A Methodology to Quantify and Transfer Variogram Uncertainty through Kriging and Simulation

Behrang Koushavand, Julián M. Ortiz and Clayton V. Deutsch

The variogram is central to any geostatistical study. Due to limited data available in the early stages of modeling, there is considerable uncertainty in all statistical parameters including the variogram. The uncertainty in the experimental should be quantified and transferred into subsequent modeling. A global simulation approach could be used for this purpose. Alternatively, an analytical approach based on the variance–covariance matrix of the experimental variogram values is developed. Experimental variogram uncertainty is transferred to variogram model uncertainty by fitting models to many realizations of the variograms. These realizations are used in estimation and simulation. Robust estimates and local uncertainty is calculated in the presence of variogram model uncertainty.

Introduction

Most geostatistical studies require a variogram model; however, a reliable variogram is difficult to infer in presence of sparse data. Fair uncertainty predictions and robust estimates motivate the quantification and use of variogram uncertainty in estimation and simulation. Variogram uncertainty has been considered by different authors. Webster and Oliver (1992), Müller and Zimmerman (1999) and Bogaert and Russo (1999) measured the variogram uncertainty in sampling schemes and suggested different methods to minimize this uncertainty. Cressie (1985), Ortiz and Deutsch (2002), and Pardo-Igúzquiza and Dowd (2001) suggested similar expressions for the covariance matrix of experimental variogram estimates to the ergodic variogram (Marchant and Lark 2004).

The semivariogram (often called the variogram for brevity) is defined under a second order stationarity assumption as (Matheron, 1971)

$$\gamma(\mathbf{h}) = \frac{1}{2} \mathrm{E}\left\{ \left[Z(\mathbf{u}+\mathbf{h}) - Z(\mathbf{u}) \right]^2 \right\} \tag{1}$$

It is estimated by the method of moments (Journel and Huijbregts, 1978):

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} \left[Z(\mathbf{u}_i + \mathbf{h}) - Z(\mathbf{u}_i) \right]^2 \tag{2}$$

where $N(\mathbf{h})$ is the number of pairs of values for the separation vector $\mathbf{h}$.

For a fixed set of K lag vectors, the variance–covariance matrix of the experimental variogram is a K×K matrix $\boldsymbol{\Sigma}$. The pq element is (Marchant and Lark, 2004):

$$[\boldsymbol{\Sigma}]_{pq} = \mathrm{Cov}\left[\hat{\gamma}(\mathbf{h}_p), \hat{\gamma}(\mathbf{h}_q) \right] \tag{3}$$

The diagonal elements give the variance of the variogram estimates:

$$[\boldsymbol{\Sigma}]_{pp} = \mathrm{Cov}\left[\hat{\gamma}(\mathbf{h}_p), \hat{\gamma}(\mathbf{h}_p) \right] = \mathrm{Var}\left(\hat{\gamma}(\mathbf{h}_p) \right) \tag{4}$$

This matrix could be calculated with a multivariate distribution model or predicted with simulation. The data are transformed to be Gaussian in both methods. Moreover, a base case initial variogram must be calculated and modeled. In the analytical method, the variance-covariance matrix between the variogram lags is calculated with Matheron's expression of fourth order Gaussian moments (Matheron, 1965). The GSLIB (Deutsch and Journel 1998) **`gamv`** program is modified to calculate the variogram variance-covariance and correlation matrices based on an initial variogram model. The program **`gamu1`** fits a Gamma distribution to the distribution of variogram values at each lag.

The second method is based on unconditional simulation. GSLIB (Deutsch and Journel 1998) **`gamv`** and **`lusim`** programs are combined. The program **`gamu2`** uses an initial variogram model to generate an

unconditional simulation at the data locations. The experimental variogram is calculated with the simulated data at each realization. This approach is similar to the one developed in Ortiz and Deutsch (2002).
The variogram uncertainty is transfered to variogram model uncertainty by fitting variogram models to each one of the variograms calculated from the simulated data with the same sample data configuration. Different simple Kriging estimates (SK) are calculated with each variogram to get local uncertainty. Sequential Gaussian simulation (`sgs`) is also used to generate different realizations with the resulting variogram models to get more robust and realistic spatial uncertainty.

Analytical Method

The expected value of $\hat{\gamma}(\mathbf{h})$ for each lag distance $\mathbf{h}$ is equal to $\gamma(\mathbf{h})$. From the definition of covariance

$$[\boldsymbol{\Sigma}]_{pq} = E\left[\hat{\gamma}(\mathbf{h}_p)\hat{\gamma}(\mathbf{h}_q)\right] - \gamma(\mathbf{h}_p)\gamma(\mathbf{h}_q) \tag{5}$$

$$= \frac{1}{4N(\mathbf{h}_p)N(\mathbf{h}_q)} \sum_{i=1}^{N(h_p)} \sum_{j=1}^{N(h_q)} E\left\{\left[Z(\mathbf{u}_i) - Z(\mathbf{u}_i + \mathbf{h}_p)\right]^2 \left[Z(\mathbf{u}_j) - Z(\mathbf{u}_j + \mathbf{h}_q)\right]^2\right\} - \gamma(\mathbf{h}_p)\gamma(\mathbf{h}_q) \tag{6}$$

Assuming that the regionalized variable is multi-Gaussian, the variogram uncertainty can be calculated from theory. Expanding expression (6), the covariance can be written as a sum of fourth order moments:

$$E\left\{\left[Z(\mathbf{u}_i) - Z(\mathbf{u}_i + \mathbf{h}_p)\right]^2 \left[Z(\mathbf{u}_j) - Z(\mathbf{u}_j + \mathbf{h}_q)\right]^2\right\}$$
$$= \left\{\begin{array}{l} E\left[Z(\mathbf{u}_i)^2 .Z(\mathbf{u}_j)^2\right] + E\left[Z(\mathbf{u}_i)^2 .Z(\mathbf{u}_j + \mathbf{h}_q)^2\right] - E\left[2.Z(\mathbf{u}_i)^2 .Z(\mathbf{u}_j).Z(\mathbf{u}_j + \mathbf{h}_q)\right] \\ +E\left[Z(\mathbf{u}_i + \mathbf{h}_p)^2 .Z(\mathbf{u}_j)^2\right] + E\left[Z(\mathbf{u}_i + \mathbf{h}_p)^2 .Z(\mathbf{u}_j + \mathbf{h}_q)^2\right] - E\left[2.Z(\mathbf{u}_i + \mathbf{h}_p)^2 .Z(\mathbf{u}_j).Z(\mathbf{u}_j + \mathbf{h}_q)\right] \\ -E\left[2.Z(\mathbf{u}_i).Z(\mathbf{u}_i + \mathbf{h}_p).Z(\mathbf{u}_j)^2\right] - E\left[2.Z(\mathbf{u}_i).Z(\mathbf{u}_i + \mathbf{h}_p).Z(\mathbf{u}_j + \mathbf{h}_q)^2\right] \\ +E\left[4.Z(\mathbf{u}_i).Z(\mathbf{u}_i + \mathbf{h}_p).Z(\mathbf{u}_j).Z(\mathbf{u}_j + \mathbf{h}_q)\right] \end{array}\right\} \tag{7}$$

This covariance is called a quadratic covariance (Matheron, 1965) and it can be calculated if $Z(\mathbf{u}_i)$, $Z(\mathbf{u}_i + \mathbf{h}_p)$, $Z(\mathbf{u}_j)$ and $Z(\mathbf{u}_j + \mathbf{h}_q)$ have a multivariate Gaussian distribution. In such case, any fourth order moment can be calculated using the pair wise covariance values as follows:

$$E\{Z_1 Z_2 Z_3 Z_4\} = C_{12}C_{34} + C_{13}C_{24} + C_{14}C_{23} \tag{8}$$

So, Equation 7 can be calculated as below:

$$E\left\{\left[Z(\mathbf{u}_i) - Z(\mathbf{u}_i + \mathbf{h}_p)\right]^2 \left[Z(\mathbf{u}_j) - Z(\mathbf{u}_j + \mathbf{h}_q)\right]^2\right\}$$
$$= \left\{\begin{array}{l} [C_{11}C_{33} + C_{13}C_{13} + C_{13}C_{13}] + [C_{11}C_{44} + C_{14}C_{14} + C_{14}C_{14}] - 2[C_{11}C_{34} + C_{13}C_{14} + C_{14}C_{13}] \\ +[C_{22}C_{33} + C_{23}C_{23} + C_{23}C_{23}] + [C_{22}C_{44} + C_{24}C_{24} + C_{24}C_{24}] - 2[C_{22}C_{34} + C_{23}C_{24} + C_{24}C_{23}] \\ -2[C_{12}C_{33} + C_{13}C_{23} + C_{13}C_{23}] - 2[C_{12}C_{44} + C_{14}C_{24} + C_{14}C_{24}] + 4[C_{12}C_{34} + C_{13}C_{24} + C_{14}C_{23}] \end{array}\right\} \tag{9}$$

Where:

$$
\begin{aligned}
C_{11} &= \mathrm{Cov}\left[Z(\mathbf{u}_i), Z(\mathbf{u}_i)\right] \\
C_{12} &= \mathrm{Cov}\left[Z(\mathbf{u}_i), Z(\mathbf{u}_i+\mathbf{h}_p)\right] \\
C_{13} &= \mathrm{Cov}\left[Z(\mathbf{u}_i), Z(\mathbf{u}_j)\right] \\
C_{14} &= \mathrm{Cov}\left[Z(\mathbf{u}_i), Z(\mathbf{u}_j+\mathbf{h}_q)\right] \\
C_{22} &= \mathrm{Cov}\left[Z(\mathbf{u}_i+\mathbf{h}_p), Z(\mathbf{u}_i+\mathbf{h}_p)\right] \\
C_{23} &= \mathrm{Cov}\left[Z(\mathbf{u}_i+\mathbf{h}_p), Z(\mathbf{u}_j)\right] \\
C_{24} &= \mathrm{Cov}\left[Z(\mathbf{u}_i+\mathbf{h}_p), Z(\mathbf{u}_j+\mathbf{h}_q)\right] \\
C_{33} &= \mathrm{Cov}\left[Z(\mathbf{u}_j), Z(\mathbf{u}_j)\right] \\
C_{34} &= \mathrm{Cov}\left[Z(\mathbf{u}_j), Z(\mathbf{u}_j+\mathbf{h}_q)\right] \\
C_{44} &= \mathrm{Cov}\left[Z(\mathbf{u}_j+\mathbf{h}_q), Z(\mathbf{u}_j+\mathbf{h}_q)\right]
\end{aligned}
$$

A FORTRAN code, **`Gamu1`** is provided to calculate this expression. The ergodic variogram function is required as an input. This is taken as the fitted variogram model.
Davis and Borgman (1982) show that the distribution of the random variable,

$$\frac{\hat{\gamma}(\mathbf{h})-\gamma(\mathbf{h})}{\sqrt{\mathrm{Var}\{\hat{\gamma}(\mathbf{h})\}}} \tag{10}$$

converges to a standard normal distribution, N(0,1), as the number of data increases to infinity. Several authors (Cressie, 1985; Ortiz and Deutsch, 2002; Pardo-Igúzquiza and Dowd, 2001) assumed the distribution of each lag is Gaussian, and defined by a mean corresponding to the ergodic variogram $\gamma(\mathbf{h})$, and the variance of each lag, given by equation (4).

Rather than assuming a Gaussian distribution for each $\hat{\gamma}(\mathbf{h})$, we use a multivariate two parameter (shape and scale) Gamma distribution that is more flexible, since a gamma distribution is a general form of a normal distribution. This distribution has only positive values that is appropriate for the squared differences of the variogram. It is also asymmetric with flexible ability to fit the actual distribution shape of squared differences. Consider a Gamma distribution for each lag:

$$
\left.
\begin{aligned}
&E\{Z_i\} = m_i = \gamma(\mathbf{h}_i) \\
&\mathrm{Cov}\{Z_i, Z_j\} = E\{Z_i.Z_j\} - m_i.m_j = \mathbf{\Sigma}_{ij} \\
&\mathrm{Var}\{Z_i\} = \sigma_i \\
&F_i(z) \sim \Gamma(m_i, \sigma_i)
\end{aligned}
\right.
\qquad i = 1,\ldots,\mathrm{nlag}
$$

Where $\gamma(\mathbf{h})$ is the variogram model, $\mathbf{\Sigma}$ is the variance covariance matrix of variogram, and nlag is the number of lags.

To simulate realizations of the variogram, a multivariate normal distribution is generated and each of distributions is back-transferred to a marginal Gamma distribution. Assume $\mathbf{Y}$ is a multivariate normal distribution: $\mathbf{Y} = \mathrm{MVN}_{\mathrm{nlag}}(\mathbf{M},\mathbf{\Sigma})$ where:

$$\mathbf{M} = \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_{\mathrm{nlag}} \end{bmatrix} = \begin{bmatrix} \gamma(\mathbf{h}_1) \\ \gamma(\mathbf{h}_2) \\ \vdots \\ \gamma(\mathbf{h}_{\mathrm{nlag}}) \end{bmatrix}$$

Cholesky Decomposition is used to generate $\mathbf{Y}_l$:

$$\begin{aligned} \mathbf{\Sigma} &= \mathbf{L}.\mathbf{U} \\ \mathbf{Y}_l &= \mathbf{L}.\mathbf{w}_l \end{aligned} \qquad l = 1, \ldots, \text{nsim}$$

Where,

$$\mathbf{w}_l = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_{nlag} \end{bmatrix}$$

$$\mathbf{w} \sim N(0,1)$$

$$\text{Cov}\{w, w\} = \mathbf{I}$$

Then, all marginal normal distributions are back-transformed to Gamma distributions:

$$\mathbf{Z}_i = F^{-1}\left(G\left(\mathbf{Y}_i\right)\right); i = 1, \ldots, \text{nlag} \qquad (11)$$

Where, G is the Normal Cumulative distribution function and F^{-1} is the Quantile function of the lag specific Gamma distribution. The step by step procedure:

- Transform the original data to normal scores.
- Compute an experimental variogram, then fit an initial variogram model,
- Compute the required fourth order moments to establish the uncertainty at each lag,
- Generate multivariate normal distribution of correlated variogram values at different lag distances by LU simulation, and then
- Back-transform these values with the Gamma distribution to the original units.

The correlation coefficient between lags may change slightly during back transformation and could be checked afterwards.

Variogram Uncertainty via Global Unconditional Simulation

A Gaussian based simulation algorithm can be used to simulate multiple realizations of the original variable at the data locations. The experimental variograms computed from the simulated realizations provides a direct assessment of variogram uncertainty. The GSLIB (Deutsch and Journel 1998) **`gamv`** and **`lusim`** programs were combined. **`Gamu2`** program unconditionally simulates at the data locations with the initial base case variogram model; new experimental variograms are calculated for each realization. This method is very fast and does not require assuming a distribution at every lag. The steps of this method are :

- Transform the original data to normal scores,
- Compute an experimental variogram, then fit an initial variogram model,
- Generate a multivariate normal distribution of correlated values at sample locations $\mathbf{Y} = \text{MVN}_{\text{ndata}}\left(\mathbf{0}, \mathbf{C}\right)$ where, $\mathbf{C}$ is covariance matrix that is calculated with initial variogram model.
- Calculate new experimental variogram at each realization with generated values.

The analytical model provides more insight into the uncertainty and the model assumptions; however, the brute-force simulation approach leads to essentially the same result.

Case Study

A subvertical tabular deposit is drilled to provide data on vein thickness and grades. The thickness values are used here. The sample locations are presented in Figure 1. Figure 2 left shows the histogram of thickness in original units. The following isotropic variogram model is fitted to the omni-directional experimental variogram of the normal score transformation of thickness:

$$\gamma(\mathbf{h}) = 0.15 + 0.85.\text{Exp}\left(\frac{\mathbf{h}}{250}\right) \quad (12)$$

The experimental variogram and fitted model are shown in Figure 2 right. Figure 3 shows the matrixes of correlation coefficient of lags which were calculated analytically by **`gamu1`** program on left and global simulation method with **`gamu2`** on right considering the variogram model and generating 1000 realizations. There are some slight differences between these two methods in this case study that may cause by multiGaussian assumption of Matherons forth order moment in theoretical method. This data set probably is not fully multiGaussian.

Figure 4 left shows the experimental variograms calculated from the generated gamma distribution and on the right, variograms calculated from global unconditional simulated values.

100 Variogram models were fitted to each of the experimental variograms generated with the simulation method with a weighted least squares criteria by a semi-auto variogram modeling software; Figure 5 left shows the fitted variogram models (gray lines), average of all fitted variograms (red line) and the initial variogram model(blue line). There is slight difference between average variogram and initial variogram model. To get better result, more variogram should be generated and fitted.

The fitted variogram models have different sill values which show uncertainty in the data variance. To prevent interference of variance uncertainty at variogram uncertainty, we standardized all fitted variograms; C0 and C were divided by sill (implicit variance). Figure 5 right shows the standardized fitted variogram models. As it is shown, the standardized variograms skewed from initial variogram model and the difference is more significant.

Histograms of lags calculated with the analytical method (**`gamu1`**) is presented in Figure 6. Gamma distributions are nicely fitted and the shapes of distribution at each lag are very similar in these two methods.

Histograms of variogram parameters and their relationships are presented in Figure 7. Figure 8 shows first 10 variogram realizations, variogram model and standardized model. Figure 9 shows simple kriging maps with different variogram models, kriging result are the same in standardized and unstandardized variogram models in Gaussian space. Variogram uncertainty has been transferred through kriging results. Variogram of this data set is quite uncertain and as it shown from Figure 9, this uncertainty has an impact on kriging.

`Sgsim` **(**Deutsch and Journel 1998) has been run for this data set with considering variogram model uncertainty. 100 realizations have been generated for each of variogram models. Figure 10 shows 10 first realizations which are generated with different unstandardized variograms on left and standardized variogram models on right.

Discussion

To compare the results with common uncertainty assessment methods, **`sgsim`** program was run with the initial variogram model to generate 1000 realizations. Variogram reproduction is presented in Figure 11. Simulation with unstandardized variograms reproduced variograms better then using standardized variograms. This data set has a clear trend and it is simulated by sgsim with initial variogram model. This trend is models with standardized variogram. Using standardized variograms reproduced histogram better that usingsedam miy unstandardized variogram.

E-type mean, variance, P10 and P90 of sgsim with varying variograms and considering only the initial variogram are presented in Figure 12. Uncertainty has increased when accounting for the uncertainty in the variogram and this is much clearer in Figure 13 which shows the histogram of E-type mean, and variance, P10 and P90 of sgsim realizations; using unstandardized variograms transfer variance uncertainty to the model, variance of the resulting variance is increased (0.10 vs. 0.6).

Conclusions

The two methods presented in this paper allow assessing the uncertainty in the variogram model. The joint uncertainty between the variogram lags is accounted for, that is, uncertainty in the fitted variogram model. This requires accounting for the covariance between variogram lags and not only assessing the variance at each lag. This is an interesting result as it allows transferring the uncertainty to the subsequent steps of the modeling (either kriging or simulation). The methodology presented includes this step of transferring the uncertainty to the subsequent modeling steps and it was shown that the resulting variogram is properly reproduced. Both analytical and numerical methods need to transfer data to Gaussian space. Different sill values in variogram models shows variance uncertainty, this also can be transfer to the model by using unstandardized variogram models. Using different variograms in SGS may cause unstable results and user should generate sufficient number of realizations to get more stable results. 100 realizations for each variogram are good enough to transfer variogram model uncertainty.

References

Bogaert, P., and Russo, D., 1999, Optimal spatial sampling design for the estimation of the variogram based on a least squares approach: Water Resour. Res., v. 35, no. 4, p. 1275–1289.

Cressie, N., 1985, Fitting variogram models by weighted least squares: Math. Geology, v. 17, no. 5, p. 563–586.

Davis, B. M., and Borgman L.E., 1982, A note on the asymptotic distribution of the sample variogram: Math. Geology, v.14, no. 2, p. 189–193

Deutsch, C. V., and Journel, A. G., 1998, GSLIB: Geostatistical software library and users guide, 2nd ed.: Oxford University Press, New York, 369 p.

Journel, A. G., and Huijbregts, C. J., 1978, Mining geostatistics: Academic Press, London, 600 p.

Marchant, B. P., and Lark, R. M., 2004, Estimation Variogram Uncertanity: Math Geology, v. 36, no, 8, p. 865-898

Matheron, G., 1965, Les variables régionalisées et leur estimation: Masson et Cie. Editeurs, Paris.

Matheron, G., 1971, The theory of regionalized variables and their applications: Ecole des Mines, Fontainbleau, France, 211 p.

Müller, W. G., and Zimmerman, D. L., 1999, Optimal designs for variogram estimation: Environmetrics, v. 10, no. 1, p. 23–37.

Ortiz, C. J., and Deutsch, C. V., 2002, Calculation of uncertainty in the variogram: Math. Geology: v. 34, no. 2, p. 169–183.

Pardo- Igúzquiza, E., and Dowd, P. A., 2001a, Variance–covariance matrix of the experimental variogram: Assessing variogram uncertainty: Math. Geology, v. 33, no. 4, p. 397–419.

Webster, R., and Oliver, M. A., 1992, Sample adequately to estimate variograms of soil properties: J. Soil Sci., v. 43, no. 1, p. 177–192.

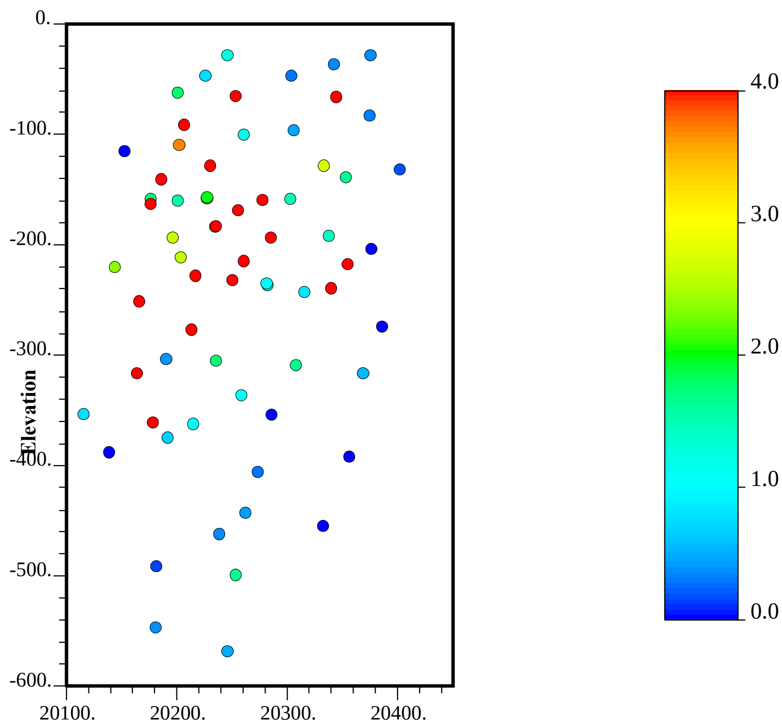


Figure 1. Location map of samples and thickness content taken from the database red.dat.

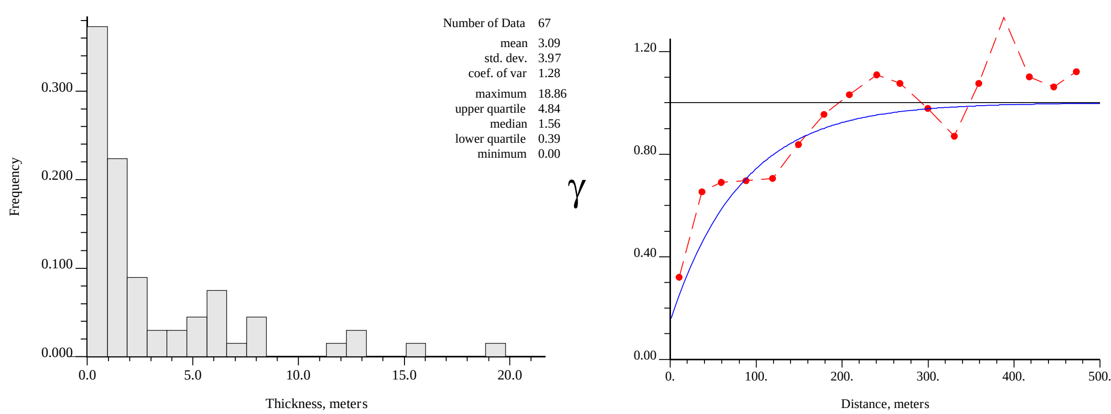
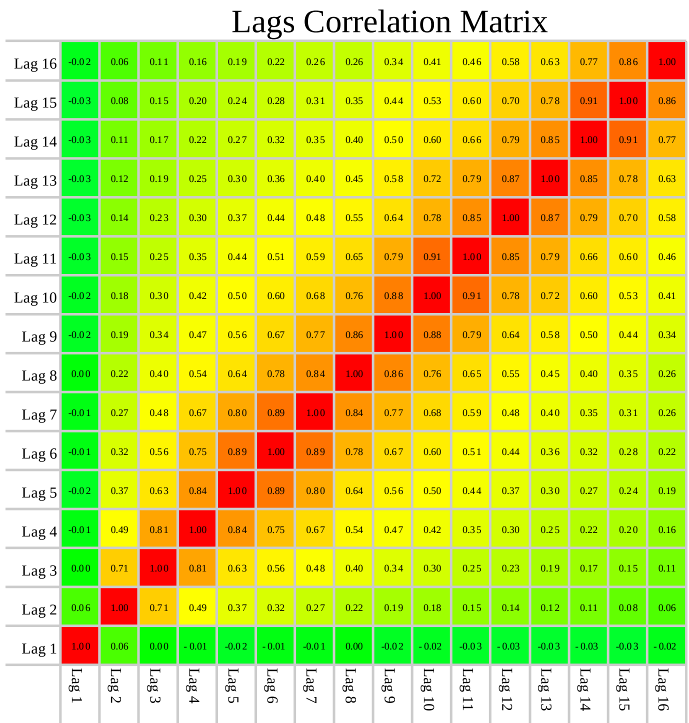


Figure 2. Histogram of thickness in meters and Experimental variogram (dashed line) and fitted model (solid line) of normal scores of thickness



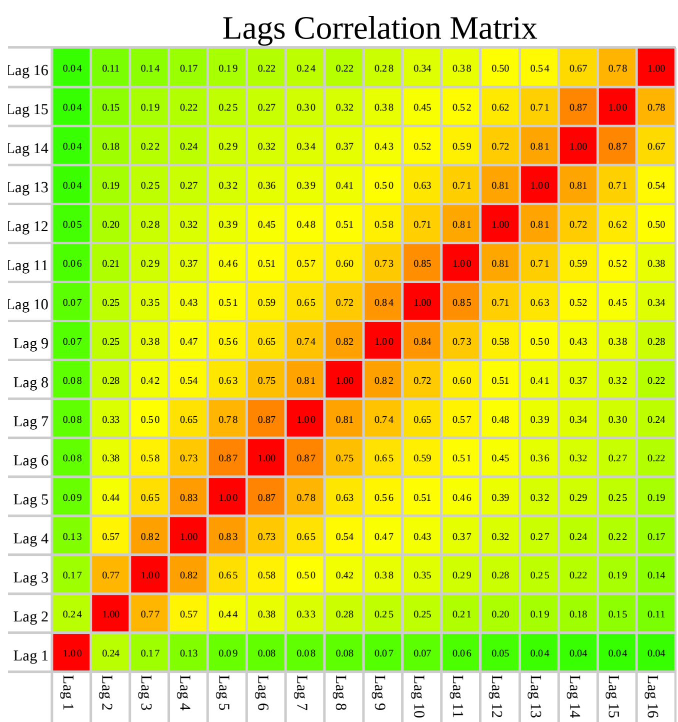


Figure 3. Correlation coefficient matrix between lags; Top: Theoretical method and Bottom: Global unconditional simulation methods

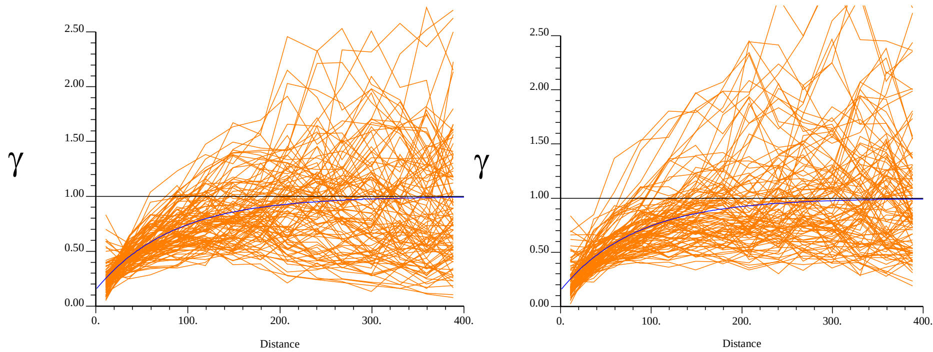


Figure 4. Generated experimental variograms and the initial variogram model; left: Theoretical method and right: Global unconditional simulation Method

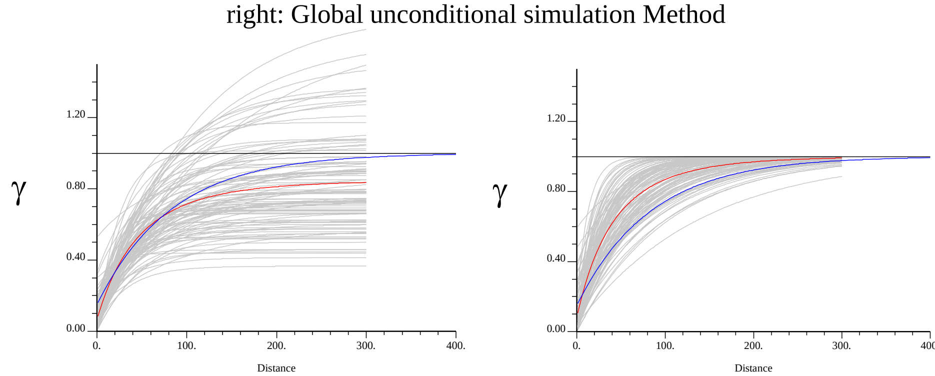


Figure 5. Fitted variogram models (gray lines), average of variograms (red line) and the initial variogram model (blue line) ; Unstandardized (left), Standardized (right)

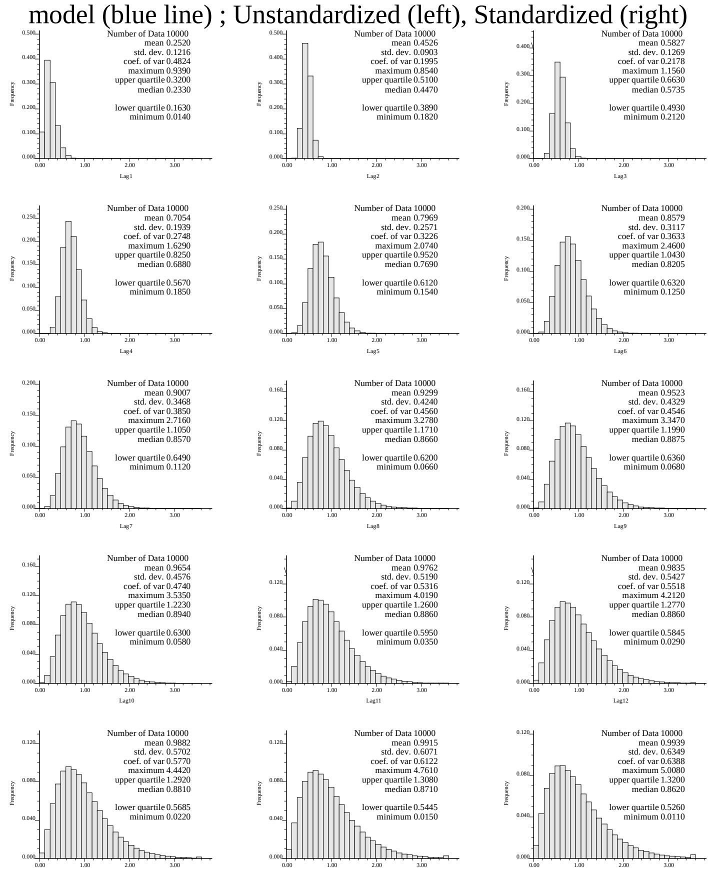


Figure 6. Histograms of variogram values at each lag; the experimental variograms are generated by **`gamu1`** program based on the initial variogram model.

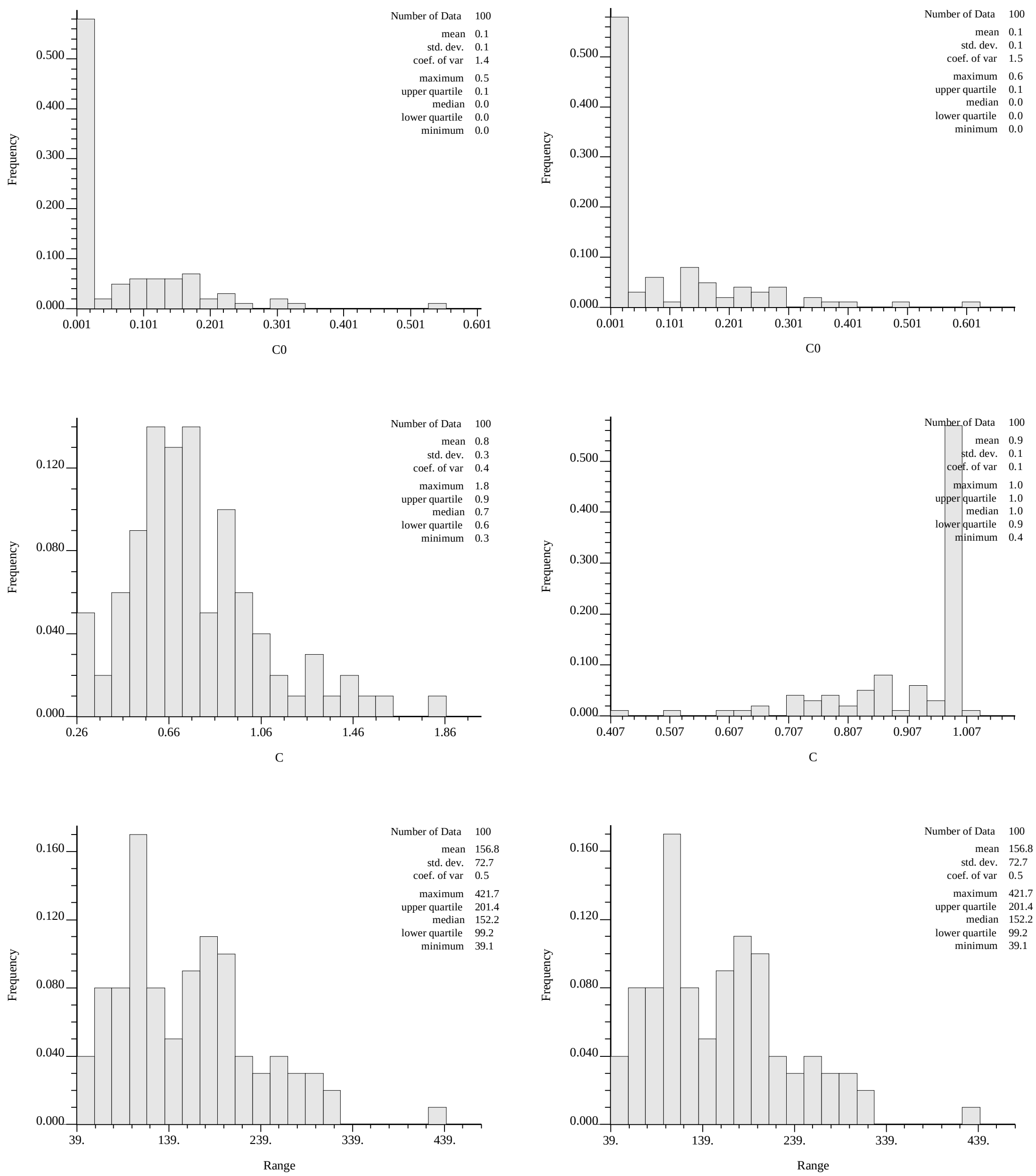


Figure 7. Histograms of variogram model parameters, Unstandardized (left), Standardized (right)

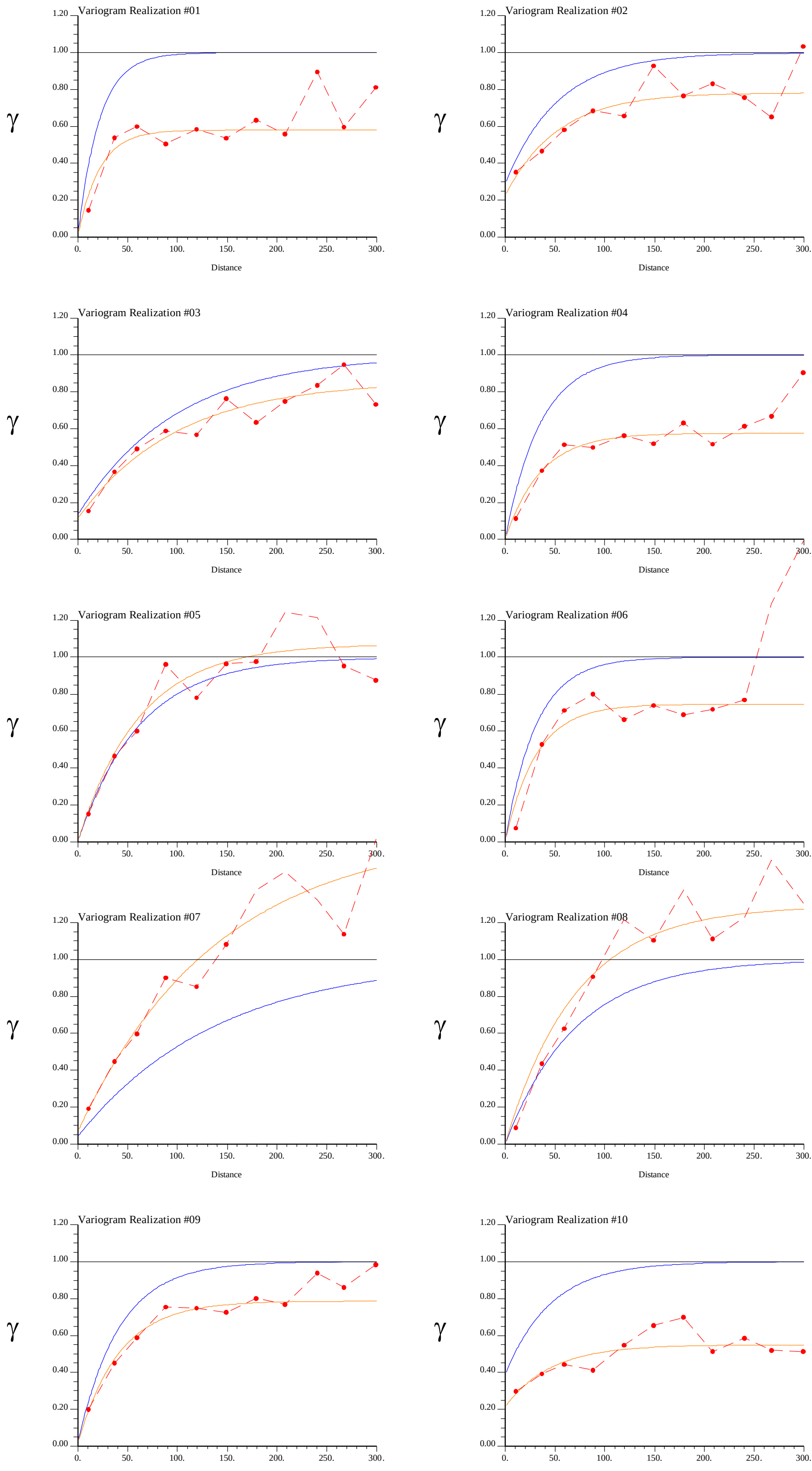


Figure 8. Generated experimental variograms and fitted variogram model for 10 first realizations.

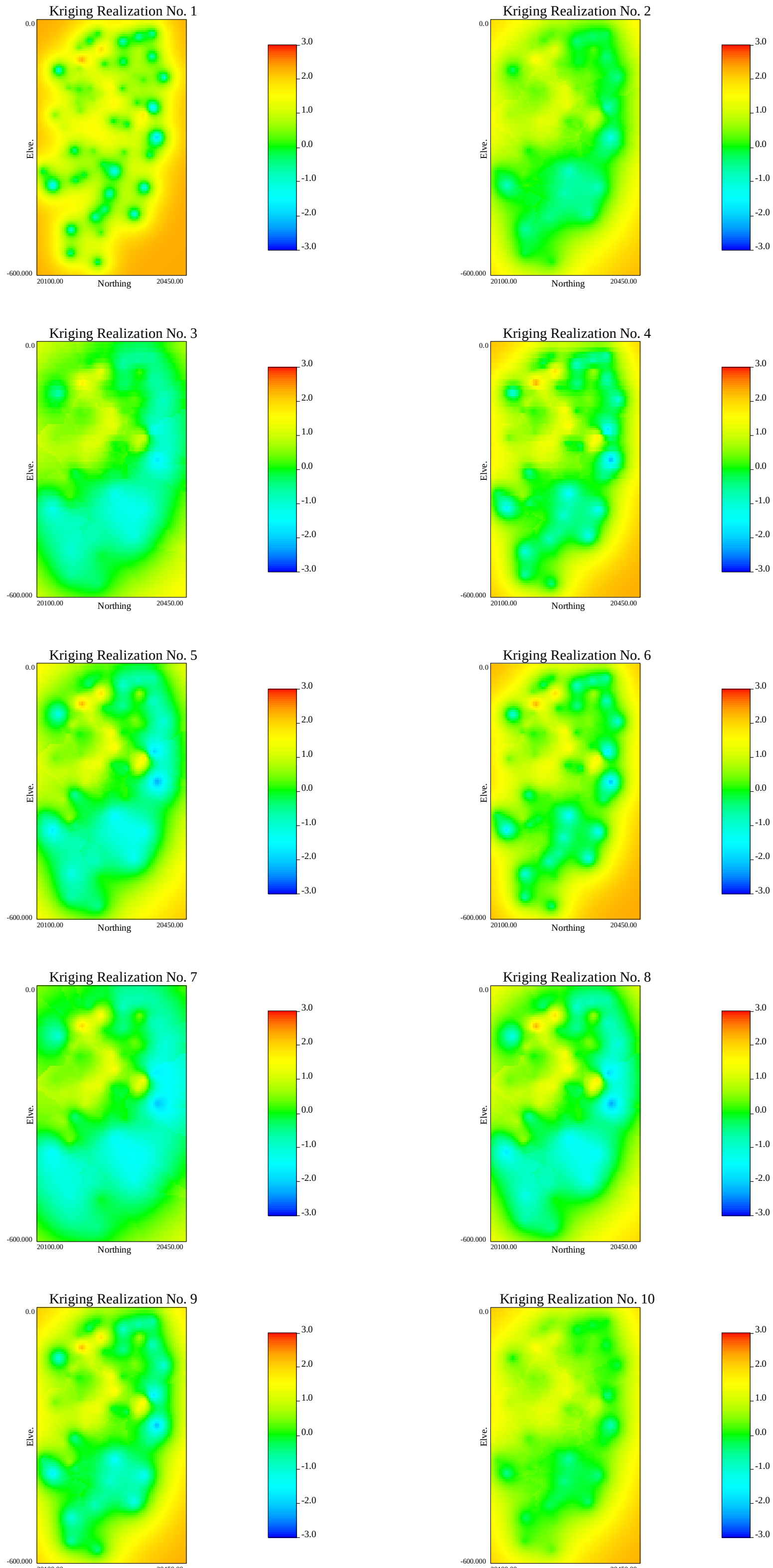


Figure 9. Kriging estimate results with different variogram models

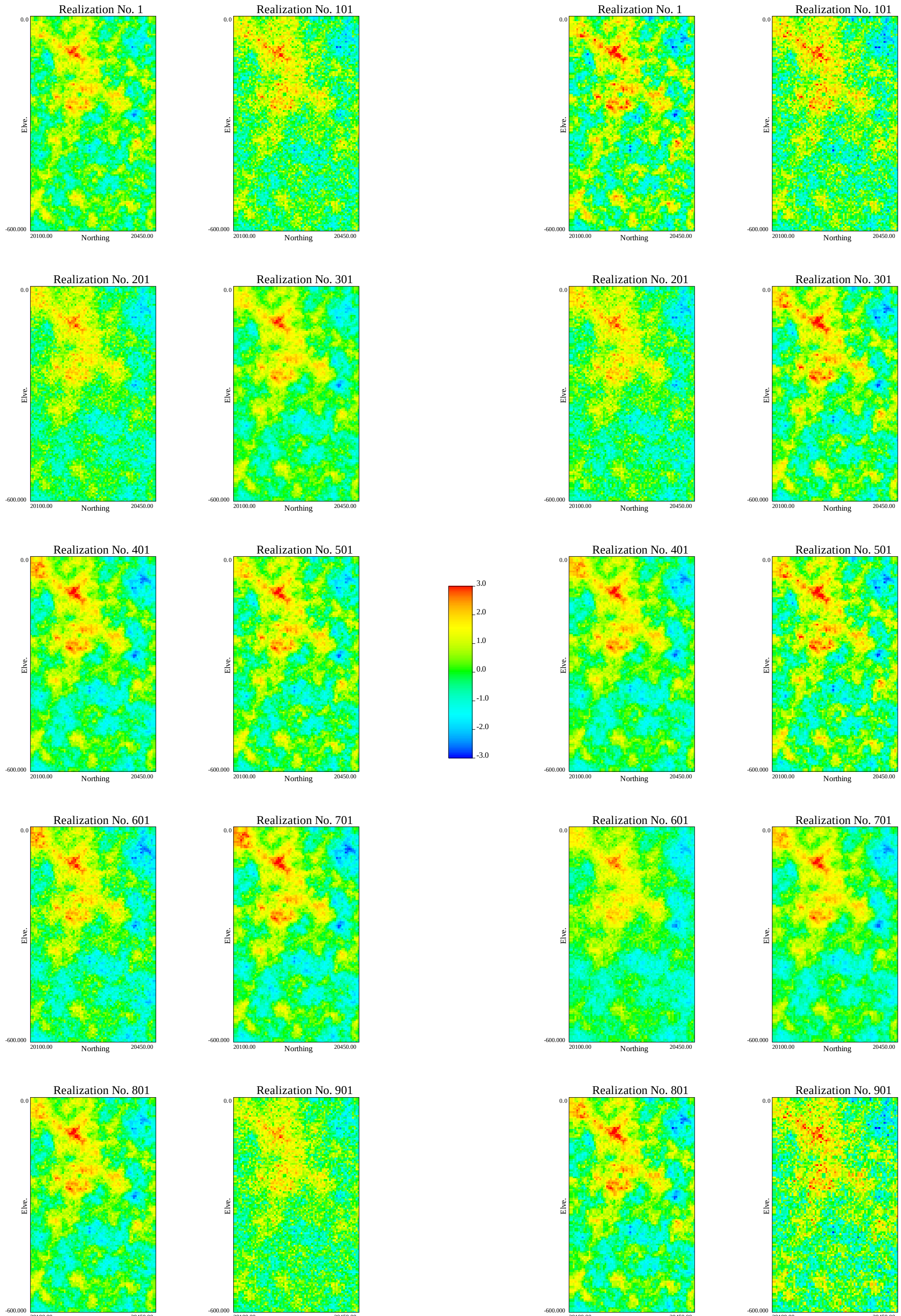


Figure 10. **`Sgsim`** realizations with unstandardized variogram models on left, Standardized on right.

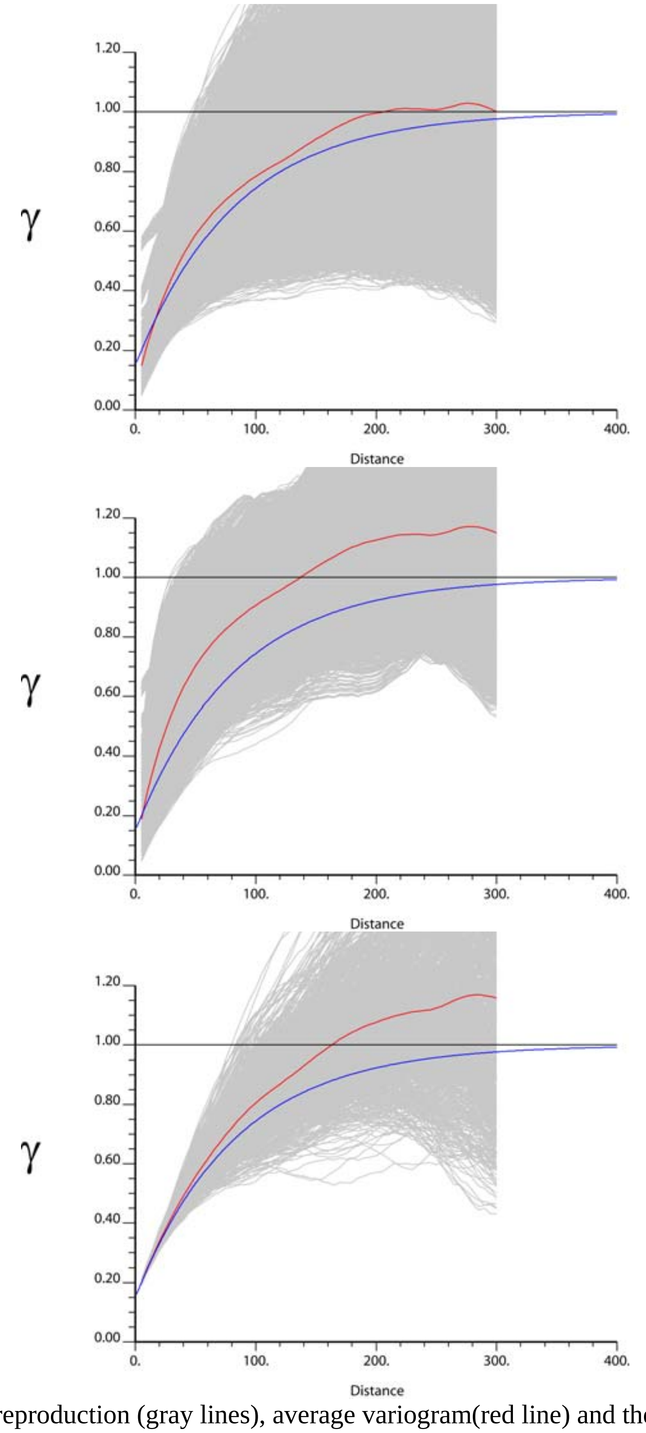


Figure 11. Variogram reproduction (gray lines), average variogram(red line) and the initial variogram (blue line); Top: Unstandardized variograms, and Middle: Standardized variograms and Bottom: **`sgsim`** W/ initial variogram models

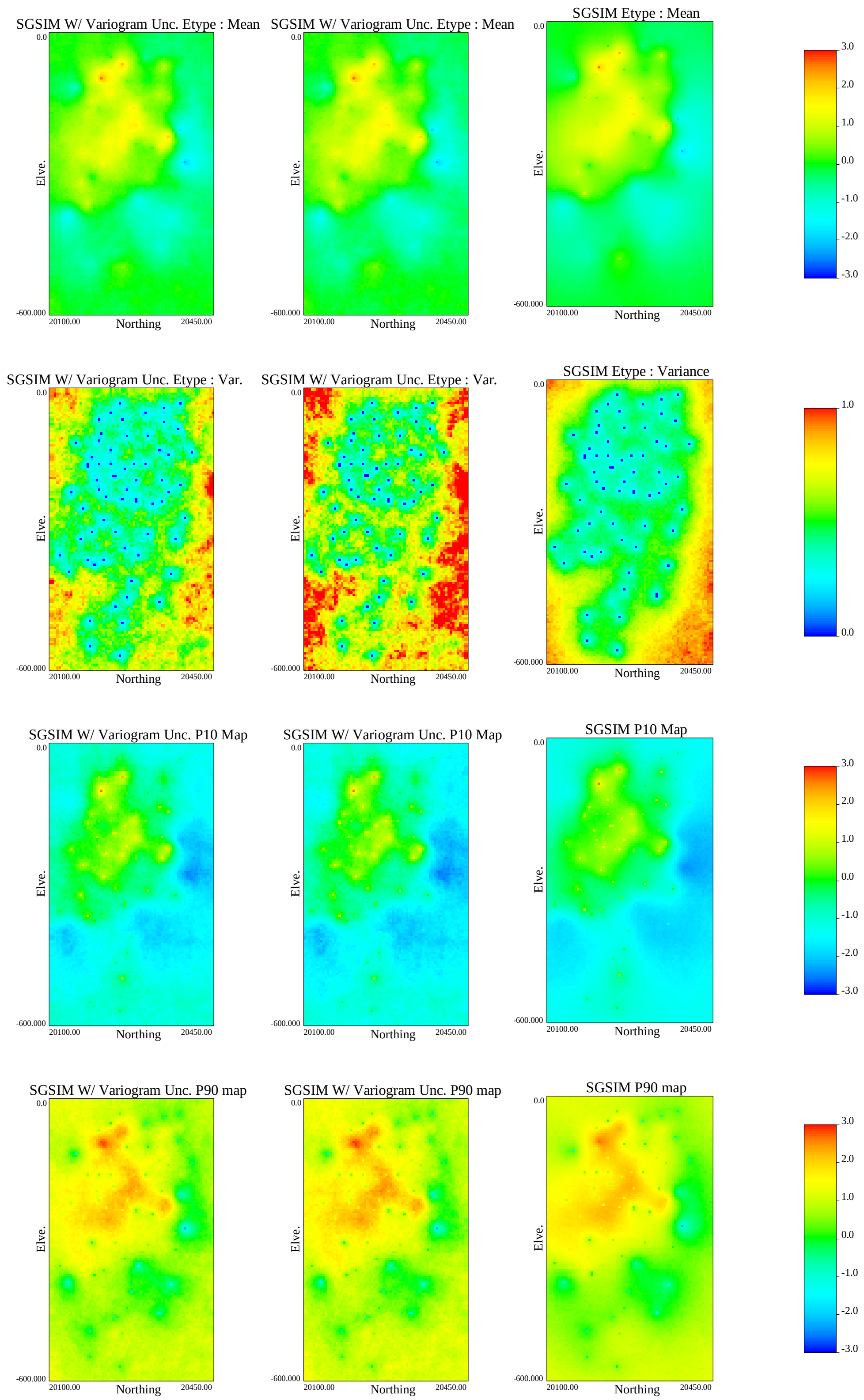


Figure 12. E-type Mean, variance, P10 and P90 maps; Left: W/ unstandardized variograms, Middle: Standardized Variograms and Right: **`Sgsim`** W/ initial variogram model

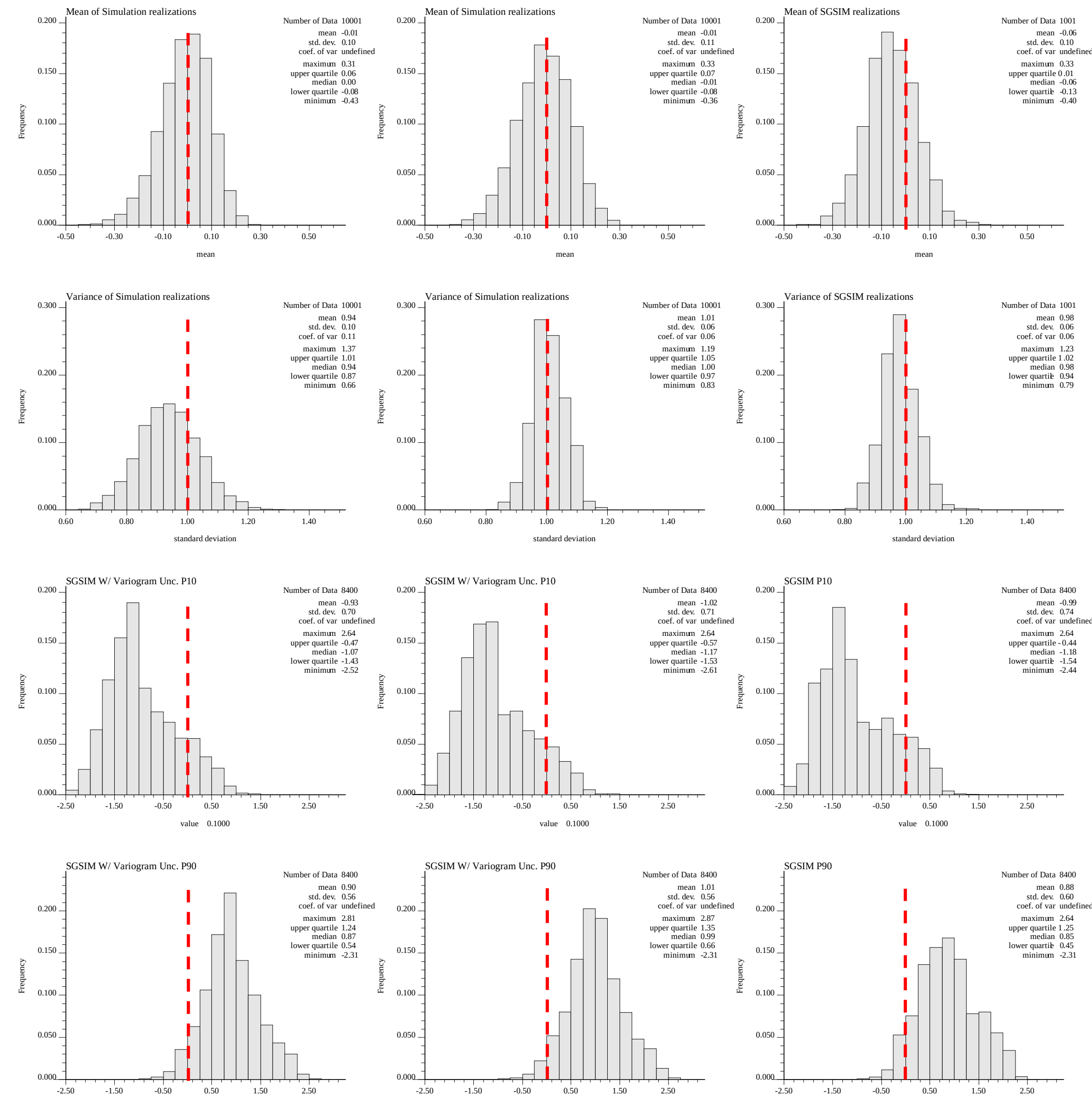


Figure 13. Histograms of Mean, Standard deviation, P10 and P90 values; Left: W/ Unstandardized Variograms, Middel: Standardized Variograms and Right: **`Sgsim`** W/ initial variogram