An efficient and accurate algorithm for generating spatially-correlated random fields

Jiannong Fang*,† and Laurent Tacher

GEOLEP-ICARE-ENAC, Ecole Polytechnique Fédérale de Lausanne,
1015 Lausanne, Switzerland

SUMMARY

This paper presents a new computer algorithm for generating spatially-correlated random fields. Such fields are often encountered in hydrology and hydrogeology and in the earth sciences and used as inputs for Monte Carlo simulations. The algorithm is designed by using a multilevel grid strategy and combining the matrix decomposition (MD) method and the screening sequential simulation (SSS) method. The idea originates from the facts: (i) the MD method accounts for all possible nodal correlation values and hence the accuracy of the method is high, but it can be extremely computationally intensive for fine meshes with large number of nodes, and (ii) the SSS method is more efficient because a small search neighbourhood for conditioning can be used due to the screening effect of measurements, however, for large separation distances, correlation values from the SSS method is significantly inferior to the values obtained by the MD method. Numerical examples are presented to demonstrate the new method. It is shown that the presented method is much more efficient than the MD method and more accurate than the SSS method. The new algorithm is also versatile: it can directly simulate fields of irregular geometry without additional efforts. Copyright © 2003 John Wiley & Sons, Ltd.

KEY WORDS: random fields; stochastic processes; sequential simulation; Monte Carlo simulation

1. INTRODUCTION

Simulation techniques for generating random fields have many important applications in hydrology and hydrogeology, as well as in the earth sciences [1]. For example, they are often used to generate input parameter fields for the Monte Carlo simulation (MCS) of some physical process. MCS is a versatile mathematical tool capable of handling stochastic problems where all other methods such as perturbation methods fail. The implementation of the method consists of numerically generating a population of the random parameters by a generation method (generator), solving the deterministic problem associated with each member of that population, and obtaining a population of the random response quantities. This population

---

*Correspondence to: Jiannong Fang, GEOLEP-ICARE-ENAC, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland.
† E-mail: jiannong.fang@epfl.ch

can then be used to obtain statistics of the prediction by ensemble averaging. As a general requirement, the generated realizations of the random parameters should honor the prescribed (measured or inferred) statistics and the generator should be computationally efficient.

A variety of different methods of generating realizations of random field exists, principally the spectral method [2], the matrix decomposition method [3–5], the turning bands method [6–9], and the screening sequential simulation method [10–12]. Among them, the matrix decomposition (MD) method and the screening sequential simulation (SSS) method are simple to implement in practice and suitable for conditional simulation. Furthermore the two methods have the advantage of allowing fields of irregular geometry without additional computational effort. The MD method essentially involves the construction of the covariance matrix of the random field and its subsequent Cholesky decomposition into an upper and lower triangular matrix. The lower triangular matrix is then multiplied by a vector of uncorrelated standard Gaussian variables, producing a stochastic field possessing the prescribed covariance function. It is recognized that the covariance function serves as an input quantity in the model. The accuracy of the method is supposed to be high due to the fact that the generation accounts for all possible nodal correlation values. However, a practical limit on the number of nodes emerges because of the Cholesky decomposition and the storage of the matrix. An advantage of the method is that the matrix decomposition step needs only to be performed once for Monte Carlo simulations because only backward substitution is required to generate each additional realization. The SSS method is based upon the fact that a one-dimensional Gaussian random field with exponential correlation function is a Markov process, which means that the value at one point only has to be conditioned on values at neighbour points. Omre et al. [11] have shown that such a process may be simulated very efficiently in $O(n)$ elementary operations. Correspondingly, the Cholesky decomposition and product in the MD method require $O(n^3)$ and $O(n^2)$ operations, respectively. In two- and three-dimensional problems the Markov property is not uniquely defined due to lack of ordering in the domain. However, a natural way of generalizing the one-dimensional method is to let the value at one point be conditioned on points within a small search neighbourhood. The conditioning is based upon the prescribed correlation function. The screening effect of measurements assures high accuracy of the SSS method for the predictions of the variance and correlation functions of generated fields. However, for large separation distances, correlation values from the SSS method is significantly inferior to the values obtained by the MD method.

Our goal here is to develop a generator for spatially-correlated random functions which will be fast and effective and reproduce accurately the prescribed spatial statistics without introducing any artificial effects. This paper presents the method and shows that the above goals have been obtained through the adoption of a multilevel grid strategy and a proper combination of the two methods briefly described above. In Section 2, the theoretical background and the implementation of the new algorithm are presented. Section 3 gives numerical examples to demonstrate the method and discusses some computational issues. Finally, a summary of the results is given in Section 4.

2. THEORY AND IMPLEMENTATION

The purpose of the new algorithm is to generate a stationary, real-valued field $y(\mathbf{x})$ where $\mathbf{x}$ denotes the spatial co-ordinate. The spatial variability of $y(\mathbf{x})$ is modelled as a space
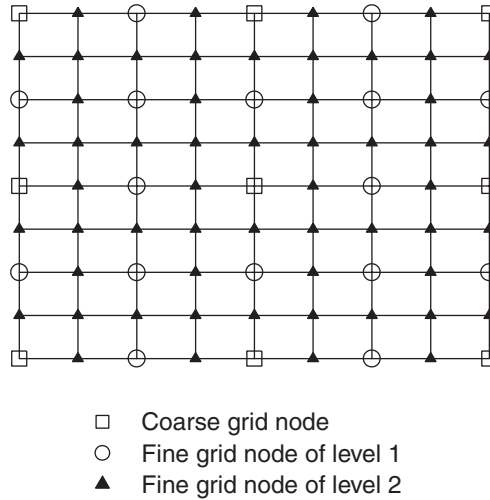
□   Coarse grid node
○   Fine grid node of level 1
▲   Fine grid node of level 2

Figure 1. Example of domain discretization by a multilevel grid.

random function (SRF) $Y(\mathbf{x})$. Here and subsequently boldface letters denote vectors, capital letters denote SRFs and lower case letters denote realizations of $Y(\mathbf{x})$. $Y(\mathbf{x})$ is characterized statistically by its moments, namely, its expected value:

$$\langle Y(\mathbf{x}) \rangle = m_Y(\mathbf{x}) \tag{1}$$

where $\langle \bullet \rangle$ denotes the expectation operator, and its covariance function:

$$C_Y(\mathbf{x}, \mathbf{x}') = \langle [Y(\mathbf{x}) - m_Y(\mathbf{x})][Y(\mathbf{x}') - m_Y(\mathbf{x}')] \rangle = C_Y(\mathbf{r} = \mathbf{x} - \mathbf{x}') \tag{2}$$

In this study we assume the random field Y follows a multivariate Gaussian distribution, for which the entire distribution is then specified by (1) and (2). While Y is generally a continuous function, the proposed algorithm consists in generating the Y field discretely over a pre-generated arbitrary multilevel grid. An example of such a grid is given in Figure 1. The grid and the associated information, i.e. co-ordinates and indices of nodes, can be supplied either by modern mesh generation softwares [13] or by user-defined programs.

At the first step, a realization $y(\mathbf{x})$ on the nodes of the coarse grid is generated using the MD method. Let $\mathbf{R} = \{R_{ij}\}$ be the covarince matrix for the random field; that is, $R_{ij} = C_Y(\mathbf{x}_i, \mathbf{x}_j)$, where $\mathbf{x}_i$ and $\mathbf{x}_j$ are the co-ordinates of node i and j in the coarse grid. It is obvious that $\mathbf{R}$ is a symmetric, positive-definite, square $N \times N$ matrix (N is the total number of nodes in the coarse grid). Therefore the matrix can be factorized as $\mathbf{R} = \mathbf{LL}^T$ by the Cholesky decomposition method, where $\mathbf{L}$ is the lower triangular decomposition of $\mathbf{R}$. The matrix $\mathbf{L}$ is then multiplied by a vector $\mathbf{z}$ of N uncorrelated, standard Gaussian variables, producing the vector $\mathbf{y} = \mathbf{Lz}$, which, after adding the mean values, represent values of the random field on the nodes of the coarse grid. It can be readily shown that the nodal values exactly recover the prescribed covariance structure. Once $y(\mathbf{x})$ on the nodes of the coarse grid is generated, it is considered as a datum and it will be used to condition the y values on the nodes of the fine grids. It is important to point out that the number of nodes in the coarse grid N must be

chosen in such a way that the performance of the Cholesky decomposition is computationally affordable at this stage.

At the next step, generation of a realization on the nodes in the fine grid of level 1 is performed by using the SSS method. Choosing a starting point, say $\mathbf{x}_{N+1}$, $Y(\mathbf{x}_{N+1})$ is conditioned on the neighbouring N nodal values from the coarse grid generated in the previous step, using the Gaussian conditioning procedure. The conditional expected value of $Y(\mathbf{x}_{N+1})$ is given by

$$\langle Y^c(\mathbf{x}_{N+1})\rangle = m_Y(\mathbf{x}_{N+1}) + \sum_{j=1}^{N} \lambda_j(\mathbf{x}_{N+1})[y(\mathbf{x}_j) - m_Y(\mathbf{x}_j)] \tag{3}$$

while the conditional variance is given by

$$\sigma_Y^{2c}(\mathbf{x}_{N+1}) = \sigma_Y^2 - \sum_{j=1}^{N} \lambda_j(\mathbf{x}_{N+1})C_Y(\mathbf{x}_{N+1}, \mathbf{x}_j) \tag{4}$$

Once these last two statistics are computed, a realization $y(\mathbf{x}_{N+1})$ is generated from any standard Gaussian random number generator which uses these two statistics as target statistics. The generated value is added to the data base for conditioning at later steps. The only difference between consecutive steps is in the number of data used for conditioning. The interpolation coefficients λ_j are the solution of the following linear system [14]:

$$\sum_{j=1}^{N} \lambda_j(\mathbf{x}_{N+1})C_Y(\mathbf{x}_j, \mathbf{x}_i) = C_Y(\mathbf{x}_{N+1}, \mathbf{x}_i); \quad i = 1, \ldots, N \tag{5}$$

The repetitive solution of (5) for each new node is computationally the most demanding step in the algorithm. The computational effort in this step can be alleviated significantly when considering the following points: (i) The decay of the covariance function is rapid due to the fact that the integral scale of Y is finite; (ii) the interpolation coefficients λ_j associated with any node $\mathbf{x}_i$ do not depend on the actual values of $y(\mathbf{x}_i)$ and $y(\mathbf{x}_j)$ but rather on the distances $\mathbf{x}_i - \mathbf{x}_j$, $j = 1, \ldots, N$; (iii) The large-scale spatial correlation is already adequately respected at an early stage by the MD method, the Y values for the fine grids can be conditioned only on the points nearby by utilizing the screening effect often employed in geostatistical applications. As a result, $Y(\mathbf{x}_{N+1})$ needs to be conditioned only to a limited search neighbourhood, and furthermore, if a fixed searching radius with a fixed generation path is used, the spatial configuration of conditioning data for each node in the fine grids is then also fixed, which leads to the fact that the set of coefficients needs to be computed once and only once and can then be used repeatedly for different realizations. A salient question is the size of the searching radius. This was resolved in our study through numerical experiments.

As a summary, the simulation procedure of the new algorithm consists of the following steps:

  (i) Compute the covariance matrix $\mathbf{R}$, factorize it as $\mathbf{R} = \mathbf{LL}^T$, and store the lower triangular matrix $\mathbf{L}$.
 (ii) Define a fixed path through all nodes in the fine grids to be simulated (normally just follow the node numbering). Search sequently the available conditioning data for each node inside the search neighbourhood. Then compute and store the interpolation coefficients for each node by solving the linear system (5).

(iii) Generate a realization $y(\mathbf{x})$ on the nodes in the coarse grid by first generating a vector $\mathbf{z}$ of N uncorrelated, standard Gaussian variables and then performing the multiplication $\mathbf{Lz}$ and adding the mean values.

(iv) Compute $\langle Y^c(\mathbf{x}_{N+1})\rangle$ and $\sigma_Y^{2c}(\mathbf{x}_{N+1})$ by using Equations (3) and (4) with the interpolation coefficients computed at step (ii). Generate a realization $y(\mathbf{x}_{N+1})$ using the conditional mean and variance just calculated as target statistics.

(v) Add $y(\mathbf{x}_{N+1})$ to the data base for conditioning of $y(\mathbf{x})$ on the next node.

(vi) Move to next node according to the path defined in step (ii) and go to step (iv).

(vii) Repeat the steps (iv)–(vi) till all nodal values are created. Go back to step (iii) for another realization.

The algorithm presented above is suitable for irregular grids such as those encountered often in the finite element discretization of irregular domains. It can also be used to the case of conditional simulation which accounts for the actual measured values of Y at a finite set of locations. The purpose can be realized in the stage of generation in the coarse grid, where the conditional statistical properties of Y conditioned on the measured data are used in the MD method instead of the unconditional ones and additional nodes corresponding to measured locations are created.

3. NUMERICAL EXAMPLES

To demonstrate and evaluate the new algorithm, random fields over a square domain are presented and discussed. Without loss of generality, the problems will be non-dimensionalized by the half of the side length of the square, and it will be assumed that, in unconditional simulation, Y has a constant mean over the domain, $m_Y = 0$, and the standard deviations of Y is $\sigma_Y = 1$. An exponential isotropic covariance function is used; $C_Y(\mathbf{r}) = \sigma_Y^2 \exp(-|\mathbf{r}|/I)$, where I is the correlation scale. The domain is uniformly divided into a two-level grid, a coarse and fine grid consisting of 21×21 and $81 \times 81 - 21 \times 21$ nodes, respectively. Hence the spacing of the coarse grid is $b = 0.1$.

In order to address model accuracy, the covariance $C_Y(\mathbf{x}_1, \mathbf{x})$ along the diagonal from the left-low corner $\mathbf{x}_1$ to the right-up corner is calculated by the new algorithm for different values of I ranging from $b/2$ to $5b$. In all cases, the results with the searching radius r_s not smaller than $2b$ show excellent reproduction of the prescribed covariance model for all separation distances. As a demonstration, the results for the case of $I = b$ is shown in Figure 2.

The computational cost of the SSS method and the presented method for generating 10 000 realizations is compared in Figure 3. The computational cost of the MD method, especially resulting from matrix decomposition operation, is very expensive for the fine mesh used here, and hence the associated CPU time is not shown in Figure 3. It is found from the figure that the SSS method and the new method with the same searching radius are comparable in the computational cost. The accuracy of the three methods is compared in Figure 4, where the two lines are superposed. Obviously, the covariance values predicted by the MD method coincide with the theoretical results. Hence the MD method has the highest accuracy. Covariance values from the new method at large separation distances is significantly superior to the values from the SSS method and comparable to the values from the MD method.
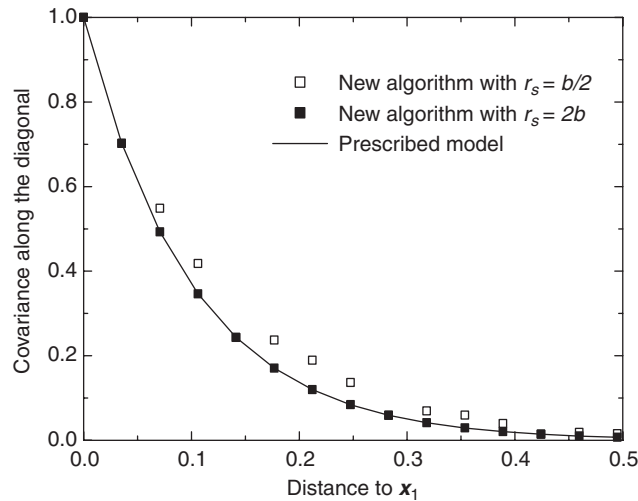
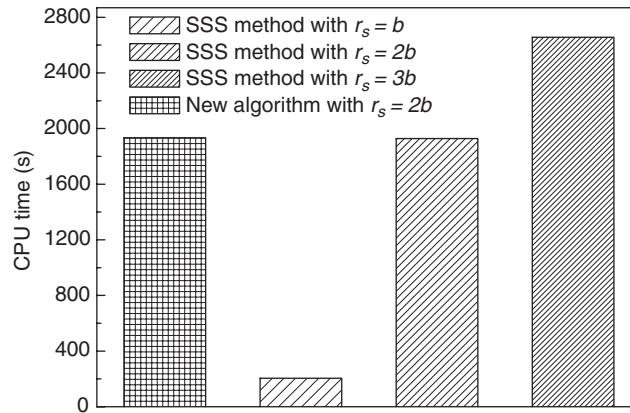Figure 2. Covariance $C_Y(\mathbf{x}_1, \mathbf{x})$ for the case of $I = b$.

Figure 3. Computational costs for generating 10 000 realizations.

Some other correlation functions have also been tested. It is found that, for the Gaussian correlation function $\exp(-r^2/I^2)$, the predictions by both the SSS method and the new method have large oscillations (negative values) at long separation distances, which cannot be removed completely by refining the mesh and/or increasing the searching radius, while for the exponential one used above and another function $1/(1 + r^2/I^2)$, there is no such a problem. A solution of this problem will be left to future work.

4. CONCLUSIONS

A new computer algorithm for generating spatially-correlated random fields is presented. The algorithm consists of using a multilevel grid strategy and combining the matrix decomposition
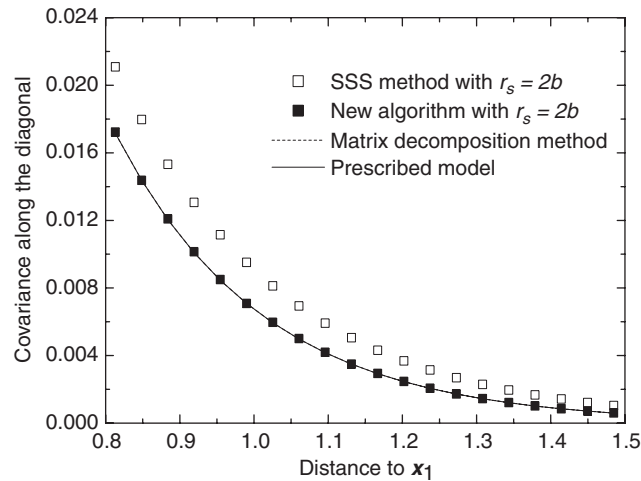
Figure 4. Covariance $C_Y(\mathbf{x}_1, \mathbf{x})$ for the case of $I = 2b$ predicted by different methods.

method and the screening sequential simulation method. Numerical results of random fields over a square domain are presented to demonstrate and evaluate the new method. In all cases, the results with the searching radius not smaller than twice of the coarse grid spacing show excellent reproduction of the prescribed covariance model for all separation distances. Concerning the computational cost and accuracy of the three methods considered, the MD method gives the highest accuracy, however the computational cost is intensive for fine meshes, the presented method and the SSS method are comparable in the computational cost and more efficient than the MD method, and the accuracy of the presented method at large separation distances is significantly superior to that of the SSS method and comparable to that of the MD method. Overall the presented method offers a new way to generate random fields efficiently and accurately. The new algorithm is also versatile: it can directly simulate fields of irregular geometry without additional efforts.

REFERENCES

1. Adler PM, Thovert JF. *Fractures and Fracture Networks*. Kluwer: Dordrecht, 1999.
2. Mejía J, Rodríguez-Iturbe I. On the synthesis of random field sampling from the spectrum: An application to the generation of hydrologic spatial processes. *Water Resources Research* 1974; **10**(4):705–711.
3. Clifton PM, Neuman SP. Effects of kriging and inverse modeling on conditional simulation of the Avra Valley aquifer in southern Arizona. *Water Resources Research* 1982; **18**(4):1215–1234.
4. Davis MW. Production of conditional simulations via the LU triangular decomposition of the covariance matrix. *Mathematical Geology* 1987; **19**(2):91–98.
5. Myers DE. Vector conditional simulation. In *Geostatistics*: *Proceedings of the Third International Geostatistics Congress, September 5–9, 1988, Avignon*, France, Armstrong M (ed.). Kluwer Academic: Boston, MA, 1989; 283–293.
6. Delhomme IP. Spatial variability and uncertainty in groundwater flow parameters: A geostatistical approach. *Water Resources Research* 1979; **15**(2):269–280.
7. Mantoglou A, Wilson IL. The turning bands method for the simulation of random fields using line generation by a spectral method. *Water Resources Research* 1982; **18**(5):1379–1394.
8. Mantoglou A. Digital simulation of multivariate two- and three-dimensional stochastic processes with a spectral turning bands method. *Mathematical Geology* 1987; **19**(2):129–149.

9. Tompson AFB, Ababou R, Gelhar LW. Implementation of the three-dimensional turning bands random field generator. *Water Resources Research* 1989; **25**(10):2227–2243.
10. Gòmez-Hernàndez JJ, Srivastava RM. ISIM3D: An ansi-c three-dimensional multiple indicator conditional simulation program. *Computers in Geosciences* 1990; **16**(4):395–440.
11. Omre H, Sølna K, Tjelmeland H. Simulation of random functions on large lattices. In *Geostatistics. Tróia '92*, Sores A (ed.). Kluwer Academic: Boston, MA, 1993; 179–199.
12. Bellin A, Rubin Y. HYDRO_GEN: A spatially distributed random field generator for correlated properties. *Stochastic Hydrology and Hydraulics* 1996; **10**:253–278.
13. Tacher L, Parriaux A. Automatic nodes generation in N-dimensional space. *Communications in Numerical Methods in Engineering* 1996; **12**:243–248.
14. Dagan G. *Flow and Transport in Porous Formations*. Springer: Berlin, 1989.