Updated Ultimate SGSIM

John G. Manchuk and Clayton V. Deutsch

In 2002 a flexible sequential simulation algorithm called Ultimate SGSIM was presented. Not all components of the program were thoroughly tested including full cokriging, the stepwise transform, self-healing, and kriging with multiple rock types. Moreover, many new techniques for geostatistical modeling have been developed since 2002. This work completes testing of Ultimate SGSIM to ensure its correctness, adds a few useful features, and incorporates two new modeling techniques: the intrinsic model for collocated cokriging and Bayesian updating for multiple secondary data. The advantage of the intrinsic model is the elimination of variance inflation caused by the Markov model. Bayesian updating has increased in popularity among industry due to its ease of implementation and ability to incorporate an abundance of information into geostatistical models.

1. Introduction

Ultimate sgsim (USGSIM) (Deutsch and Zanon, 2002) is an algorithm that was developed to offer a higher level of flexibility to practitioners using the sequential Gaussian simulation (SGSIM) approach for modeling (Deutsch and Journel, 1992). SGSIM was originally developed with a few options including simple kriging, ordinary kriging, collocated cokriging with a single secondary variable, locally varying mean, and external drift. Its limitations were identified and incorporated into USGSIM along with other modeling options including: simulation involving multiple rock-types with hard and soft boundaries; the stepwise conditional transformation (SCT) for multiple continuous variables (Leuangthong, 2003); full cokriging using a linear model of coregionalization (LMC); LU simulation (Davis, 1987); and locally variable angles of anisotropy. Two options mentioned in the 2002 USGSIM paper were never implemented and these include the self healing option to deal with variance inflation and locally varying correlation coefficients for collocated corkiging.

Many additional modeling techniques have been developed at the CCG, but are not immediately accessible to practitioners. Some developments have been added to USGSIM with minor changes to the parameters. Additions include the intrinsic model (IM) for collocated cokriging (Babak and Deutsch, 2007) and Bayesian updating (BU) (Zanon and Deutsch, 2004). The intrinsic model does not suffer from variance inflation and should replace the Markov model; however, it is more computationally expensive. It is added as an additional kriging option. BU is an approach to incorporate secondary data. The linear relationship between the variables being modeled (the primary data) and the secondary data is used to update the estimate and estimation variance from kriging. It is possible to incorporate many secondary data sources for each primary variable. The last addition was made to simplify the parameter inputs for cases involving an LMC. Users can either put the LMC definition directly into the parameter file or provide the name of the output file from `varfit_lmc`, a program that automatically models a valid LMC for full cokriging (Larrondo, Neufeld and Deutsch, 2003; Neufeld and Deutsch, 2004; Jewbali, 2009).

Not all of the options available in USGSIM were tested including full cokriging, SCT, modeling with multiple rock types, and the two options not implemented (self healing and locally varying correlations). Further testing is presented in this work and any identified bugs are reported and corrected. The self healing option is not implemented because the new intrinsic model does not suffer from variance inflation. The SCT is not tested since this transformation is performed external to USGSIM. Locally varying correlations are not implemented, but the correlation matrix for multivariate problems can now depend on the rock type. This paper is organized into sections that describe each component of USGSIM that is new or requires testing, the associated parameters, and testing results. Synthetic data based on the Walker Lake data was created for testing.

2. Parameter Changes

Most input parameters for USGSIM have remained unchanged from the 2002 paper. Parameters are input using blocks indicated by keywords and these include: *MAIN*, *SRCH*, *VARG*, *DATA*, *TRAN*, *TRND*, *SECV*, *ROCK*, and *MULT*. Changes have been made to the *VARG*, *SECV* and *MULT* blocks.

VARG: For multivariate modeling, performing full cokriging requires an LMC that can be quite large and tedious to input into a parameter file. For example, a problem with $M = 5$ variables involves $M(M+1)/2 = 15$ variogram models per rock type. A program called `varfit_lmc` for generating a valid LMC model from a set of experimental variograms writes the models to a text file. Instead of having to input these variograms directly into the parameter file for USGSIM, they can be read in directly from the varfit_lmc output files. The *VARG* block of parameters is defined below with 1 variogram model shown:

```
START OF VARG:
3                       -number of variograms
1 1 1                   -rock type, variable 1, variable 2
1 0.1                   -nested structures, nugget effect
1 0.9 45.0 0.0 0.0      -structure type, variance, ang1, ang2, ang3
   100.0  50.0  25.0    -a_hmax, a_hmin, a_hvert
```

To read in a varfit_lmc output file, instead of specifying the number of variograms, the number of files to read is identified instead and the value is made negative. Subsequent lines include the rock type and the associated file:

```
START OF VARG:
-2                      -number of varfit files
1 varfit_lmc_1.out      -rock type, varfit output file name
2 varfil_lmc_2.out      -rock type, varfit output file name
```

SECV: The original version of USGSIM allowed one secondary variable to be read in for collocated cokriging; however, when modeling using BU there are often several secondary variables involved. The *SECV* block of parameters has been changed so that multiple secondary variables can be read in:

```
START OF SECV:
3                 -number of secondary variables
Sec1.out          -file with secondary
1 2               - columns for variables
Sec2.out          -file with secondary
3                 - columns for variables
1                 -normal score transform
```

The program will read in secondary file names and column numbers until the number of columns found equals the number of secondary variables. An additional update has been added for the normal score transform option. The following options are available: 0 – no transformation; 1 – normal score transform the entire secondary variable; 2 – normal score transform the secondary variable by rock type if a rock type model was specified in the *ROCK* parameters.

MULT: Several multivariate options can be specified in the *MULT* parameter block including: 1 – independent; 2 – collocated; 3 – stepwise, and; 4 – full cokriging. Two additional options have been added including: 5 – the intrinsic model for collocated cokriging developed by Babak and Deutsch (2007) and; 6 – Bayesian updating. The ability to use a different correlation matrix for each rock type was also added to the parameters:

```
START OF MULT:
4                 -multivariate option
0                 -number of classes for stepwise
Stepcon.dat       -file with stepwise transform data
1 2 3             -columns for variables
1                 -rock type for correlation matrix
0.4 0.7 0.5       -correlation: 1-2 1-sec1 1-sec2
    0.6 0.3       -                 2-sec1 2-sec2
        0.7       -                        sec1-sec2
```

Multivariate options are numbered 1 through 6 as described previously. A correlation matrix for each rock type can be input by repeating the parameters from rock type down. To apply the same correlations to all rock types, input a rock type of zero. Note that if full cokriging is being done, the correlations between primary variables are reset based on the LMC. For BU, a second multivariate option can be input to indicate how primary variables are handled and includes independent, collocated, or full cokriging.

Currently, collocated cokriging, full cokriging, and the intrinsic model are limited to one secondary variable. Multiple secondary variables are only used when performing BU. However, it is possible to use multiple secondary in collocated cokriging with the other options. It may also be possible to use full cokriging together with the intrinsic model when multiple primary and secondary variables are involved. These are areas of future development.

3. Test Data

The Walker Lake data is used to conduct testing of `USGSIM`. There are two variables involved, U and V. A synthetic rock-type model with three rock types was constructed along with three exhaustive secondary data (Figure 1). Models constructed during testing are on a 260 by 300 grid of 1 m square cells. 195 locations were sampled using the exhaustive U and V models in a regular pattern with some noise added in to the locations (Figure 2). An arbitrary LMC is used:

- Rock type 1: a zero nugget effect spherical variogram is used with an azimuth of 90 degrees, maximum range of 100 m and minimum of 25 m. The cross variogram has a correlation of 0.87.
- Rock type 2: Similar to rock type 1 except an exponential structure is used with an azimuth of 45 degrees and the correlation is 0.63.
- Rock type 3: Similar to rock type 2 except an azimuth of 0 degrees and a 0.02 nugget effect is used. The correlation is 0.67.

4. Full cokriging

When multivariate modeling problems are encountered, full cokriging can be used to incorporate the complete correlation structure of all variables that is contained in the LMC. This approach has several advantages including: heterotopic sample data is possible; estimates consider more available information; there are no issues of variance inflation; and correlation is reproduced at all distances, rather than at $\mathbf{h} = 0$, where $\mathbf{h}$ is a lag vector. Kriging is more computationally demanding since the system of equations is M times larger than a univariate problem, with M being the number of variables involved. During simulation, a spatial search is done to find the k-nearest neighbours for kriging; so the system has rank $k \times M$. Generating a single realization can be time consuming and concessions must be made and usually involve reducing k; however, this leads to degradation of the statistical properties of the resulting models.

Full cokriging was tested using multiple rock types and no secondary data. A bug that caused some of the variogram models from the LMC of one rock type to get overwritten by variogram models of a different rock type was identified and fixed This was originally discovered by Shaun Hackett with Xstract Mining Consultants. It led to an invalid LMC and singular kriging systems. One way to determine if poorly conditioned cokriging systems are occurring is to plot the realization (Figure 3). Systems that are singular or indefinite can lead to extreme estimates and negative variance and are set to undefined by USGSIM. If the debugging level is set to 1 or larger in the *MAIN* parameter block, these error conditions are written to the debug file. Another bug that was identified involved the correlation matrix for full cokriging. It was only being reset based on the LMC when a secondary variable was being used. If the incorrect correlation matrix was input in the *MULT* block, then it was being used for correlating the realizations at $\mathbf{h} = 0$.

Generating a realization of variables U and V took 10.84 seconds using 20 nearest neighbours. Each kriging system has rank 40 for this bivariate case. The resulting maps are shown in Figure 4 that shows a clear difference in anisotropy in each rock type as expected. Correlations between U and V in each rock type are 0.82, 0.61, and 0.70 in rock types 1, 2 and 3 respectively. This is close to the input values of 0.87, 0.63, and 0.67.

5. Intrinsic Model

The IM is a cokriging approach to incorporate secondary data in a similar way as collocated cokriging; however, instead of using the secondary datum strictly at the grid block being estimated, the secondary data at all conditioning data locations are used. For one secondary variable, this results in the system of equations in Equation (1), where $\mathbf{R}$ is the covariance matrix between all conditioning data, $\mathbf{r}$ is the vector of covariance between the estimate and conditioning data, ρ is the correlation between the primary and secondary, λ_p and λ_s are the solution weights applied to the primary and secondary, and λ_0 is the weight applied to the collocated secondary.

$$\begin{bmatrix} \mathbf{R} & \rho\mathbf{R} & \rho\mathbf{r} \\ \rho\mathbf{R}^T & \mathbf{R} & \mathbf{r} \\ \rho\mathbf{r}^T & \mathbf{r}^T & 1 \end{bmatrix} \begin{bmatrix} \lambda_p \\ \lambda_s \\ \lambda_0 \end{bmatrix} = \begin{bmatrix} \mathbf{r} \\ \rho\mathbf{r} \\ \rho \end{bmatrix} \quad (1)$$

The disadvantage is the increased computational cost when compared to traditional collocated cokriging. Kriging with one primary and one secondary variable is as expensive as full cokriging with two primary variables collocated with a secondary. Generating a realization of U and V using Secondary A with the IMC and 20 nearest neighbours took 15.1 seconds. Correlations between simulated U and V are 0.81, 0.64, and 0.77 for rock types 1, 2 and 3 respectively. Correlations between the primary and Secondary A are summarized in Table 1.

Table 1: Correlations between primary and Secondary A.

Rock Type	Pair	Target	Simulated
1	U-Secondary A	0.264	0.309
	V-Secondary A	0.226	0.234
2	U-Secondary A	0.286	0.265
	V-Secondary A	0.276	0.222
3	U-Secondary A	0.512	0.545
	V-Secondary A	0.483	0.490

Correlations for all cases are close. The advantage of the intrinsic model is that it does not cause variance inflation as with collocated cokriging. 25 realizations were generated and the variance was calculated. Using the intrinsic model, the average variance for U and V was 0.96 and 0.88; using collocated cokriging, the variances were 1.02 and 0.92. Collocated cokriging took 4.7 seconds per realization using the 20 nearest neighbours. Variances lower than 1 are caused by the domain size of individual rock types relative to the variogram. To confirm, the same realizations were generated using a range of 10 m instead of 100 m in the major direction. Variances using the IM were 0.99 and 0.95 and using collocated cokriging were 1.18 and 1.12 for variables U and V respectively. An ad-hoc procedure such as a variance reduction factor is required for the later.

6. Bayesian Updating

The multivariate methods including full cokriging, collocated cokriging, and IM are limited to cases with one secondary variable in their current implementations. In cases when several secondary variables are available, BU is an approach to utilize all of them to better predict the primary variables. In BU, the secondary variables are merged into a likelihood based on how they correlate with the primary variables. Deriving the likelihood distributions is identical to kriging (Equation (2)), where $\mathbf{P}$ is the correlation matrix between the secondary, $\boldsymbol{\rho}_k$ is the vector of correlations between the k^{th} primary variable and the secondary, and $\boldsymbol{\nu}_k$ are the solution weights. Defining $\mathbf{s}(\mathbf{u})$ as the vector of secondary values at a location $\mathbf{u}$, the likelihood mean, $L_k(\mathbf{u})$, is defined by Equation (3) and the likelihood variance, $L_k^2(\mathbf{u})$, by Equation (4), where N is the number of secondary variables. Recall M is the number of primary variables.

$$\mathbf{P}\mathbf{v}_k = \boldsymbol{\rho}_k, k = 1,...,M \quad (2)$$

$$L_k(\mathbf{u}) = \sum_{j=1}^{N} \nu_{kj} s_j(\mathbf{u}), k = 1,...,M \quad (3)$$

$$L_k^2(\mathbf{u}) = 1 - \sum_{j=1}^{N} \nu_{kj} \rho_{jk}, k = 1,...,M \quad (4)$$

As a realization of USGSIM proceeds, distributions of uncertainty for the primary variables are computed using simple kriging or cokriging. These are denoted $\mu_k(\mathbf{u})$ and $\sigma_k^2(\mathbf{u})$ for the estimate and estimation variance respectively. These are often called the prior mean and prior variance when implementing BU. To compute the updated mean and updated variance from the likelihood and prior, Equations (5) and (6) are used, where $U_k(\mathbf{u})$ is the updated mean and $U_k^2(\mathbf{u})$ the updated variance.

$$U_k(\mathbf{u}) = \frac{L_k \sigma_k^2 + \mu_k L_k^2}{(1 - L_k^2)(\sigma_k^2 - 1) + 1}, k = 1,...,M \quad (5)$$

$$U_k^2(\mathbf{u}) = \frac{L_k^2 \sigma_k^2}{(1-L_k^2)(\sigma_k^2 - 1) + 1}, k = 1, ..., M \quad (6)$$

The Likelihood and updating computations are incorporated after kriging. In the case of multiple primary variables, correlated Gaussian values are simulated and conditioned using the updated distributions. In cases where collocated cokriging or full cokriging are used to handle multiple primary variables, no secondary variables are used in building and solving the kriging equations. There is a risk of incorporating a secondary variable twice, resulting in a false reduction in the estimation variance. Other features such as a trend and locally varying mean are possible. BU is another form of collocated cokriging and can lead to global variance inflation.

A map of the likelihood mean and variance calculated using the three secondary variables is shown in Figure 5 for variable U. The map is very similar for variable V since the correlation between U and V is high. Features from the secondary data can be seen in the likelihood mean. The likelihood variance is constant within rock types, with values of 0.208 in rock type 1, 0.827 in rock type 2, and 0.714 in rock type 3.

7. Conclusions

Several new features have been added to `USGSIM` and tested. During testing, two bugs worth reporting were identified: 1 – the LMC was being overwritten when cokriging with multiple rock types; 2 – the correlation matrix was not being reset from the LMC when cokriging without a secondary variable. All tests done in this work involved multiple rock types and did not reveal additional problems. Enhancements to USGSIM include the following:

- Ability to read the LMC from `varfit_lmc` output files.
- Correlation matrices can now depend on rock type.
- The intrinsic model for collocated cokriging that does not suffer from variance inflation. This is limited to one secondary variable.
- Bayesian updating. Kriging of primary variables can be done independently or with full cokriging. Simulation can be done independently or correlated. Multiple secondary variables can be specified in the parameter file.
- The normal score transform for secondary variables can be done as a whole or by rock type.

All enhancements have been tested; however, users must be aware that `USGSIM` is research code. There will be instances where a certain configuration of parameters causes the program to fail. Such instances indicate something is wrong in the parameters or a technique is attempted that is not available. Future testing will reveal these instances and can be incorporated into the program with the exact problem reported to users. It is very useful to the CCG that anyone using the program reports problems after thoroughly checking that it is not a simple problem with the parameter file.

There is always room for improvement, even for an *ultimate* program. Many modeling techniques exist and can be incorporated. `USGSIM` performs a variety of some of the more popular techniques. A few areas for further enhancement were identified in this work. Full cokriging and collocated cokriging are limited to one secondary. The program has been enhanced to read in multiple secondary and their use in collocated or full cokriging could be added. Furthermore, the intrinsic model could be coded to work with multiple secondary. The idea of full cokriging with multiple secondary using the intrinsic model was mentioned; however, the conditioning of kriging systems is a concern. This would be the ideal way to incorporate multiple primary and secondary, but it would be more computationally expensive; however, it may be possible to reduce this cost with parallelization (Boisvert, 2010).

References

Boisvert, J.B., 2010, Parallel Programming. Centre for Computational Geostatistics, 12, 409

Babak, O. and Deutsch, C.V., 2007, An intrinsic model of coregionalization that solves variance inflation in collocated cokriging. Centre for Computational Geostatistics, 9

Davis, M.W., 1987, Production of conditional simulations via the LU triangular decomposition of the covariance matrix. Mathematical Geology 19(2): 91-98

Deutsch, C.V. and Journel, A.G., 1992, GSLIB: geostatistical software library and user's guide. Oxford University Press, 384 pp.

Deutsch, C.V. and Zanon, S., 2002, Ultimate sgsim: non-stationary sequential Gaussian cosimulation by rock type. Centre for Computational Geostatistics, 4

Jewbali, A., 2009, Finding the nearest positive definite matrix for input to semi automatic variogram fitting (varfit_lmc). Centre for Computational Geostatistics, 11, 402

Larrondo, P.F., Neufeld. C., and Deutsch, C.V., 2005, Varfit: A program for semi-automatic variogram modeling: Centre for Computational Geostatistics, 5

Leuangthong, O., 2003, Stepwise conditional transform for multivariate geostatistical simulation. University of Alberta, 204 pp.

Neufeld, C. and Deutsch, C.V., 2004, Developments in semi-automatic variogram fitting. Centre for Computational Geostatistics, 6=

Zanon, S.D. and Deutsch, C.V., 2004, Direct prediction of reservoir performance with Bayesian updating under a multivariate Gaussian model. In 5th Canadian International Petroleum Conference, Calgary, Alberta, Canada, June 2004. Petroleum Society of the Canadian Institute of Mining, Metallurgy & Petroleum

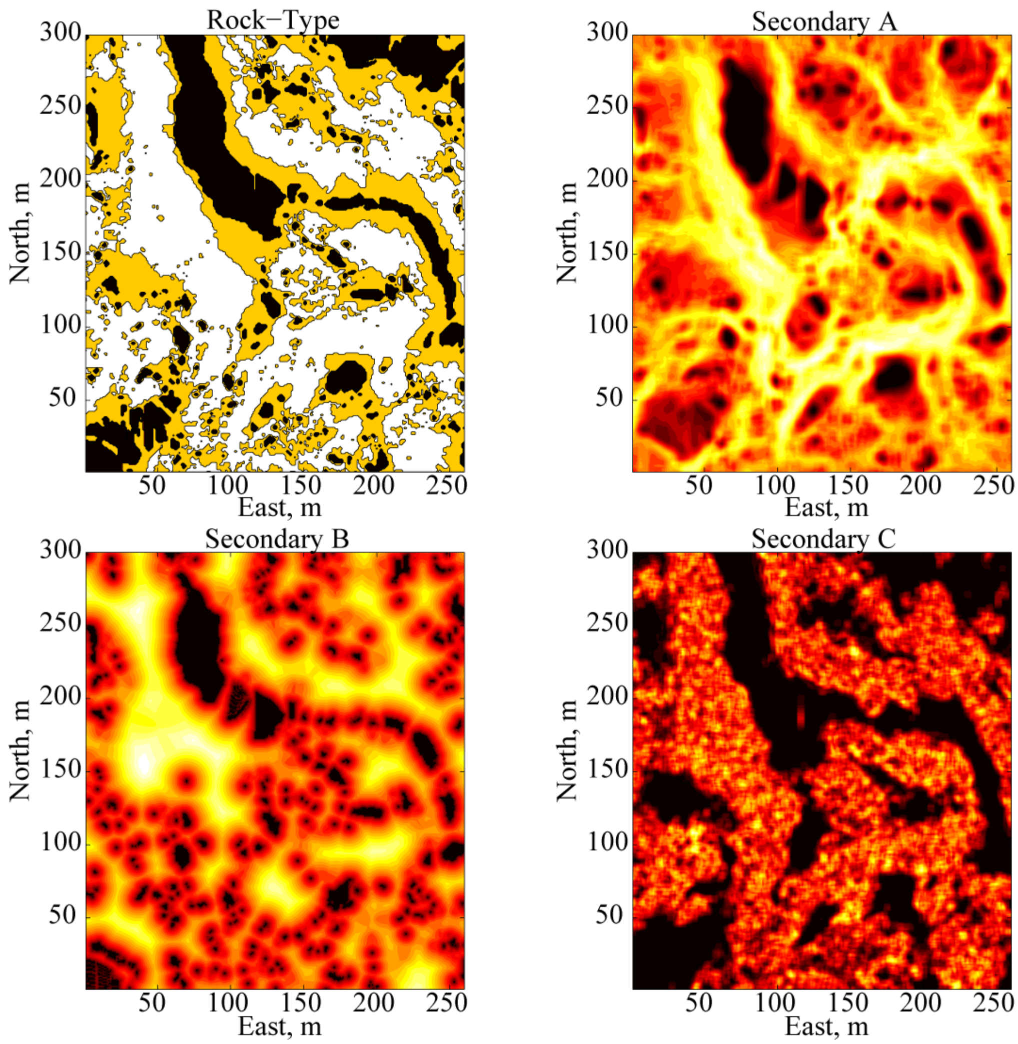

Figure 1: Synthetic rock type model and secondary fields. Rock types: 1-black, 2-gray, 3-white.

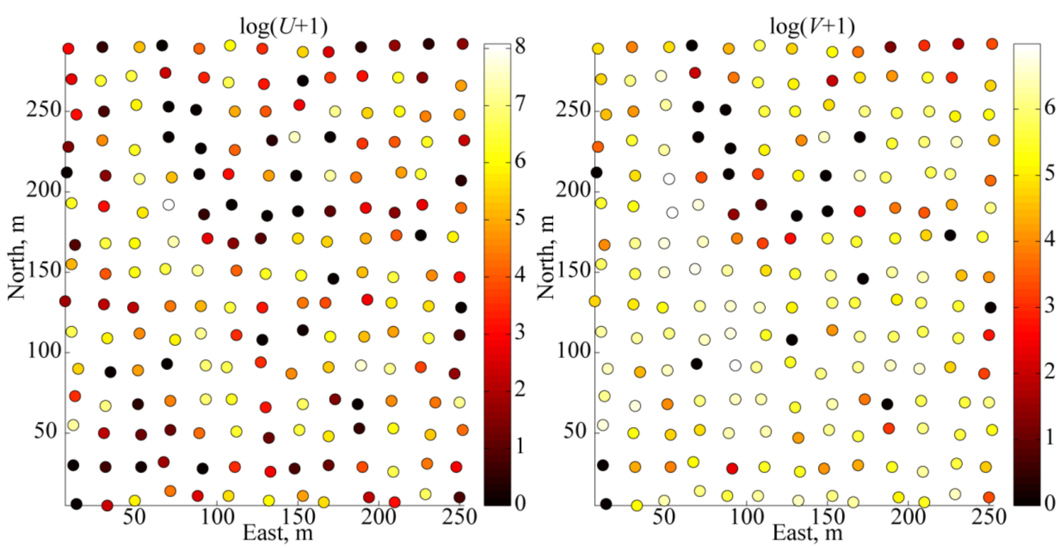

Figure 2: Walker Lake sample data.

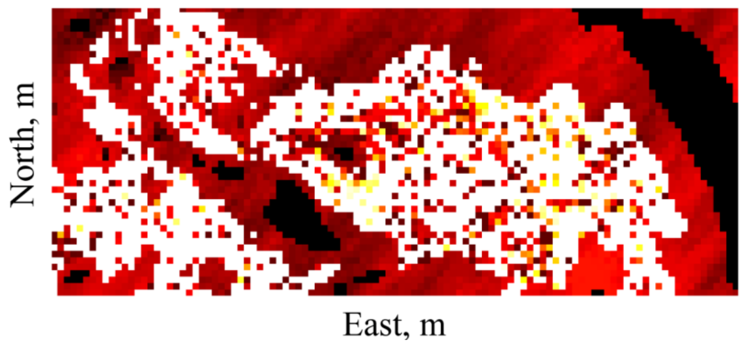

Figure 3: Portion of a realization of variable U with poorly conditioned systems in rock type 3.

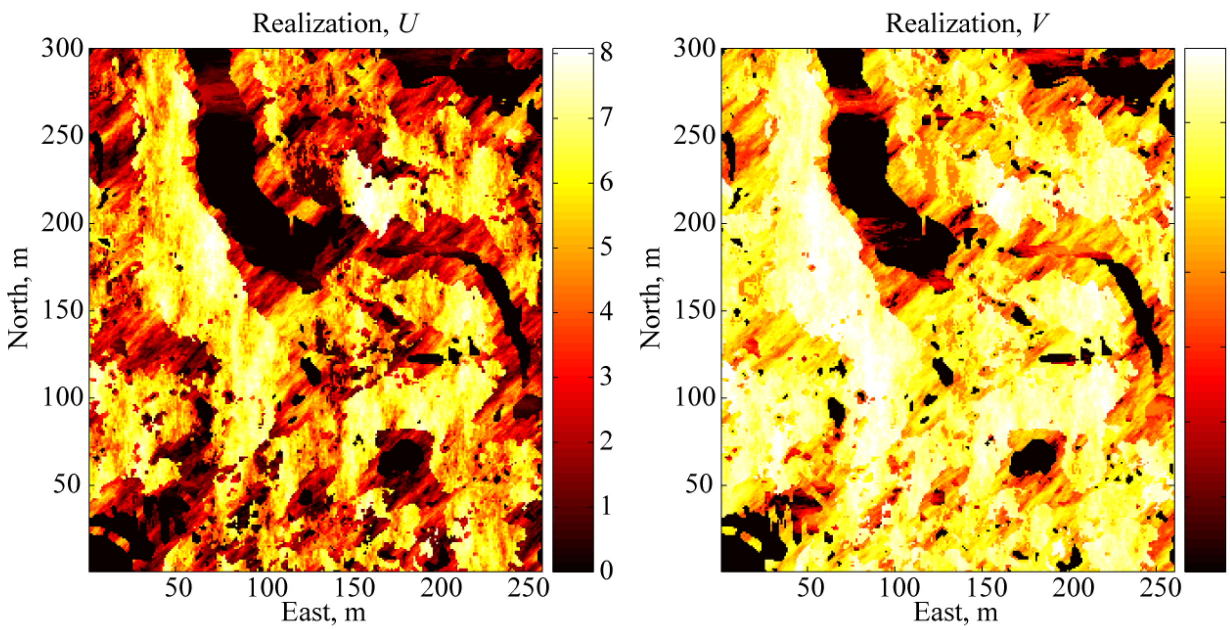

Figure 4: A realization using full cokriging by rock type.

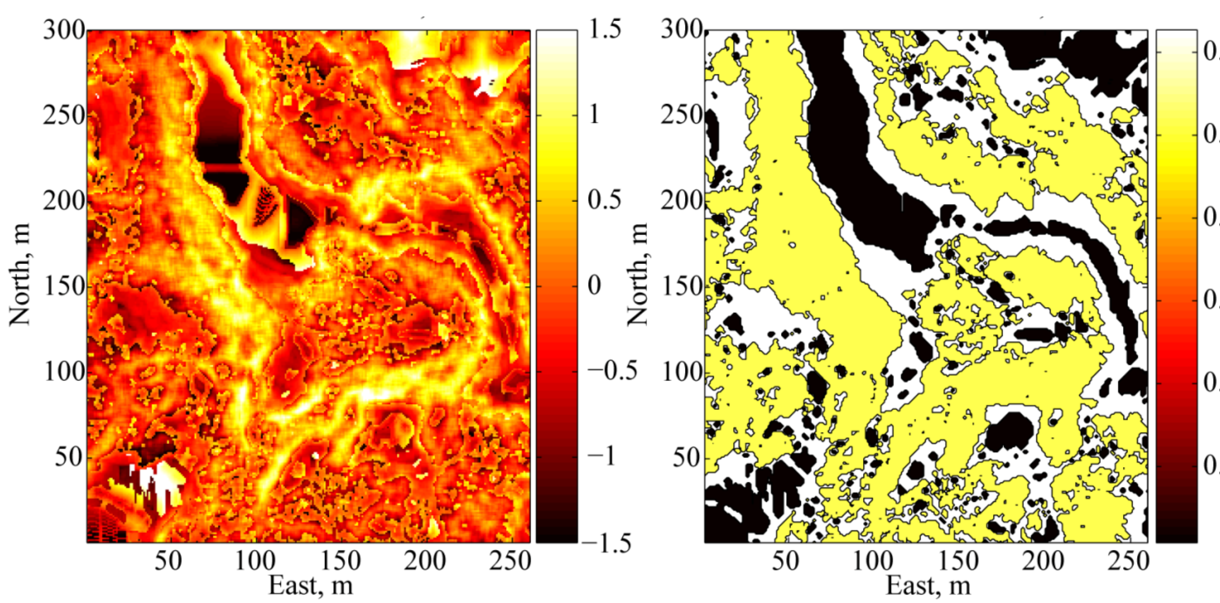

Figure 5: Likelihood mean (left) and variance (right) for variable U.