Multivariate direct block simulation using a modified Blusim algorithm

Arja Jewbali (arja.jewbali@riotinto.com)
Resource Estimation Geologist
Rio Tinto Iron Ore

The Rio Tinto iron ore group currently estimates chemical variable grade in mineral resources using Ordinary Kriging (OK.) OK produces a smoothed globally unbiased estimate, but does not capture local variability which can have significant impact on mine planning and scheduling. The dimensions and the multivariate nature of iron ore deposits make point simulations of whole deposits impractical. With a direct block simulation method the size and speed limitations of point simulations can potentially be overcome, resulting in a viable method for capturing local variability and for providing risk measures into the mine plan and schedules. Blusim is an algorithm developed at the CCG for assessing uncertainty of recoverable reserves at block resolution, however Blusim does not reproduce the spatial grade relationships between blocks and as such cannot be used to assess uncertainty of quarterly schedules (which contain multiple blocks) or assess stockpile variability . This paper describes the enhancements made to the Blusim algorithm so that the correct spatial relationships between blocks are reproduced in realizations. The Walker Lake data set is used to demonstrate the enhanced Blusim algorithm.

Introduction

Blusim (Ortiz and Deutsch, 2007) is an algorithm developed at the CCG for assessing uncertainty of recoverable reserves at a block scale. The algorithm accommodates multiple correlated variables through the use of a linear model of coregionalization or LMC (Boisvert and Deutsch, 2007). The Blusim algorithm discretizes the blocks into points and considers all blocks to be independent. For each block, LU simulation is used to generate realizations on the points, which areaveraged to get the block grade. Realizations are then discarded. Because the blocks are independent, thespatial relationship between the blocks is not reproduced. However within a block the spatial relationships between the discretization points are reproduced. The Blusim algorithm proceeds as follows (Ortiz and Deutsch, 2007):

1. Transform the data to normal scores (standard Gaussian distribution);
2. Fit the LMC using the normal scores data;
3. Select the first block and build the covariance matrix (a random path is not required);
4. Decompose the covariance matrix in a lower and upper triangular form (Cholesky decomposition: Equation 1);

$$\begin{bmatrix} C_{11} & C_{13} \\ C_{31} & C_{33} \end{bmatrix} = \begin{bmatrix} L_{11} & \\ L_{31} & L_{33} \end{bmatrix} \begin{bmatrix} U_{11} & U_{13} \\ & U_{33} \end{bmatrix} \qquad \text{Equation 1}$$

Where,

- C_{11} is the matrix with the covariances between the data locations.
- C_{13} is the matrix with the covariances between the data locations and the discretized points inside the block.
- C_{33} is the matrix with the covariances between the discretized points inside the block.

5. The simulated values at each discretized location inside the block are generated by solving the system in Equation 2;

$$\begin{bmatrix} L_{11} & \\ L_{31} & L_{33} \end{bmatrix} \begin{bmatrix} W_1 \\ W_3 \end{bmatrix} = \begin{bmatrix} Z_1 \\ Z_3 \end{bmatrix} \qquad \text{Equation 2}$$

Where,

- $W_1 = L_{11}^{-1} Z_1$ is a vector of weights that are solved for.
- Z_1 is a vector containing the conditioning data values
- W_3 is a vector of standard normal random numbers
- Z_3 is a vector that will contain the simulated values for every discretized location inside the block.

6. The simulated values at the discretized locations are back-transformed and averaged to get the simulated block value;
7. If more simulated values are required, a new vector of random numbers W_3 is generated;
8. Move on to the next block (step 3).

A modified Blusim algorithm

This modification of the Blusim algorithm is based on Godoy (2003) and Boucher and Dimitrakopoulos (2009) and involves the incorporation of previously simulated blocks into the simulation process. Godoy (2003) presents the approach for a single variable, while Boucher and Dimitrakopoulos (2009) apply it to the simulation of independent MAF factors. However, the modified Blusim algorithm as described below is geared towards the joint simulation of multiple correlated variables using only the routine normal scores transform and LMC as follows:

1. Transform the data to normal scores (standard Gaussian distribution);
2. Fit the LMC using the normal scores data;
3. Generate a random path that visits each block once;
4. Select the first block and build the covariance matrix;
5. Decompose the covariance matrix in a lower and upper triangular form (Cholesky decomposition: Equation 3);

$$\begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{bmatrix} = \begin{bmatrix} L_{11} & & \\ L_{21} & L_{22} & \\ L_{31} & L_{32} & L_{33} \end{bmatrix} \begin{bmatrix} U_{11} & U_{12} & U_{13} \\ & U_{22} & U_{23} \\ & & U_{33} \end{bmatrix} \qquad \text{Equation 3}$$

Where,

- C_{11} is the matrix with the covariances between the data locations (this is a point to point covariance).
- C_{13} is the matrix with the covariances between the data locations and the discretized points inside the block (this is a point to point covariance).

- C_{33} is the matrix with the covariances between the discretized points inside the block (this is a point to point covariance).
- C_{12} is the matrix with the covariances between the data locations and the previously simulated blocks (this is a point to block covariance).
- C_{23} is the matrix with the covariances between the previously simulated blocks and the discretized points inside the block being simulated (this is a block to point covariance).
- C_{22} is the matrix with the covariances between the previously simulated blocks (this is a block to block covariance).

9. Next the simulated values at each discretized location inside the block are generated by solving the system of equations in Equation 4;

$$\begin{bmatrix} L_{11} & & \\ L_{21} & L_{22} & \\ L_{31} & L_{32} & L_{33} \end{bmatrix} \begin{bmatrix} W_1 \\ W_2 \\ W_3 \end{bmatrix} = \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix} \qquad \text{Equation 4}$$

Where,

- $W_1 = L_{11}^{-1} Z_1$ is a vector of weights, that are solved for.
- $W_2 = L_{22}^{-1}[Z_2 - C_{21} C_{11}^{-1} Z_1]$ is also a vector of weights that are solved for.
- Z_1 is a vector containing the conditioning data values.
- W_3 is a vector of standard normal random numbers.
- Z_3 is a vector that will contain the simulated values for every discretized location inside the block.
- Z_2 is a vector containing the previously simulated block values (Gaussian values).

10. Generate the Gaussian simulated values at each discretized location and average them to get the simulated block Gaussian value. This is used for further conditioning during the simulation process;
11. The Gaussian simulated values at each discretized location are back-transformed and averaged to get the simulated block value. These block values are not used during the simulation process and can be written out and discarded;
12. Move on to the next block.

The modified Blusim algorithm requires point-point, point-block and block-block covariances. The point-point covariances can be calculated using the fitted LMC. The point-block covariances can be calculated by discretizing the block v and calculating the average covariance between point u_1 and the discretization points within v as shown in Equation 5 below (Journel and Huijbregts, 1978, Liu and Journel, 2009).

$$C(u_1, v) = \frac{1}{N} \sum_{i=1}^{N} C(u_1, u_i) \qquad i \in v \qquad \text{Equation 5}$$

For calculation of the block to block covariance between blocks v_1 and v_2, both blocks are discretized and the average covariance between the discretized points is calculated using Equation 6 below (Journel and Huijbregts, 1978, Liu and Journel, 2009).

$$C(v_1,v_2) = \frac{1}{N_1 N_2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} C(u_i,u_j) \qquad i \in v_1 \; and \qquad j \in v_2 \qquad \text{Equation 6}$$

Walker Lake Data

The Walker Lake data (Isaaks and Srivastava, 1989) is an exhaustive set of data with 78,000 points for two continuous variables, U and V. The exhaustive data set was sampled at a 15 m spacing to produce a set of 357 sample data. Descriptive statistics of the sample data are shown in Table 1 and Figure 1 is a plot of the sample locations. The statistics indicate that there are differences between the exhaustive and sample data sets. The differences are larger for variable U than variable V. For variable U, the mean and standard deviation of the sub sampled data set is lower compared to the exhaustive data. The higher percentiles are also different indicating a change in the histogram compared to the exhaustive data.

Table 1: Statistics of the exhaustive and sub sampled data sets for Variables U and V

	Exhaustive data set		Sample data set	
Variable	U	V	U	V
Number of data	78,000	78,000	357	357
Mean	266.04	277.98	244.84	276.84
Standard deviation	488.45	249.84	416.25	253.96
Minimum	0.0	0.0	0.0	0.0
Maximum	9499.51	1631.16	2692.34	1160.29
Coefficient of variation	1.84	0.90	1.70	0.92
25^{th} percentile	6.67	67.80	5.24	59.26
50^{th} percentile	56.90	221.25	59.36	214.78
75^{th} percentile	316.35	429.35	285.14	436.19
Pearson correlation	0.646		0.665	

Both variables U and V from the sample set were transformed into normal scores and a LMC depicted in Figure 2 was fit. The LMC consisted of a nugget effect and two spherical structures using the values detailed in Equation 7, where Sph_1 is a spherical structure with a maximum range of 50 metres in the 346° direction and a minimum range of 25 metres in the 76° direction; Sph_2 is a spherical structure with a maximum range of 65 metres in the 346° direction and a minimum range of 42 metres in the 76° direction.

$$\gamma\left(\begin{bmatrix} h_U & h_{UV} \\ h_{UV} & h_V \end{bmatrix}\right) = \begin{bmatrix} 0.3 & 0.17 \\ 0.17 & 0.1 \end{bmatrix} + \begin{bmatrix} 0.53 & 0.34 \\ 0.34 & 0.23 \end{bmatrix} Sph_1 + \begin{bmatrix} 0.17 & 0.34 \\ 0.34 & 0.67 \end{bmatrix} Sph_2 \qquad \text{Equation 7}$$

Simulation using the enhanced Blusim algorithm

Multivariate direct block simulation of both variables U and V for 10 m by 10 m blocks was performed using the LMC model listed in Equation 7 above. The following parameters were used for the direct block simulation:

- Fifty realisations.
- A maximum of 28 samples (per variable).
- The number of previously simulated blocks to be used was set to 15 (per variable).
- Isotropic search with a radius of 300 meters.
- A block discretization of five by five nodes per block.

Table 2 is a listing of the parameters of the direct block simulation grid.

Table 2 Dimensions of the direct block simulated field

Dimension	Grid origin	Block size (metres)	Number of blocks
Easting	5.0	10	26
Northing	5.0	10	30

After simulation, the modified Blusim algorithm writes out both the Gaussian and back-transformed block values. These were then used for validation in both data and Gaussian space.

Comparison of direct block simulation to reality in Gaussian space

In Gaussian space the variance of the block realizations can be compared against the variance derived from a regularization of the fitted LMC (Equation 7) for 10 by 10 meter blocks (Table 3). For all 50 realizations the means, standard deviations and U-V cross-correlations were calculated and summarized with histogramsdisplayed in Figure 3. In Gaussian space there are some fluctuations in the mean, standard deviation and cross-correlation, however on average the realizations reproduce the target values (as given by the regularized LMC).

Table 3 Block standard deviations and correlation derived from regularization (Gaussian space)

Statistic	Variable	Value
Standard deviation	U	0.75
	V	0.86
Correlation	U-V	0.87

To verify reproduction of block spatial variability in Gaussian space the variograms of the block realizations are compared with the variogram model derived from regularization of the fitted LMC, which provides the theoretically correct model for blocks. Figure 4 contains variogram plots of all 50 realizations and also the regularized LMC variogram model. In general, variogram reproduction compares with theory, although some realizations do have marginally longer ranges for variable V.

Comparison of direct block simulation to reality in data space

The point data from the exhaustive Walker Lake data set was converted to 10 meter by 10 meter blocks by reblocking the exhaustive information using the grid parameters used for the direct block simulation. Table 4 is a listing of the statistics of the reblocked exhaustive point data.

Table 4: Statistics of the reblocked exhaustive data set for Variables U and V

	Reblocked exhaustive data set	
Variable	U	V
Number of data	780	780
Mean	266.04	277.98
Standard deviation	316.84	216.09
Minimum	0.0	0.0
Maximum	2062.44	1247.47

Coefficient of variation	1.19	0.78
25th percentile	42.39	102.97
50th percentile	161.21	239.0
75th percentile	357.79	403.90
Pearson correlation	0.75	

The direct block realizations were compared to the reblocked reality through inspection of realization images and summary statistics. This comparison depends on how well the 15m sub sampled data set reproduces the characteristics of the exhaustive data set. Figure 5 contains plots of the reblocked exhaustive data and two randomly chosen realizations. Visual inspection of the plots for both variables reveals that the block realizations display similar patterns as seen in the reblocked exhaustive data set.

Figure 6 contains histograms of the means, standard deviations and U-V cross-correlations for all 50 realizations. On these plots the average statistics of the exhaustive data and the conditioning data are highlighted as red and yellow lines respectively. On average for variable V, the realizations reproduce the mean and variance as indicated by the reblocked exhaustive data. For variable U, reproduction is satisfactory given the mean of the conditioning data does differ from the exhaustive mean as a function of the sampling process. On average the mean of the realizations is 252.5, which is 3% higher than the mean of the 15 m subsample data. On average reproduction of the cross-correlation between U and V at the block scale is satisfactory, although the correlation is slightly higher (0.78) compared to the correlation derived from the reblocked exhaustive data (0.75).

Figure 7 contains Q-Q plots that compare the quantiles of the reblocked exhaustive data to the realization quantiles for both variables. These plots confirm that general histogram reproduction in the realizations is satisfactory with variable V having some differences in the upper bins of the histogram due to the absence of high values in the 15m sub sampled data set.

Figure 8 contains variogram plots where in data space the variograms of the block realizations are compared with the variogram derived from the reblocked exhaustive data. Variograms of the block realizations correspond favorably with the variogram of the reblocked exhaustive data, however the reblocked exhaustive data has slightly longer ranges (especially in the major direction) not seen in the realizations. This feature is not observed in the 15m sample data set from which the LMC is derived and as a consequence the realizations do not reproduce it.

Conclusion

The Blusim algorithm has been modified to also reproduce the spatial relationship between the blocks. The modified algorithm discretizes a block and generates simulated values for every discretization location inside the block by conditioning on the surrounding data and on previously simulated blocks. Afterwards the values for the discretization points are averaged to get the block simulated value and the point values are discarded. Reproduction of histograms and variograms by the modified Blusim algorithm is shown to be reasonable. The advantages of the modified Blusim algorithm are:

- Only a point covariance/variogram model is required and the block-block and point-block covariances are calculated using an averaging process. Variogram regularization to the relevant blocksize is not necessary.
- Simulation still proceeds on a point level i.e. the discretization points inside the block are simulated and thus the point normal score transform can still be used. A Gaussian anamorphosis at block level is not required.
- An added benefit of the algorithm is that because simulation proceeds at a point level and block-point and block-block covariances are calculated using a discretization, the algorithm can be programmed to handle different block sizes in one simulation, that is, the simulated field does not have to consist of regular blocks; the blocks can be of different size and shape-as long as the level of discretization is sufficient to calculate the covariances to an acceptable level of precision.

- How long does it take.

References

Boisvert, J., and Deutsch, C.V., 2007., Multivariate block LU simulation. Ninth annual report of the Centre for Computational Geostatistics. Paper 305.

Boucher, A., and Dimitrakopoulos, R.D., 2009, Block simulation of multiple correlated variables. Mathematical Geosciences 41, pp 215-237.

Godoy, M., 2003, The effective management of geological risk in long-term scheduling of open pit mines. PhD thesis, University of Queensland, Brisbane, 256pp

Isaaks, E., and Srivastava, R.M., 1989., Applied geostatistics. Oxford University Press Inc, New York. 561 pp.

Journel, A., and Huijbregts, C.J., 1978., Mining geostatistics. Academic Press Limited. 610 pp.

Liu,Y., and Journel, A., 2009., A package for geostatistical integration of coarse and fine scale data. Computers and Geosciences 35, pp 527-547.

Ortiz, J.M., and Deutsch, C.V., 2007., Local recoverable reserves prediction with block LU simulation. Ninth annual report of the Centre for Computational Geostatistics. Paper 304.

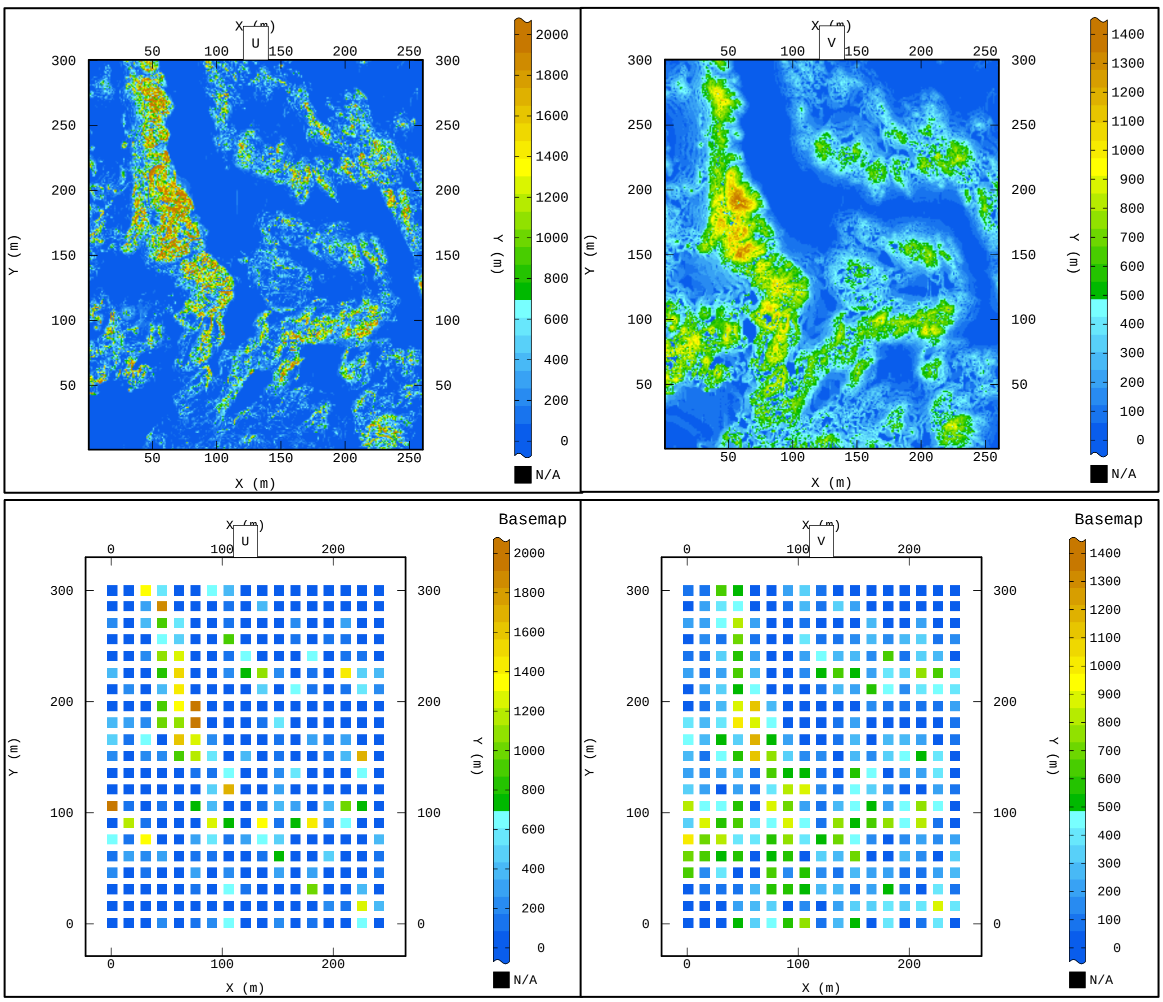

Figure 1: Exhaustive data set (top) and sample locations (bottom) for variable U (left) and variable V (right)

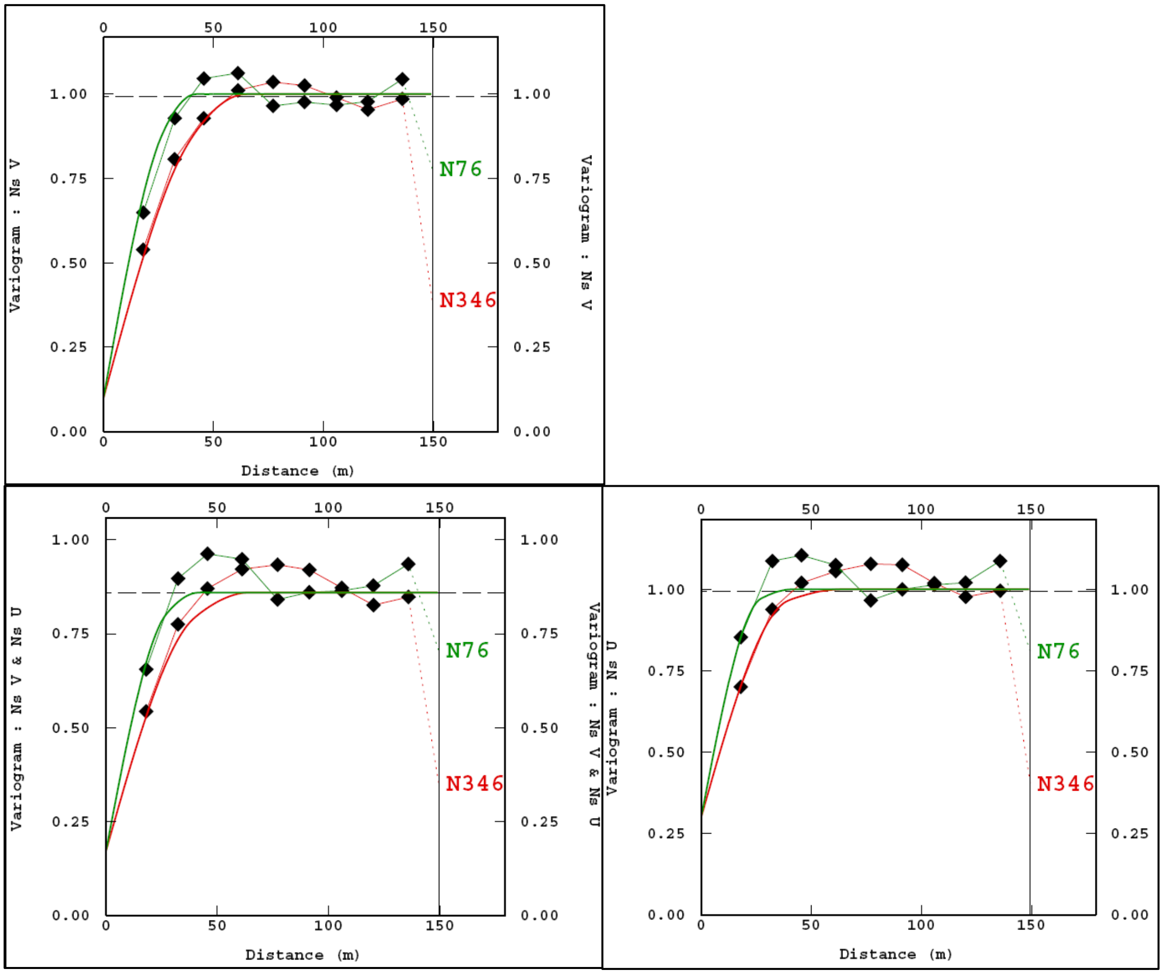

Figure 2: LMC fitted to the normal scores of variables U and V (major direction in red and minor in green)

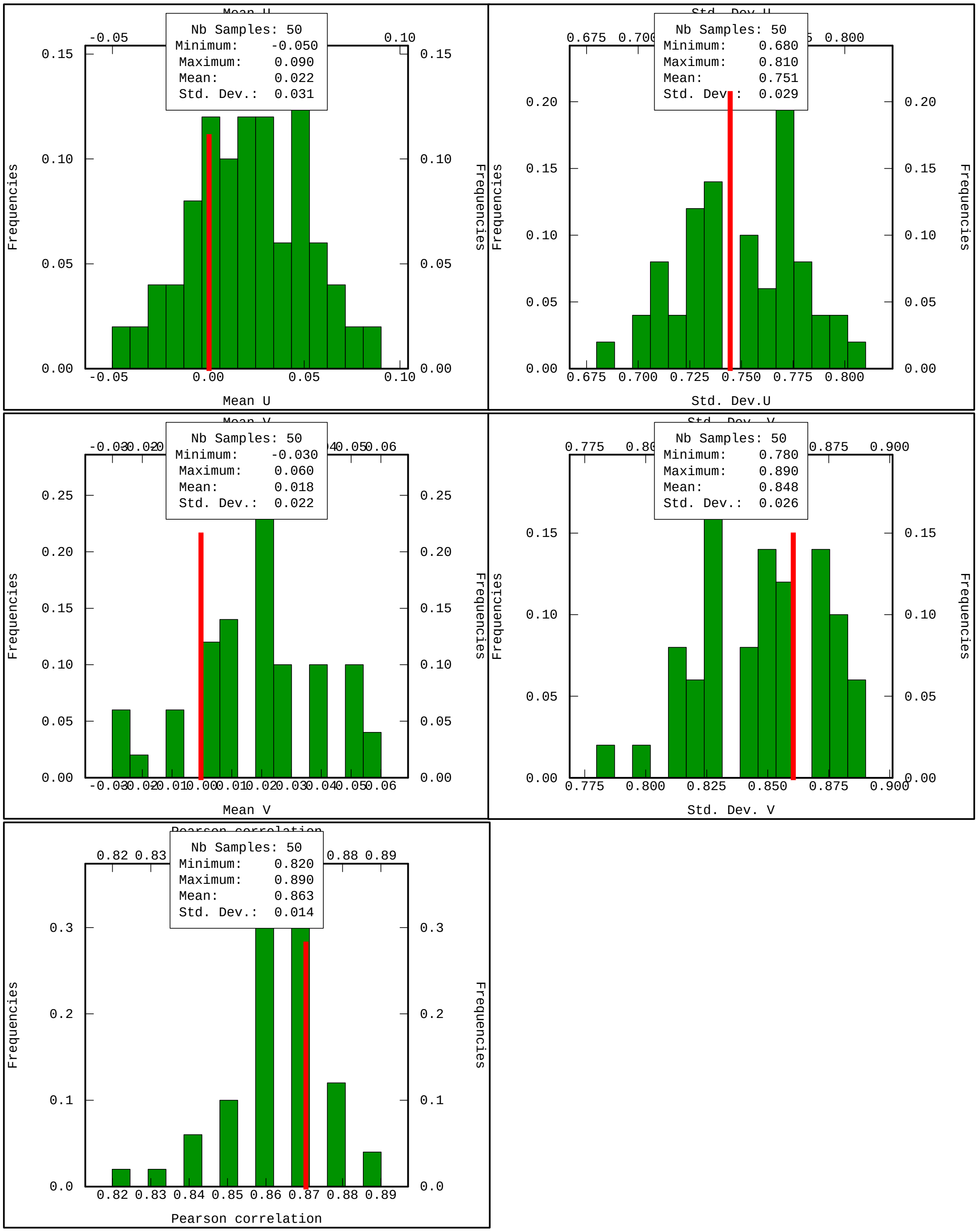

Figure 3: Histograms of direct block realization means, standard deviations and U-V correlations in Gaussian space (the red line represents the values derived from regularization of the fitted LMC)

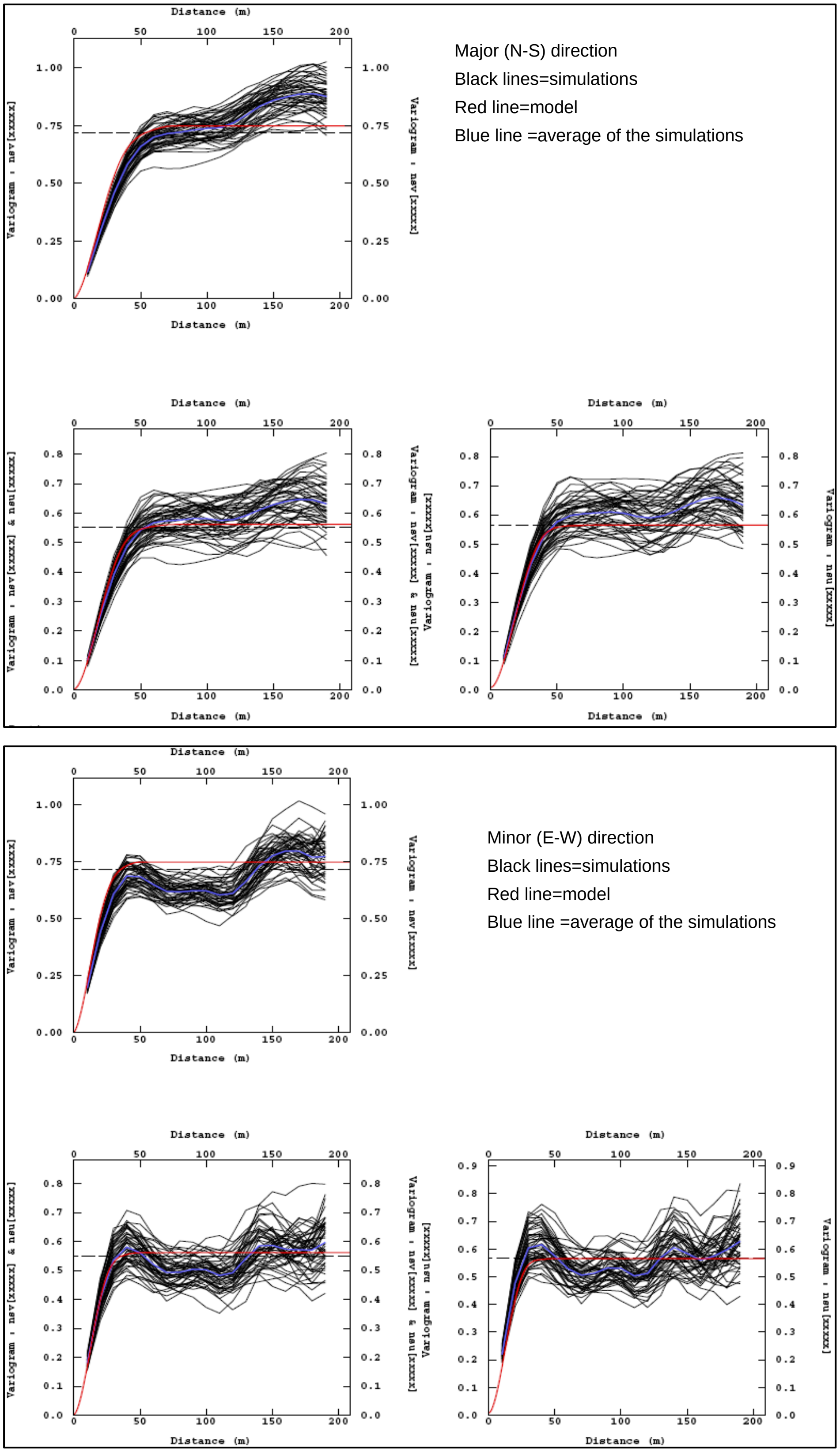

Figure 4: Variograms of direct block realizations (black) against theoretical block variogram from regularized LMC model in Gaussian space (red)

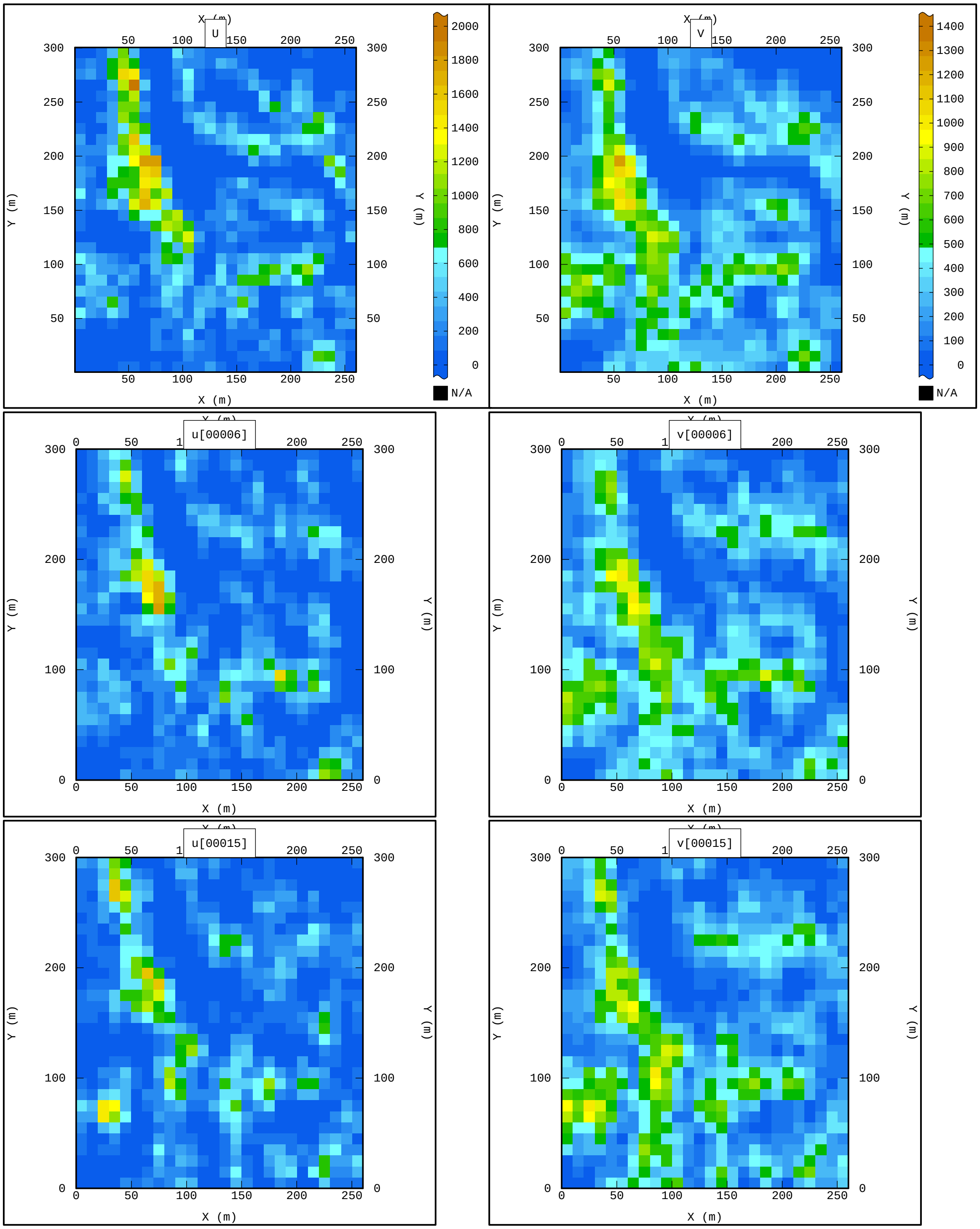

Figure 5: Plots of the reblocked exhaustive data set (top) and two block realizations (middle and bottom) for variable U (left) and variable V (right)

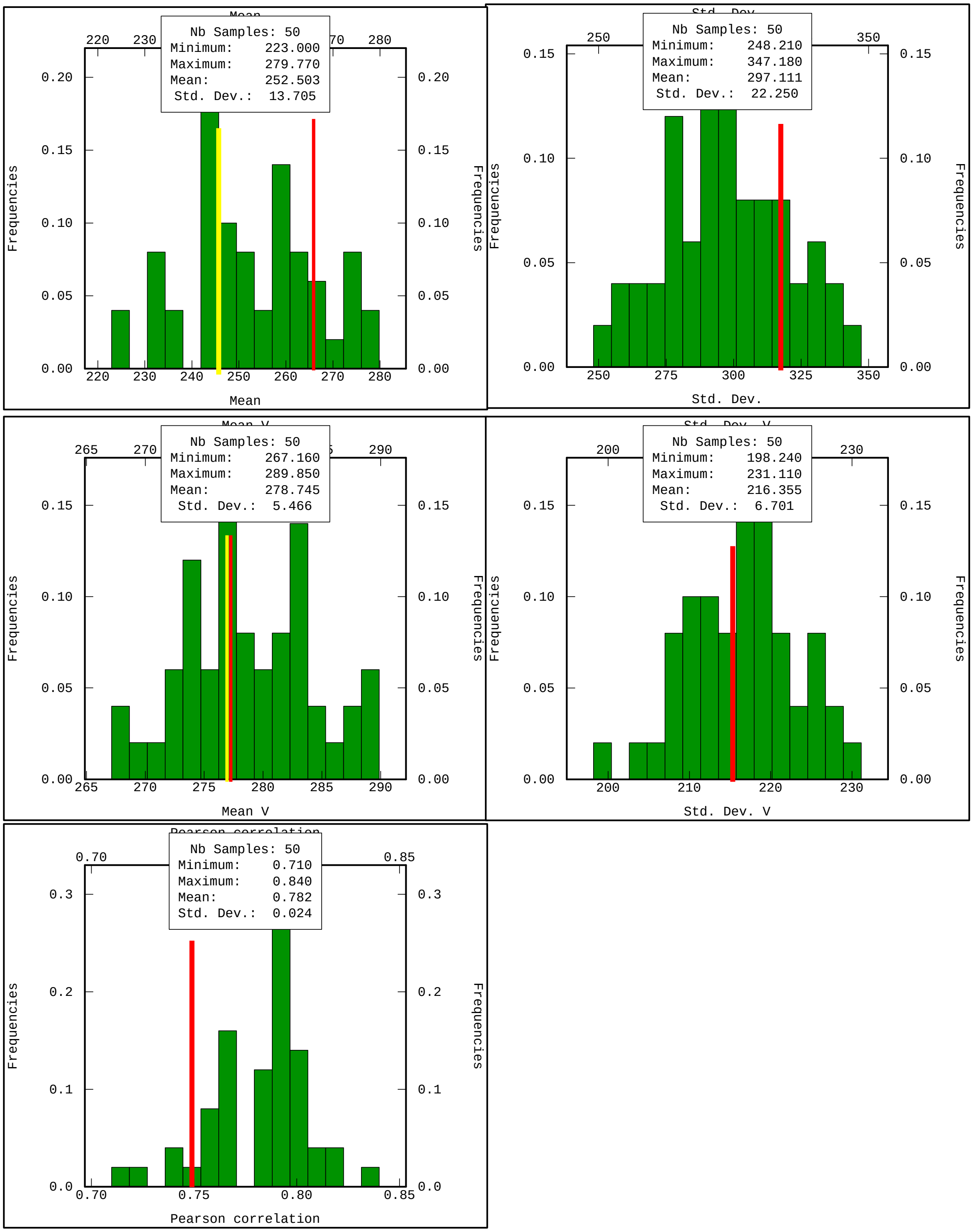

Figure 6: Histograms of direct block realization means, standard deviations and U-V correlations in data space (the red line represents the values derived from the reblocked exhaustive data set, the yellow line is the mean of the sub sampled 15 m data set)

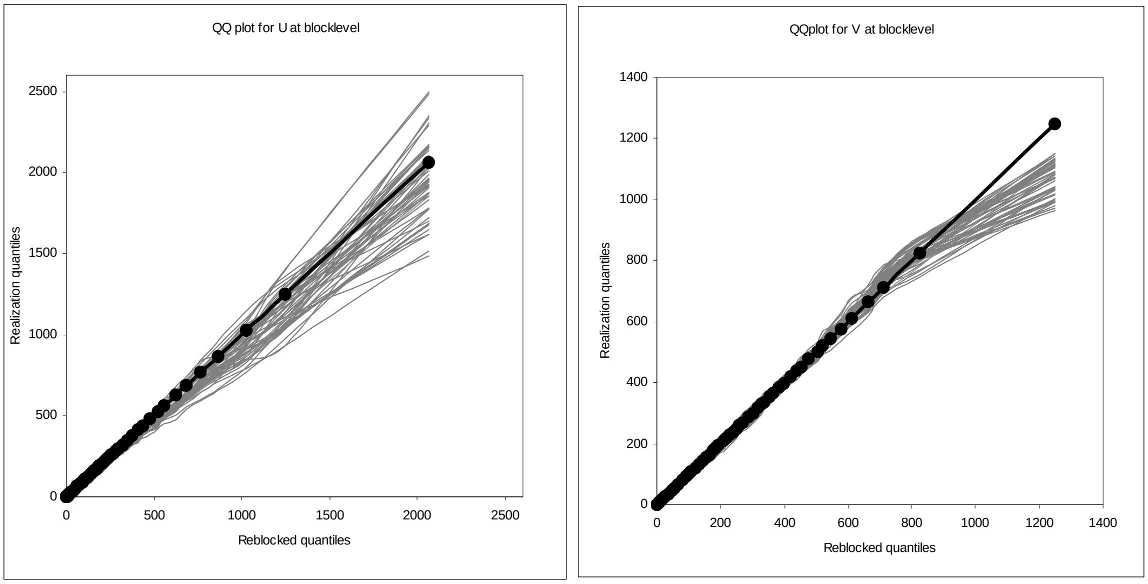

Figure 7: QQ plots of the block realizations (grey lines) against the reblocked exhaustive data set (thick black line)

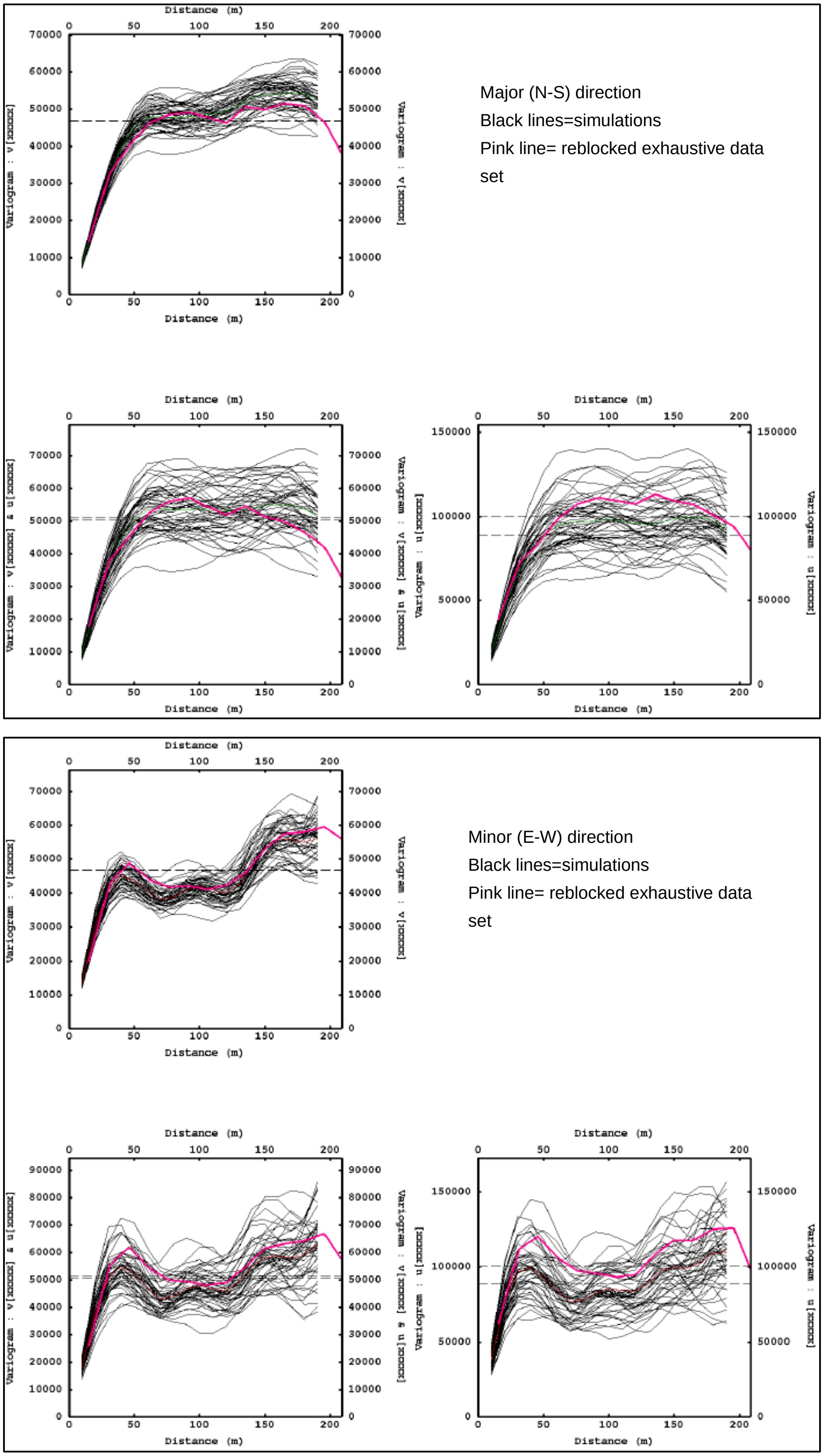

Figure 8: Variograms of direct block realizations (black) against reblocked exhaustive data set (pink) variogram in data space