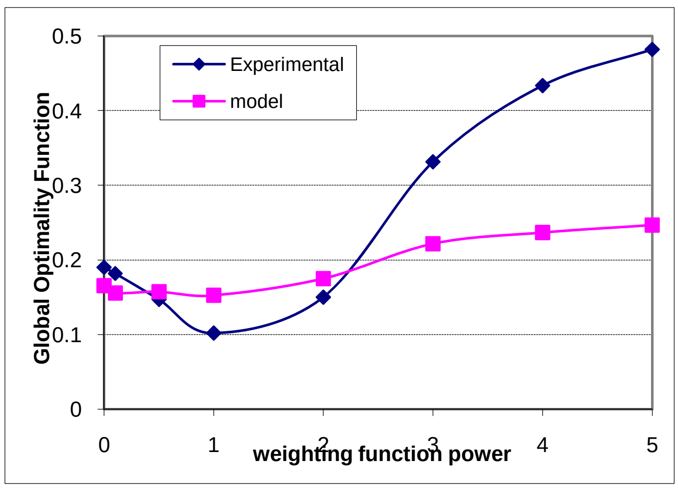


Figure 9: Global absolute differences between true and weighted local variogram.

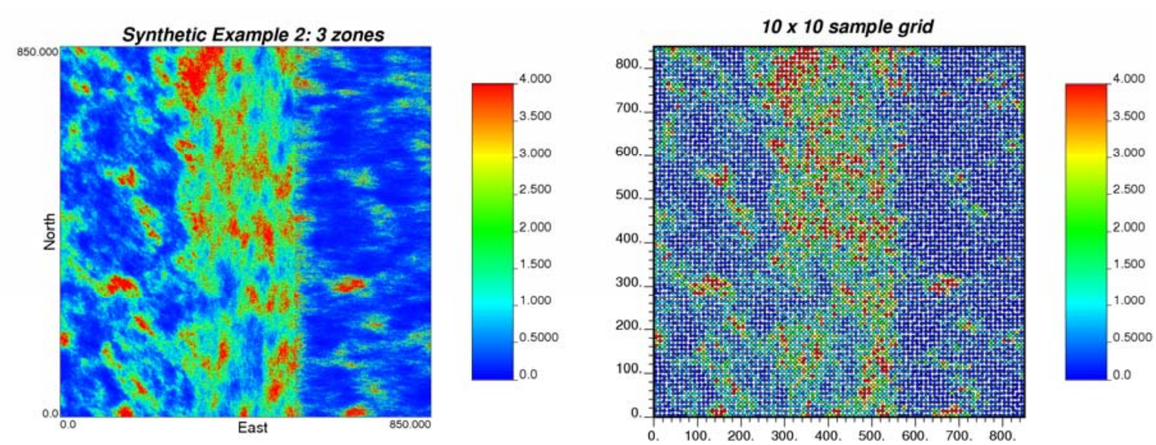


Figure 10: synthetic image and 10x 10 pixels sample grid used in the second example.

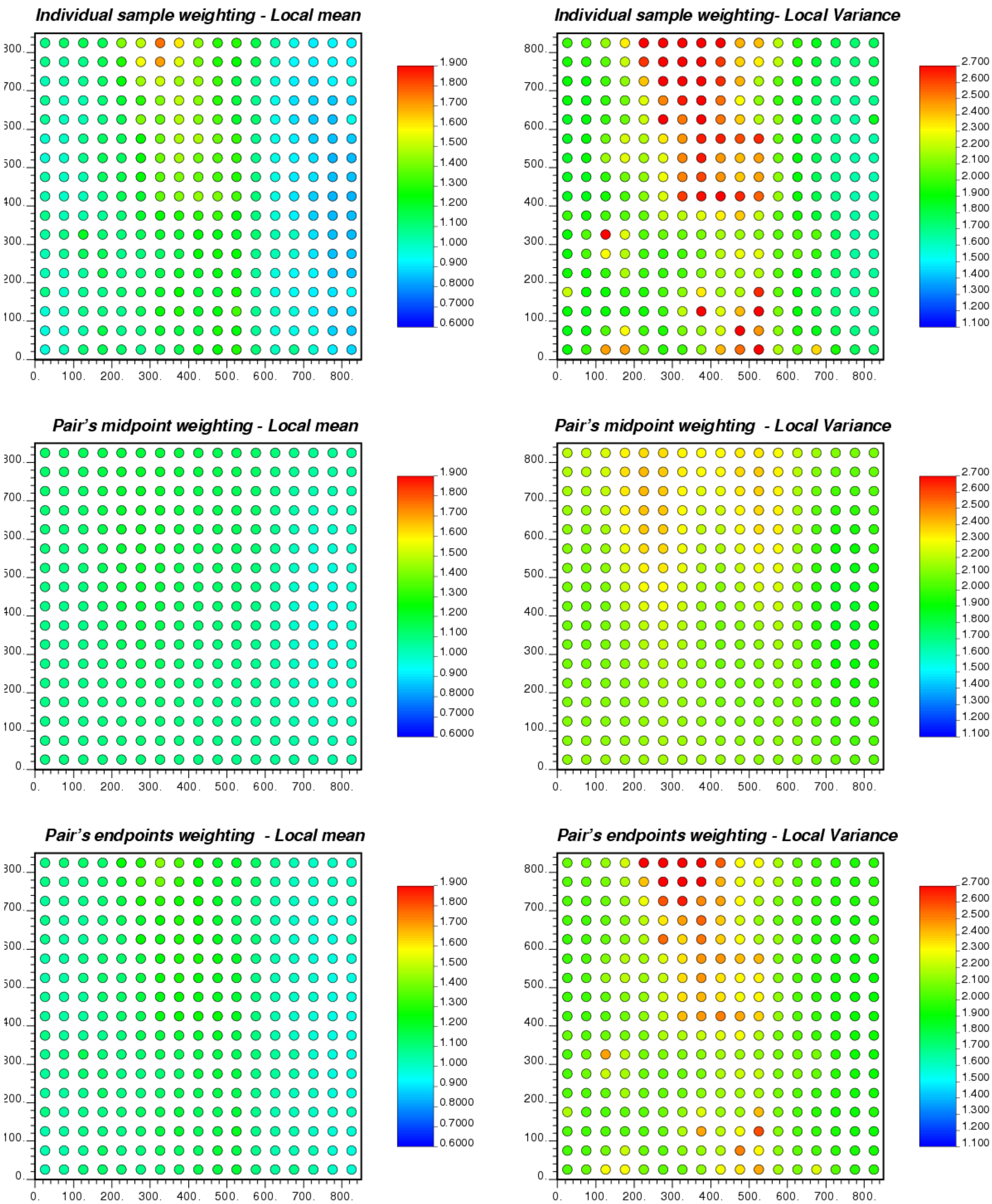
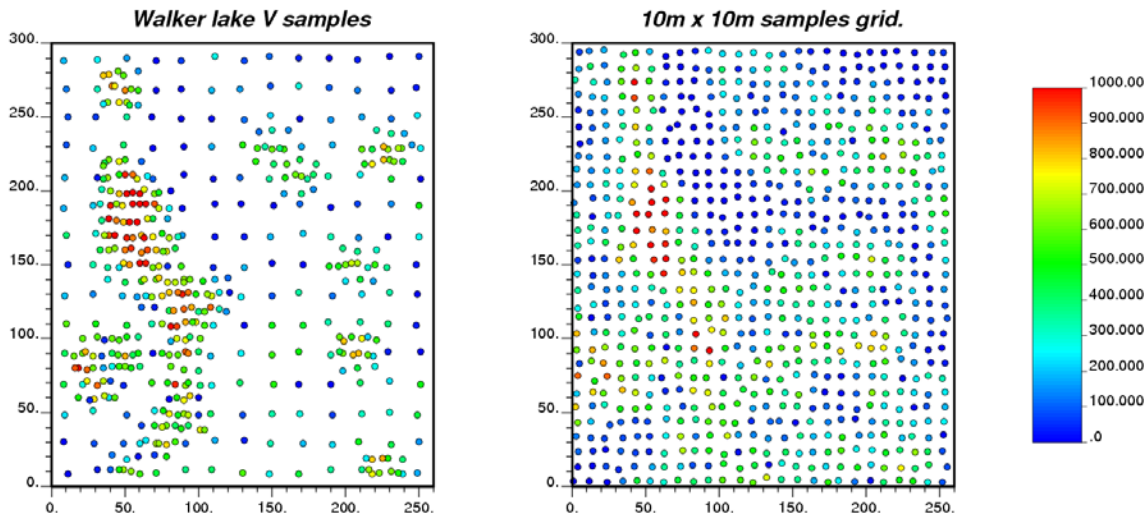


Figure 11: Local mean and variances calculated using inverse distance weighting of individual samples (top), of pair's midpoints (centre) and of pair's endpoints (bottom). A power of 1 was used for the inverse distance weighting function.



(a) **(b)**

Figure 12: Clustered (a) and 10m x 10m samples grid (b) from the walker data set.

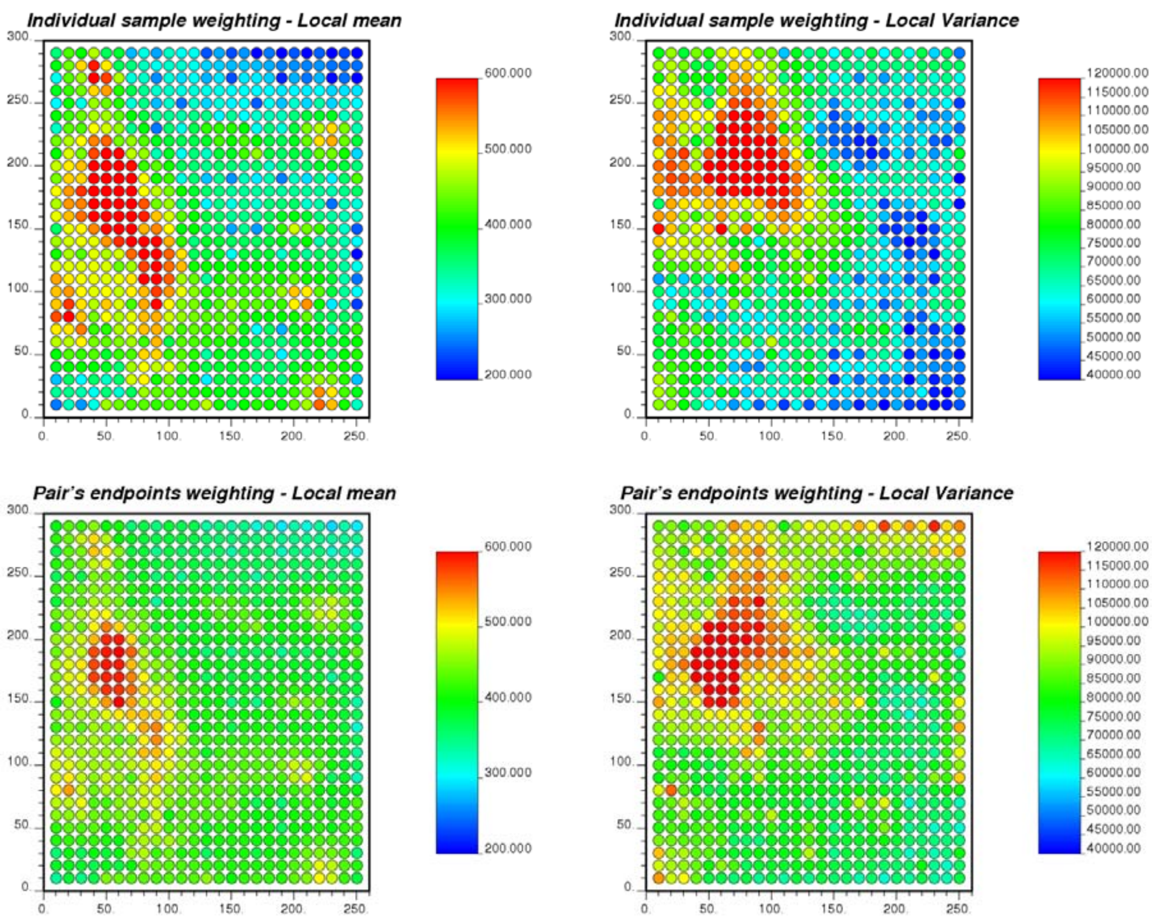


Figure 13: Local means and variances calculated for the clustered data set by weighting the distance to individual samples (top) and by weighting by the distances to the pair's endpoints.

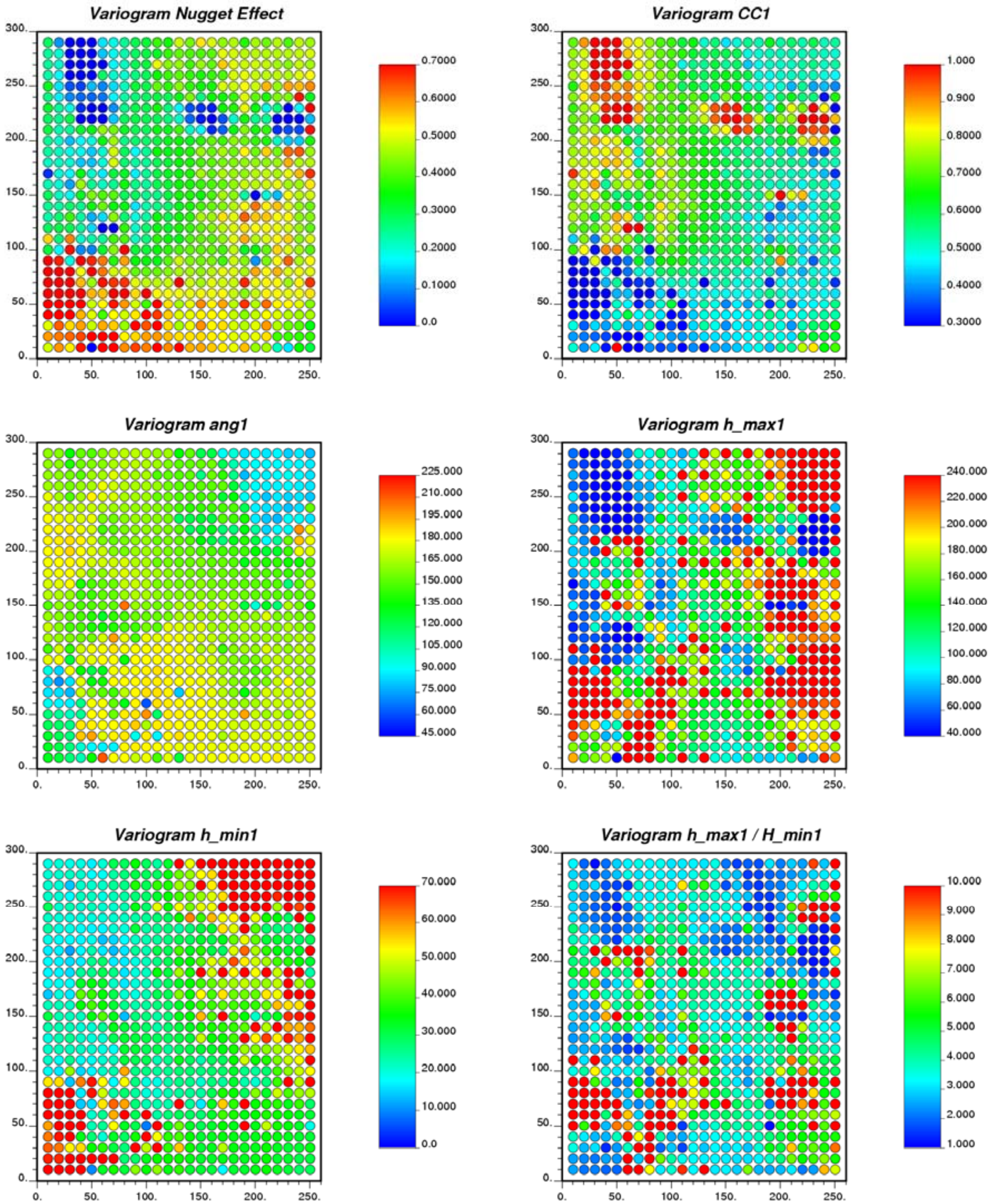


Figure 14: Local parameters for the model fitted to the non stationary variograms calculated from the clustered data set.

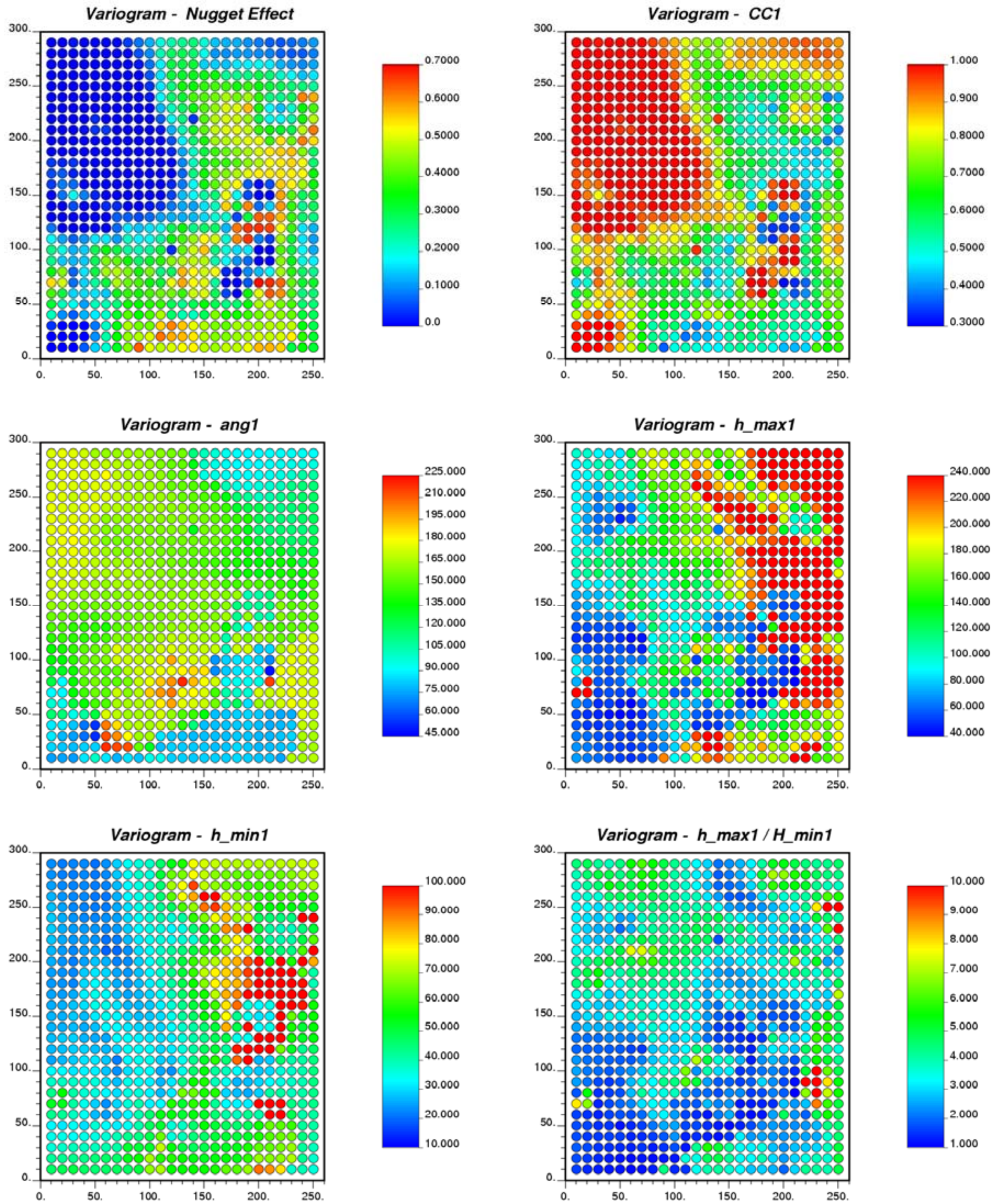


Figure 16: Local parameters for the model fitted to the non stationary variograms calculated from the 10m x 10m sample grid.