# TBSIM: A computer program for conditional simulation of three-dimensional Gaussian random fields via the turning bands method $\stackrel{\leftrightarrow}{\sim}$

Xavier Emery<sup>a,\*</sup>, Christian Lantuéjoul<sup>b</sup>

<sup>a</sup>Department of Mining Engineering, University of Chile, Avenida Tupper 2069, Santiago 837 0451, Chile <sup>b</sup>Centre de Géostatistique, Ecole des Mines de Paris, 35 rue Saint-Honoré, 77 300 Fontainebleau, France

#### Abstract

The simulation of spatially correlated Gaussian random fields is widespread in geologic, hydrologic and environmental applications for characterizing the uncertainty about the unsampled values of regionalized attributes. In this respect, the turning bands method has received attention among practitioners, for it allows multidimensional simulations to be generated at the CPU cost of one-dimensional simulations.

This work provides and documents a set of computer programs for (i) constructing three-dimensional realizations of stationary and intrinsic Gaussian random fields, (ii) conditioning these realizations to a set of data and (iii) back-transforming the Gaussian values to the original attribute units. Such programs can deal with simulations over large domains and handle anisotropic and nested covariance models.

The quality of the proposed programs is examined through an example consisting of a non-conditional simulation of a spherical covariance model. The artifact banding in the simulated maps is shown to be negligible when thousands of lines are used. The main parameters of the univariate and bivariate distributions, as well as their expected ergodic fluctuations, also prove to be accurately reproduced.

Keywords: Geostatistics; Stochastic simulation; Spectral method; Ergodic fluctuations; Multivariate normal distribution

# 1. Introduction

Geostatistical simulation is increasingly used in earth and environmental sciences for assessing the uncertainty about the unsampled values of a spatial attribute, either at a single location (local uncertainty) or jointly over several locations (spatial uncertainty). For instance, it is helpful to estimate the probability that the average value of the attribute over a given area exceeds a critical level, a problem that arises in ore reserve evaluation, agricultural land management, groundwater contamination or polluted soil management.

Continuous attributes are often modeled by nonlinear transforms of Gaussian random fields, for which the spatial distribution is multivariate

<sup>&</sup>lt;sup>☆</sup>Code available from server at http://www.iamg.org/ CGEditor/index.htm.

<sup>\*</sup>Corresponding author. Tel.: + 56 2 978 4498;

fax: +5626723504.

*E-mail address:* xemery@cec.uchile.cl (X. Emery).

Gaussian (in short, multigaussian). The simulation algorithms developed in this respect can be classified in two families:

- (1) Exact algorithms, for which the statistical properties of the simulated field match that of the desired model. This family comprises the LU decomposition of the covariance matrix (Davis, 1987), convolution methods such as the autoregressive and moving average models (Black and Freyberg, 1990; Chilès and Delfiner, 1999), and spectral approaches based on discrete and fast Fourier transforms (Pardo-Igúzquiza and Chica-Olmo, 1993; Dietrich and Newsam, 1993, 1996; Wood and Chan, 1994; Chilès and Delfiner, 1997). The LU algorithm is applicable to any configuration of the simulated locations and any covariance model, but CPU time becomes prohibitive when more than a few thousands locations are considered. Although computationally efficient, the other aforementioned methods can simulate only gridded locations and are not applicable to any covariance model (in particular, the presence of largescale structures may be problematical); also, memory storage requirements become important when millions of locations have to be simulated.
- (2) Approximate algorithms, which produce realizations of random fields whose spatial distribution is close to multigaussian. In practice, they are often the only alternative when dealing with huge conditioning datasets or with simulation domains containing millions of nodes. Approximations take place because of algorithmic simplifications (e.g. neighborhood restrictions) or because the algorithm relies on the central limit theorem and convergence to multinormality is only asymptotic. In general, this is a minor concern for practical applications, but it may be a crucial issue in studies that aim to assess statistical procedures. To date, the most widespread approximate algorithms are the sequential Gaussian (Deutsch and Journel, 1992) and continuous spectral methods (Shinozuka and Jan, 1972). Another option is the turning bands method (Matheron, 1973), which achieves 2D or 3D simulations via a series of one-dimensional simulations along lines that sweep  $R^2$  or  $R^3$ .

Although it is one of the oldest techniques for simulating multidimensional random fields, the

turning bands method is still scarcely used in geostatistical applications and the existing free packages are often limited in their implementation or applicability. As an example, the first *GSLIB* version (Deutsch and Journel, 1992) offers a basic program for simulating a regular grid in  $\mathbb{R}^3$ , based on an earlier code by Journel and Huijbregts (1978, pp. 538–545). However, apart from generating non-conditional realizations, this program uses a restricted number of lines (15) and allows only isotropic spherical and exponential covariance models to be reproduced.

A more complete free library is the R package RandomFields by Schlather (2001). It uses either a continuous spectral or a circulant-embedding algorithm for line simulation. Also, a large number of covariance models are available and the realizations can be made conditional to a set of data, which makes the package appealing. However its limitations are twofold. The first one concerns the line simulation process: the continuous spectral method is not efficient for simulating covariances that are not smooth at the origin (Lantuéjoul, 2002, p. 192), while the circulant-embedding algorithm (Dietrich and Newsam, 1993) requires discretizing each line into regular intervals and therefore loses accuracy in the covariance reproduction (see discussion in Section 2.2). Second, several facilities that are helpful for practical applications are not implemented, such as a back-transformation from normal scores to original variable, the definition of a moving neighborhood for conditioning kriging, or the use of dual kriging to quicken calculations when working in a unique neighborhood.

Such limitations have motivated the present work, which aims at supplying and documenting computer programs for simulating three-dimensional Gaussian random fields via the turning bands method. A second objective is to illustrate the efficiency and reliability of this method, through an analysis of the bivariate distributions and statistical fluctuations of the simulated values. The proposed programs are based on a former code published by Lantuéjoul (1994, pp. 169–177), which has been completed to include the following features:

- possibility to handle nested and anisotropic covariance models,
- possibility to simulate either gridded or scattered locations on large domains,
- accurate reproduction of the desired covariance model (without approximation),

- availability of the most commonly used covariance models,
- conditioning to a set of existing data,
- definition of either a unique or a moving neighborhood for conditioning kriging,
- back-transformation from normal values to original units,
- change of support (upscaling) of the realizations,
- improvements in the simulation code to quicken calculations.

In the following, the simulation of stationary or intrinsic random fields is considered. In the intrinsic case, the traditional covariance function does not exist and one works with a generalized covariance; also, the multivariate Gaussian assumption does not concern the random field itself, but its increments or generalized increments (Dimitrakopoulos, 1990; Pardo-Igúzquiza and Dowd, 2003). Henceforth the term 'covariance' will stand for either the traditional or the generalized covariance; this function is assumed invariant under translation of the coordinates, i.e. it only depends on the separation vector between the current locations.

# 2. The turning bands algorithm

#### 2.1. General presentation

The principle of the turning bands method is to simplify the simulation problem in  $\mathbb{R}^d$  (with, in general, d = 2 or 3) into a problem in  $\mathbb{R}$ , which allows one to perform multidimensional simulations at the CPU cost of one-dimensional simulations. Let us consider a random field { $X(t), t \in \mathbb{R}$ } with zero mean and continuous covariance function  $C_X$ , and a random vector U with a uniform distribution over the unit sphere  $S_d$  of  $\mathbb{R}^d$ . For any location  $\mathbf{x} \in \mathbb{R}^d$ , let us put:

$$\forall \mathbf{x} \in \mathsf{R}^d, \quad Y(\mathbf{x}) = X(\langle \mathbf{x}, \mathbf{U} \rangle), \tag{1}$$

where <,> is the standard inner product in  $\mathbb{R}^d$ . Eq. (1) defines a random field  $\{Y(\mathbf{x}), \mathbf{x} \in \mathbb{R}^d\}$  with zero mean and isotropic covariance  $C_Y$  such that

$$C_Y(r) = \int_{\mathsf{S}_d} C_X(\langle \mathbf{h}, \mathbf{u} \rangle) \varpi_d(\mathrm{d}\mathbf{u}), \tag{2}$$

where **h** is a vector of  $\mathbb{R}^d$ , *r* its modulus and  $\varpi_d$  the uniform distribution over  $S_d$ .

Matheron (1973, p. 462) proved that relation (2) is a one-to-one mapping between the set of continuous and isotropic covariances in  $R^d$  and

that of continuous covariances in R. This property allows one to replace the simulation of a multidimensional random field with covariance  $C_Y$  by the simulation of a one-dimensional field with covariance  $C_X$ . The expression of  $C_X$  as a function of  $C_Y$ is often complicated when d = 2 (planar simulation) (Brooker, 1985; Gneiting, 1998), but it has a simple solution in three dimensions, which is the framework considered in this work. By using spherical coordinates, formula (2) can be inverted as

$$C_X(r) = \frac{\mathrm{d}}{\mathrm{d}r} [r C_Y(r)],\tag{3}$$

which determines the covariance  $C_X$  associated with a given isotropic covariance  $C_Y$  in  $\mathbb{R}^3$ .

# 2.2. Comments on the line simulation process

There exists a wide choice of methods for simulating a one-dimensional random field with prescribed covariance  $C_X$ . Several options have been proposed by Mantoglou and Wilson (1982), Mantoglou (1987), Christakos (1987), Tompson et al. (1989), Lantuéjoul (1994) and Dietrich (1995), among others. Exact Gaussian simulation algorithms such as the moving average and the circulant embedding methods are suited to the simulation of regularly spaced locations, which is usually not the case for the projection of the nodes to be simulated onto a line with random orientation, even if these nodes are on a regular grid in  $\mathbb{R}^3$ . A discretization of the line at a given mesh and a migration of the projected nodes are therefore required to perform the one-dimensional simulation. This procedure provokes a bias in the reproduction of the covariance model, as well as difficulties in memory management (and increase in computing time) if the discretization mesh is very small with respect to the size of the projected three-dimentional domain.

These impediments can be avoided by resorting to continuous simulation algorithms along the line. A common example is the continuous spectral method, which provides an accurate simulation of the target covariance without need for regularly spaced locations. This method has been implemented in program TBSIM.M for simulating covariance models that are smooth at the origin. Regarding other models, ad hoc algorithms have been preferred, allowing an efficient and accurate simulation of the one-dimensional random field at any set of irregularly spaced locations. Note that the use of multigaussian simulation algorithms is needless since multivariate normality is lost anyway in the spreading process from R to  $R^3$  [Eq. (1)].

### 2.3. Available covariance models

Fifteen models are implemented in program TBSIM.M. Their definition and the way to obtain the one-dimensional simulations are described hereafter.

(1) *Spherical covariance* with sill *C* and scale factor (range) *a*:

$$C_Y(r) = C\left(1 - \frac{3}{2a}\frac{r}{a} + \frac{1}{2a^3}\right) \mathbf{1}_{r \leqslant a}.$$
 (4)

The associated one-dimensional random field  $\{X(t), t \in \mathbb{R}\}$  is simulated in the following fashion (Lantuéjoul, 1994, p. 157):

- choose an offset on the real axis at random (uniformly in [0,*a*]) and divide this axis in intervals with length *a*;
- within each interval, draw a linear function with an equal probability of being increasing or decreasing; the slope sign is independent from one interval to another one (Fig. 1).

(2) *Exponential covariance* with sill *C* and scale factor *a* (practical range 3*a*):

$$C_Y(r) = C \, \exp\left(-\frac{r}{a}\right). \tag{5}$$

The one-dimensional random field can be simulated thanks to a Poisson point process in R (Lantuéjoul, 1994, p. 156). Here, we use a faster method, consisting in viewing the exponential covariance as a scale mixture of spherical models.



Fig. 1. Realization of a 1D random field associated with a 3D spherical model

Let us put

$$C_Y(r) = \int_r^{+\infty} C\left(1 - \frac{3}{2u}r + \frac{1}{2u^3}r^3\right) f(u) \,\mathrm{d}u \tag{6}$$

for a probability density function f(.) in  $R_+$  to be determined. By differentiating with respect to r and after simplification, one finds

$$\forall u > 0, f(u) = \frac{u}{3 a^2} \exp\left(-\frac{u}{a}\right) \left\{\frac{u}{a} + 1\right\}.$$
(7)

One recognizes the probability density function of a gamma random variable **G** with parameters (3,*a*) with probability 1/3 and (2,*a*) with probability  $\frac{1}{3}$ . The simulation of a random field with an exponential covariance can therefore be performed in the following fashion:

- simulate the random variable G;
- simulate a random vector U with a uniform distribution over the sphere S<sub>3</sub>;
- divide the line with orientation U into intervals with the same length G. Within each interval, draw a linear function with positive or negative slope with equal probability (Fig. 1).

(3) Gamma covariance with sill C, scale factor a and parameter b > 0:

$$C_Y(r) = C\left(1 + \frac{r}{a}\right)^{-b}.$$
(8)

This model is simulated as a scale mixture of exponential covariances [Eq. (5)]. The scale factors are the reciprocals of gamma random variables with shape parameter b (Chilès and Delfiner, 1999, p. 70).

(4) *Stable covariance* with sill *C*, scale factor *a* and parameter  $b \in [0,2]$ :

$$C_Y(r) = C \exp\left\{-\left(\frac{r}{a}\right)^b\right\}.$$
(9)

When b > 1, this model is simulated as a scale mixture of Gaussian covariances (see model type n<sup>o</sup> 6 hereafter); the scale factors are the reciprocals of square-rooted stable random variables with parameter b. Mixing Gaussian covariances is also valid when  $b \le 1$ , but this approach is not efficient because the covariance becomes convex at the origin. In this case, it is preferable to see the stable model as a mixture of exponential covariances with scale factors inversely proportional to stable random variables with parameter b. Note that the case b = 1corresponds to the exponential model and b = 2 to the Gaussian model. (5) *Cubic covariance* with sill *C* and scale factor (range) *a*:

$$C_Y(r) = C \left( 1 - 7\frac{r^2}{a^2} + \frac{35}{4}\frac{r^3}{a^3} - \frac{7}{2}\frac{r^5}{a^5} + \frac{3}{4}\frac{r^7}{a^7} \right) 1_{r \le a}.$$
(10)

This model is obtained in a way similar to the spherical covariance, but with a cubic function instead of a linear function on each interval (Chilès and Delfiner, 1999, pp. 488, 649).

(6) Gaussian covariance with sill C and scale factor a (practical range  $a\sqrt{3}$ ):

$$C_Y(r) = C \exp\left\{-\left(\frac{r}{a}\right)^2\right\}.$$
(11)

This covariance can be simulated by applying the continuous spectral method, which constitutes a particular case of the turning bands method. It consists in putting

$$\forall \mathbf{x} \in \mathsf{R}^3, \, Y(\mathbf{x}) = \sqrt{2C} \, \cos\left(2\pi < \mathbf{x}, \mathsf{T} > + \Phi\right), \quad (12)$$

where  $\Phi$  is a uniform phase in  $[0,2\pi]$  and **T** a random frequency distributed according to the spectral measure of  $C_Y$ . In this case, **T** is a Gaussian vector in  $\mathbb{R}^3$ .

(7) Cardinal sine covariance with sill C and scale factor a:

$$C_Y(r) = C\left(\frac{a}{r}\right)\sin\left(\frac{r}{a}\right).$$
(13)

Again, the continuous spectral method is used for simulating this covariance. Here, the spectral measure in  $\mathbb{R}^3$  is uniform on the sphere centered on **0** with radius 1/a.

(8) *J-Bessel covariance* with sill C, scale factor a and parameter  $b \ge 0.5$ :

$$C_Y(r) = C2^b \Gamma(b+1) \left(\frac{r}{a}\right)^{-b} J_b\left(\frac{r}{a}\right).$$
(14)

In  $\mathbb{R}^3$ , the spectral density of  $C_Y$  is

$$\forall \mathbf{t} \in \mathsf{R}^{3}, f(\mathbf{t}) = \frac{\Gamma(b+1)a^{3}(1-a^{2}|\mathbf{t}|^{2})^{b-3/2}}{\Gamma(b-1/2)\pi^{3/2}} \mathbf{1}_{|\mathbf{t}|<1/a}.$$
(15)

If **T** is a random vector with this density, then  $a^2$   $|\mathbf{T}|^2$  is a beta random variable with parameters (3/2,b-1/2). Note that the cardinal sine model [Eq. (13)] belongs to the J-Bessel family and corresponds to the particular case b = 1/2. The case b < 1/2 is not allowed in three dimensions.

(9) *K-Bessel covariance* with sill *C*, scale factor *a* and parameter b > 0:

$$C_Y(r) = \frac{C}{\Gamma(b)} 2^{1-b} \left(\frac{r}{a}\right)^b K_b \left(\frac{r}{a}\right).$$
(16)

This covariance is simulated as a mixture of Gaussian covariances whose scale factors are square-rooted gamma random variables with shape parameter b. Indeed, by using the Sonine–Schläfli representation of the K-Bessel function (Matheron, 1965, p. 290), Eq. (16) becomes

$$C_Y(r) = \frac{C}{\Gamma(b)} \int_0^{+\infty} \exp\left\{-\left(\frac{r}{2 a\sqrt{u}}\right)^2\right\} \exp(-u) u^{b-1} du.$$
(17)

The previous approach is not efficient when the parameter *b* is less than 1/2, as the covariance is not smooth at the origin. In this case, it is preferable to view the K-Bessel covariance as a mixture of exponential models whose scale factors are square-rooted beta random variables with parameters (b,1/2-b) (Matheron, 1965, p. 47):

$$C_Y(r) = \frac{2C\sqrt{\pi}}{\Gamma(b)\Gamma(1/2 - b)} \int_0^1 \exp\left(-\frac{r}{au}\right) u^{2b-1} (1 - u^2)^{-b-1/2} \, \mathrm{d}u.$$
(18)

Also note that the case b = 0.5 corresponds to the exponential model.

(10) Generalized Cauchy covariance with sill C, scale factor a and parameter b > 0:

$$C_Y(r) = C \left\{ 1 + \left(\frac{r}{a}\right)^2 \right\}^{-b}.$$
(19)

This model is a scale mixture of Gaussian covariances. The scale factors are the reciprocals of square-rooted gamma random variables with shape parameter b.

(11) *Exponential-sine covariance* with sill C and scale factor a:

$$C_Y(r) = C \sin\left\{\frac{\pi}{2}\exp\left(-\frac{r}{a}\right)\right\}.$$
 (20)

By truncating a Gaussian random field with such a covariance to the median threshold (y = 0), one obtains the indicator of a random set with proportion 0.5 and exponential covariance (Lantuéjoul, 2002, p. 216). For the one-dimensional simulation, the continuous spectral method is used. The spectral density of the exponential-sine covariance is a mixture of Gaussian densities with random variances:

$$\forall \mathbf{t} \in \mathsf{R}^{3}, f(\mathbf{t}) = \int_{0}^{+\infty} \left(\frac{2a^{2}u}{2\pi}\right)^{3/2} \exp(-a^{2}|\mathbf{t}|^{2}u)h(u) \,\mathrm{d}u,$$
(21)

where h(.) is a probability density function given by a series expansion:

$$\forall u \in \mathsf{R}_{+}, h(u) = \frac{1}{\sqrt{\pi u}} \sum_{n=0}^{+\infty} \frac{(-1)^{n}}{(2n)!} \left(\frac{\pi}{2}\right)^{2n+1} \exp\{-(2n+1)^{2}u\}$$
(22)

Random variables with density h(.) are simulated via an acceptance-rejection method (Lantuéjoul, 2002, p. 63), by noting that h(.) is an alternate series, hence less than the term corresponding to n = 0 in Eq. (22), which is  $\pi/2$  times the standard gamma density with shape parameter 1/2.

(12) *Linear generalized covariance* with slope *C* and scale factor *a*:

$$C_Y(r) = -C\frac{r}{a}.$$
(23)

This covariance is simulated along the real axis by drawing a random field that increases or decreases by unit jumps at regularly spaced discontinuity points (Chilès and Delfiner, 1999, p. 507). Another option is to use a power covariance model (next type) with exponent b = 1.

(13) Power generalized covariance with slope C, scale factor a and parameter b > 0:

$$C_Y(r) = (-1)^{k+1} C\left(\frac{r}{a}\right)^b,$$
 (24)

where k is an integer (order of the intrinsic random field) such that  $2k < b \le 2k+2$ . This model is invariant by the turning bands operator [Eq. (3)], i.e. the corresponding covariance  $C_X$  is also a power model with the same exponent, but not the same slope. When b is less than or equal to 1, a locally stationary representation of the power covariance can be simulated continuously along the real axis, by means of a mixture of stationary triangular covariances:

$$\forall r \in [0, a], 1 - \left(\frac{r}{a}\right)^b = \int_{r/a}^1 \left(1 - \frac{r}{au}\right) F(\mathrm{d}u), \qquad (25)$$

with  $\forall u \in [0, 1[, F(u) = (1 - b)u^b \text{ and } F(1) = 1$ . The triangular model is simulated by dividing the real axis into regular intervals and assigning to each interval a constant value drawn at random from a given distribution, independently from one interval to another one.

When *b* is greater than 1, an adaptation of the continuous spectral approach is considered, with a one-dimensional random field  $\{X(t), t \in \mathsf{R}\}$  of the form

$$\forall t \in \mathsf{R}, X(t) = \theta(R) \cos\left(2\pi Rt + \Phi\right), \tag{26}$$

where  $\Phi$  is a uniform phase in  $[0,2\pi]$  and *R* is the ratio between two independent gamma random variables with shape parameter 0.5. Function  $\theta(.)$  is defined by

$$\forall r \in \mathsf{R}_{+}, \theta(r) = \sqrt{\frac{-C\pi^{0.5-b}\Gamma(b+2)\Gamma(\frac{b+3}{2}-k)}{2^{2k-3}\Gamma(k-\frac{b}{2})\Gamma(b+2-2k)}} \frac{1+r}{r^{b+0.5}}.$$
(27)

Note that Eq. (27) is undefined if *b* is an even integer. In this particular case, the random field  $\{X(t), t \in \mathsf{R}\}$  is simply a monomial of degree k = b/2.

(14) Mixed power generalized covariance with slope C, scale factor a and parameter  $b \in [0,2]$ :

$$C_Y(r) = \frac{C}{\ln(r/a)} \left\{ 1 - \left(\frac{r}{a}\right)^b \right\}.$$
 (28)

This model is a mixture of power covariances whose exponents are random variables uniformly distributed in ]0,b[.

(15) Spline generalized covariance with slope C, scale factor a and parameter b = 2k with  $k \in \mathbb{N}$ :

$$C_Y(r) = (-1)^{k+1} C\left(\frac{r}{a}\right)^b \ln\left(\frac{r}{a}\right).$$
<sup>(29)</sup>

The associated one-dimensional random field  $\{X(t), t \in \mathbb{R}\}$  is still given by Eq. (26), except that function  $\theta(.)$  is now defined as

$$\forall r \in \mathsf{R}_{+}, \theta(r) = \sqrt{\frac{C\pi^{1-b}\Gamma(b+2)}{2^{b-1}}\frac{1+r}{r^{b+0.5}}}.$$
 (30)

This expression can be obtained by considering the spline covariance with parameter b as the derivative of the power covariance given in Eq. (24) with respect to the exponent b (Chilès and Delfiner, 1999, p. 264).

#### 2.4. Ergodicity and multivariate normality

Formula (1) is not fully satisfactory as it leads to non-ergodic three-dimensional random fields: each realization exhibits a zonal anisotropy oriented along vector **U**, so that its experimental covariance does not match the theoretical isotropic model (Chilès and Delfiner, 1999, p. 474). To solve this issue, one has to repeat the spreading process many times, say N, and put

$$Y(\mathbf{x}) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i(\langle \mathbf{x}, \mathbf{U}_i \rangle), \qquad (31)$$

where { $\mathbf{U}_i$ , i = 1..., N} are independent directions randomly distributed over the sphere  $S_3$  and { $X_i$ , i = 1..., N} are independent one-dimensional random fields, each of them with covariance  $C_X(.)$ fulfilling Eq. (3). Given the realization { $\mathbf{U}_i = \mathbf{u}_i$ , i = 1..., N}, the covariance of { $Y(\mathbf{x}), \mathbf{x} \in \mathbb{R}^3$ } for a separation vector **h** with modulus *r* is

$$C_{Y}(r) = \frac{1}{N} \sum_{i=1}^{N} C_{X}(<\mathbf{h}, \mathbf{u}_{i} >).$$
(32)

This constitutes a discrete approximation of Eq. (2): the theoretical covariance model (an integral over the unit sphere  $S_3$ ) is reproduced experimentally by Monte Carlo integration. Eq. (31) explains the name of the method, as the simulated random field is the sum of several random fields generated along lines whose spatial orientation scans the unit sphere of  $\mathbb{R}^3$ .

Provided that the simulation algorithm used along the lines is accurate, the turning bands method reproduces the theoretical covariance model without bias, irrespective of the number N of lines chosen by the user. Even so, it belongs to the class of approximate methods because the spatial distribution of the simulated random field (or, in the intrinsic case, of its increments) is neither ergodic nor multigaussian. Such properties are only ensured asymptotically, i.e. if N is infinitely large, by virtue of the central limit theorem.

## 2.5. Simulation of nested and anisotropic models

A Gaussian random field with nested covariance model can be simulated as the sum of independent Gaussian random fields, each of them associated with one basic structure of the nested model. Program TBSIM.M also allows the simulation of covariances with a geometric anisotropy, which can be transformed into isotropic covariances via a rotation and rescaling of the coordinate axes. For the one-dimensional simulation, a different scale factor is assigned to each line, according to its orientation on the sphere  $S_3$ . Extensions of the turning bands method to the simulation of more general anisotropic models (not implemented in TBSIM.M) can be found in Lantuéjoul (2002, p. 192).

# 2.6. How should the lines be distributed on the sphere?

So far, the presentation assumed that the lines had a uniform distribution over the sphere  $S_3$ . However, Eq. (32) ensures that the covariance is also honored if the line directions form an equidistributed (not necessarily uniform) sequence over  $S_3$ , i.e. if the proportion of directions falling in a given sub-domain  $B \subseteq S_3$  converges to the fraction of  $S_3$  covered by B when the number of directions becomes infinite. In practice, the convergence is faster when the lines are regularly distributed, as it amounts to calculating the integral in Eq. (2) by a regular discretization of the sphere (Chilès and Delfiner, 1999, p. 476).

The maximal number of regular directions in the three-dimensional space is equal to 15. This number relies on the icosahedron approximation (Journel and Huijbregts, 1978, p. 503) and has led several authors to propose the use of 15 regular lines for the turning bands algorithm. Actually, the lines do not need to be regularly distributed (the previous presentation assumes that U is a uniform random vector over the sphere), so a number greater than 15 is by no means forbidden.

Freulon and de Fouquet (1991) examined several sequences of line directions and showed that the ergodic properties of the simulation substantially improved by choosing the directions from a van der



Fig. 2. A set of 300 equidistributed directions over the unit sphere of  $R^3$  (van der Corput sequence).

Corput sequence (Fig. 2) instead of uniform directions. The construction of such a sequence is explained in Lantuéjoul (1994, p. 164) and has been implemented in subroutine VDC.M. The directions are defined up to a random rotation in order to avoid repeating the same set of directions for all the realizations.

# 2.7. How many lines should be used?

Several factors should be considered when choosing the number of lines to use for simulating each structure of the nested model:

- (1) The distribution of the lines over the sphere (almost regular or purely random, see previous discussion).
- (2) The relative importance of the basic structure in the covariance model: the greater the contribution of the structure (fraction of the total sill), the more lines.
- (3) The criterion used to decide whether the multigaussian model is well reproduced or not. Lantuéjoul (1994, p. 159) gives clues on how to quantify the closeness of the simulated spatial distribution to a multivariate Gaussian distribution.
- (4) The type of covariance and the associated simulation technique along the lines
  - The spectral method is suitable for simulating covariance models whose spectral densities rapidly vanish at infinity (e.g. a Gaussian covariance); few lines are then needed to sample correctly such densities.
  - Concerning the simulation methods based on a division of the lines into intervals (e.g. for simulating the spherical and exponential covariances), more lines should be used if the scale factor increases. Indeed, the average number of discontinuities (Fig. 1) within a given segment [**x**,**x**'] of R<sup>3</sup> is proportional to the number of lines and inversely proportional to the scale factor. If they are not numerous enough, these discontinuities may not compensate and provoke artifacts in the simulated maps.

In two dimensions, Mantoglou and Wilson (1982) suggest using 4 to 16 lines, Chilès (1977) recommends the use of 180 lines (one per degree of angle), while Gneiting (1999) has found that 64 lines lead to an excellent covariance reproduction. In three

dimensions, the 15 regular lines are clearly insufficient and several hundreds of lines should be used, say N = 1000, as confirmed by the experience of several authors (Tompson et al., 1989; Freulon and de Fouquet, 1991). Lantuéjoul (2002, p. 198) showed that hundreds of lines are necessary to make sure that the moments up to order 4 of any weighted average of the simulated values are close enough to the expected Gaussian moments.

Actually, the number of lines is not the limiting factor of the turning bands method, as the simulation along the lines is very fast. In general, a greater CPU time is devoted to post-process the realizations, in particular concerning the conditioning to a set of existing data (see next section).

#### 3. Conditioning the realizations to Gaussian data

To facilitate practical applications, program TBSIM.M allows conditioning of the realizations to a set of normal data. Non-conditional realizations are converted into conditional realizations through a kriging step (Journel and Huijbregts, 1978, p. 495). Let  $\{Y_S(\mathbf{x}), \mathbf{x} \in \mathbb{R}^3\}$  be a non-conditional simulation. In the stationary case (finite variance models), the random field defined by

$$\forall \mathbf{x} \in \mathsf{R}^3, \, Y_{CS}(\mathbf{x}) = Y_S(\mathbf{x}) + [Y(\mathbf{x}) - Y_S(\mathbf{x})]^{SK} \quad (33)$$

constitutes a conditional simulation, i.e. it reproduces the distribution of  $\{Y(\mathbf{x}), \mathbf{x} \in \mathbb{R}^3\}$  conditional to the Y-data. In formula (33), the superscript SK stands for a simple kriging (with mean zero and covariance  $C_Y$ ) from the values at the data locations. To summarize, the conditional simulation is constructed as follows:

- draw a non-conditional realization at the target location x and at the data locations, via the turning bands method;
- (2) compute the deviations (residuals) between the data values and simulated values at the data locations;
- (3) perform a simple kriging of the residual from its values at the data locations
- (4) add the result to the non-conditional realization, in accordance with Eq. (33).

Although this approach is less straightforward than the direct conditioning performed by the sequential Gaussian method, its advantages are threefold (Emery, 2004, p. 413). First, the kriging weights needed in step (3) are the same when several realizations are required, hence only *one* kriging is necessary to make all the realizations honor the data. Second, the kriging matrix only involves the original conditioning data, not the simulated values, therefore it is smaller than in the sequential algorithm and its inversion is faster. The dual formulation of kriging advantageously reduces the amount of calculations when a unique neighborhood is used. Third, if the simulation is performed on a fine grid, numerical instabilities in the kriging systems are avoided.

In the intrinsic case (infinite variance models), conditioning is ensured by substituting intrinsic kriging for simple kriging in Eq. (33) (Dimitrakopoulos, 1990, p. 369). The other steps of the conditioning process are unchanged.

#### 4. Program description

The proposed programs are ASCII files written in MATLAB language (version 5.0 or later). The main routine, called TBSIM.M, allows one to draw several realizations of a Gaussian random field over a regular grid or over scattered locations in  $\mathbb{R}^3$  condition these realizations to Gaussian data, back-transform the values from the Gaussian scale to that of the original attribute, and regularize the simulated values to a block support. The covariance model can consist of several nested structures (see Section 2.3 for a list of the available structure types), each of them with a possible geometric or zonal anisotropy. The nugget effect is treated apart as it amounts to adding a white noise to the simulated random field.

Regarding the back-transformation from Gaussian values (y) to original values (z), a piecewise interpolation is performed according to a conversion table  $\{(z_i, y_i), i = 1 \dots p\}$  specified by the user:

$$\forall i \in \{1, ..., p-1\} \quad \forall y \in [y_i, y_{i+1}], z = a_i + b_i y,$$
 (34)

with

$$b_i = \frac{z_{i+1} - z_i}{y_{i+1} - y_i}$$
 and  $a_i = z_i - b_i y_i$ .

For tail extrapolation, an exponential function is used, which requires the definition of two positive scalars  $\lambda$  and  $\lambda'$ , as well as the extreme values  $z_{min}$  and  $z_{\text{max}}$  of the original attribute:

$$\begin{cases} \forall y \in ] -\infty, y_1], z = z_{\min} + b_0 \exp(\lambda y), \\ \forall y \in [y_p, +\infty[, z = z_{\max} + b_p \exp(-\lambda' y). \end{cases}$$
(35)

with 
$$\begin{cases} b_0 = (z_1 - z_{\min}) \exp(-\lambda y_1) \\ b_p = (z_p - z_{\max}) \exp(\lambda' y_p) \end{cases}$$

To quicken the algorithm, grid simulation is performed row-wise. At each step, one or several rows of the 3D grid are projected onto the lines and their values are simulated, conditioned to the data, back-transformed and written in the output file; the maximum number of nodes to consider at each step is defined by the user. This way, the 3D grid is not loaded in the Matlab workspace and the program is able to deal with large simulation domains. Two options are available for the conditioning kriging step: if the search radius is set to  $\infty$ , a dual kriging with a unique neighborhood is used, otherwise a conventional kriging is performed in a moving neighborhood, in association with a super-block search strategy (Deutsch and Journel, 1992, p. 31).

The reader is referred to the header of the program file for a description of the input parameters of TBSIM.M. No output variable is generated in the Matlab workspace. Instead, an external ASCII file is created, which contains one realization per column. For grid simulations, the ordering of the nodes follows the GSLIB conventions (Deutsch and Journel, 1992): point by point to the east, then row by row to the north, and finally level by level upward.

Alternatively, program TBSIM.M can be used with a parameter file: in this case, no input argument is required and the user is prompted for the name of the parameter file. If no name is entered, a default file (*tbsim.par*) is assumed. If the file does not exist, a blank file *tbsim.par* is created. When resorting to the parameter file mode, the coordinates of the locations to be simulated (if these locations are not gridded), the data coordinates and data values must be in external ASCII files without headers. TBSIM.M can also be compiled and converted to a stand-alone executable that works like a GSLIB program.

The generation of random variables requires the availability of the Matlab Statistics Toolbox (routines for simulating uniform, normal, gamma and beta random variables). TBSIM.M uses twelve additional subroutines, the description of which can be found in the corresponding program files: BACKTR, COVA, CREATE\_PARAMFILE, DUAL, KRIGE, PICK-SUPR, SEARCH, SETDUAL, SETROT, SU-PERBLK, TBMAIN and VDC.

# 5. Investigating the quality of the simulation program: example of a spherical covariance model

This section examines the quality of the realizations obtained by program TBSIM.M, through a particular example consisting of a non-conditional simulation over a domain of size  $1024 \times 1024 \times 1$ . The covariance model is composed of an isotropic spherical model with range 50 and sill 0.9 plus a nugget effect with sill 0.1. In the following, three kinds of checks are performed:

- a visual appreciation of the artifact banding due to the use of a finite number of lines;
- an analysis of the statistics of the simulated univariate and bivariate distributions;
- an inspection of the fluctuations of the first moments over a set of realizations.

# 5.1. Artifact banding

Fig. 3 presents two realizations in grayscale representations, the first one obtained by using 15 lines, the second one corresponding to 1000 lines.

As stated earlier, the use of 15 lines is clearly insufficient and leads to artifacts (striping) in the simulated map. However, such artifacts are no longer perceptible when using 1000 lines. In practice, to choose a suitable number of lines, users can construct a few non-conditional realizations and check whether their 'texture' is acceptable or not. Based on the current example and on the comments in Section 2.7, the following heuristic rule is deemed appropriate for simulating a spherical covariance model with relative sill C (sill of the spherical model divided by the total sill):

$$N \approx 20C \max_{\mathbf{u} \in S_3} \left\{ \frac{a_{\mathbf{u}}}{mesh(\mathbf{u})} \right\},\tag{36}$$

where  $a_{\mathbf{u}}$  is the range along the direction spanned by vector  $\mathbf{u}$  and *mesh*( $\mathbf{u}$ ) the minimal distance between consecutive nodes (or locations to be simulated) along that direction. This formula can also be applied to the exponential and cubic models, the construction of which is close to that of the spherical model.

# 5.2. Reproduction of the univariate and bivariate distributions

To validate the quality of the algorithm beyond a visual check of the simulated maps, a set of 100 nonconditional realizations is drawn with the same model and same domain as in the previous



Fig. 3. Two realizations of random fields simulated with the turning bands algorithm by using 15 and 1000 lines, respectively. The covariance consists of an isotropic spherical model plus a nugget effect.

X. Emery, C. Lantuéjoul



Fig. 4. Validation of the reproduction of the univariate and bivariate distributions, from 100 non-conditional realizations obtained by using 1000 lines. The covariance consists of an isotropic spherical model plus a nugget effect.

subsection, by using 1000 lines for each realization. The simulated univariate distributions are then compared to the expected standard Gaussian distribution via a quantile–quantile plot. Also, the bivariate distributions are examined through three parameters: the simulated variograms, madograms and median indicator variograms (Fig. 4). For a standard Gaussian random field with correlogram  $\rho(\mathbf{h})$ , one has the following relationships (Lantué-joul, 2002, p. 208; Emery, 2005a, p. 166):

- variogram:  $1-\rho(\mathbf{h})$ ;
- madogram (first-order variogram):

$$\sqrt{1-\rho(\mathbf{h})/\pi};$$

• median indicator variogram:

$$1/\pi \arcsin \sqrt{1 - \rho(\mathbf{h})/2}$$
.

In each case, one observes that the average simulated statistics match the theoretical model almost perfectly. These tests are easy to perform and helpful for investigating the quality of a simulation algorithm, see for instance Emery (2004) for an application to the sequential Gaussian algorithm.

#### 5.3. Ergodic fluctuations

Another way to validate a simulation algorithm is to examine the ergodic fluctuations in the simulated

statistics. The amplitude of such fluctuations depends on the size of the simulated domain and on the spatial distribution of the random field (in the present case, a multigaussian distribution). In the following subsections, we focus on the fluctuations of the spatial averages and of the simulated variograms.

# 5.3.1. Fluctuations of the spatial averages over a set of increasing sub-domains

Let V be the entire simulated domain. Suppose that it can be partitioned into P sub-domains  $\{W_p, p = 1...P\}$  with the same shape, size and orientation as a reference domain W. Let us consider the dispersion variance  $D^2(W|V)$  of the average value of the simulated random field over these sub-domains. By applying Krige's relationship, the following approximate identity is established (Matheron, 1989, p. 85):

$$D^{2}(W|V) \approx A\left(\frac{1}{|W|} - \frac{1}{|V|}\right),\tag{37}$$

where |V| and |W| are the volumes of V and W, and A is the integral range of the correlogram model (in the present case, A = 1413.7). Eq. (37) holds provided that A is finite and non-zero and that |W| is much greater than A (Lantuéjoul, 1991). It can be re-written as follows:

$$D^2(W|V) \approx \frac{A}{|V|}(P-1).$$
(38)

In log-log coordinates, the points plotting the dispersion variance of W within V as a function of P-1 should be asymptotically aligned with slope 1 when P decreases. In practice the dispersion variance is estimated as the average of the experimental variances of the sub-domains over the 100 realizations. In Fig. 5A, the test is applied to two types of sub-domains: squares and stripes with length 1024. In both cases, the points are almost aligned along the theoretical asymptote for small values of P, which corroborates the consistency of the observed fluctuations with the multigaussian model.

#### 5.3.2. Fluctuations of the regional variogram

Let  $\Gamma_{\nu}(\mathbf{h})$  be the probabilistic version of the regional variogram over domain V and  $\gamma(\mathbf{h})$  the theoretical variogram model. Then one has (Matheron, 1989, p. 81):

$$\operatorname{var}[\Gamma_{V}(\mathbf{h})] = \frac{1}{2K_{\mathbf{h}}^{2}(\mathbf{0})} \int [\gamma(\mathbf{u} + \mathbf{h}) + \gamma(\mathbf{u} - \mathbf{h}) - 2\gamma(\mathbf{u})]^{2} K_{\mathbf{h}}(\mathbf{u}) d\mathbf{u},$$
(39)

where  $K_{\mathbf{h}}(.)$  is the geometric covariogram of the intersection of V and V shifted by  $-\mathbf{h}$ .

The regional variogram  $\Gamma_{\nu}(\mathbf{h})$  is the average of many identically distributed random variables (the squared differences between pairs of simulated values). For lag distances  $|\mathbf{h}|$  much smaller than



Fig. 5. Validation of the fluctuations for (A) the spatial averages over a set of increasing sub-domains, and (B) the simulated regional variograms.

the diameter of the simulated domain V (say,  $|\mathbf{h}| < 100$  in the case under study), these variables are not independent but are located in an area that is several hundreds times the integral range of the covariance model. By assuming a mixing property in that area, the central limit theorem states that the distribution of  $\Gamma_{\nu}(\mathbf{h})$  at small distances is almost Gaussian, with mean  $\gamma(\mathbf{h})$  and variance given by Eq. (39). Consequently, the distribution of the dispersion of the regional variogram over a set of v realizations is proportional to a  $\chi^2$  random variable with v-1 degrees of freedom (in the present case v - 1 = 99). This result allows one to define a 95% probability interval around the expected variance [Eq. (39)] and check that the observed dispersion of the simulated regional variograms lies within such an interval (Fig. 5B).

Lantuéjoul (1994) showed that the turning bands method produces fluctuations that depart from the theoretical expression given in Eq. (39). However, the departure tends to zero when the number of lines increases, so it can be controlled in order to be smaller than a given accuracy level. When using one thousand of lines as in this example, the bias is negligible.

### 6. Conclusions

Thirty years after being developed, the turning bands approach is still scarcely used in geostatistical applications, although it is more efficient than most other methods for simulating Gaussian random fields. The non-conditional simulation step is generally very fast as it achieves the three-dimensional simulation through a series of one-dimensional simulations. Furthermore, a single kriging run is required to make all the realizations conditional to a set of data. By setting adequately the parameters of the algorithm (in particular, by using several hundreds or thousands of lines), the artifact banding is almost imperceptible in the realization maps, whereas the simulated statistics of the univariate and bivariate distributions and the associated statistical fluctuations are consistent with the theoretical multigaussian model.

The proposed computer programs allow one to simulate continuous attributes over large domains and can handle nested and anisotropic covariance models. They may also be a basis for the simulation of categorical attributes through the truncated Gaussian and plurigaussian algorithms, or for that of continuous attributes represented by Gaussianrelated models (Lantuéjoul, 2002; Emery, 2005b).

#### Acknowledgments

The authors are grateful to two anonymous reviewers for their valuable comments on an earlier version of this paper.

## Appendix A. Supplementary materials

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.cageo.2006.03.001

#### References

- Black, T.C., Freyberg, D.L., 1990. Simulation of one-dimensional correlated fields using a matrix-factorization moving average approach. Mathematical Geology 22 (1), 39–62.
- Brooker, P.I., 1985. Two-dimensional simulation by turning bands. Mathematical Geology 17 (1), 81–90.
- Chilès, J.P., 1977. Géostatistique des phénomènes non stationnaires (dans le plan). Doctoral Thesis, Université de Nancy I, Nancy, 152pp.
- Chilès, J.P., Delfiner, P., 1997. Discrete exact simulation by the Fourier method. In: Baafi, E.Y., Schofield, N.A. (Eds.), Geostatistics Wollongong '96. Kluwer Academic, Dordrecht, pp. 258–269.
- Chilès, J.P., Delfiner, P., 1999. Geostatistics: Modeling Spatial Uncertainty. Wiley, New York 695pp.
- Christakos, G., 1987. Stochastic simulation of spatially correlated geo-processes. Mathematical Geology 19 (8), 807–831.
- 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., Journel, A.G., 1992. GSLIB: Geostatistical Software Library and User's Guide. Oxford University Press, New York 340pp.
- Dietrich, C.R., 1995. A simple and efficient space domain implementation of the turning bands method. Water Resources Research 31 (1), 147–156.
- Dietrich, C.R., Newsam, G.N., 1993. A fast and exact method for multidimensional Gaussian stochastic simulations. Water Resources Research 29 (8), 2861–2869.
- Dietrich, C.R., Newsam, G.N., 1996. A fast and exact method for multidimensional Gaussian stochastic simulations: extension to realizations conditioned on direct and indirect measurements. Water Resources Research 32 (6), 1643–1652.
- Dimitrakopoulos, R., 1990. Conditional simulation of intrinsic random functions of order k. Mathematical Geology 22 (3), 361–380.
- Emery, X., 2004. Testing the correctness of the sequential algorithm for simulating Gaussian random fields. Stochastic Environmental Research and Risk Assessment 18 (6), 401–413.

- Emery, X., 2005a. Variograms of order  $\omega$ : a tool to validate a bivariate distribution model. Mathematical Geology 37 (2), 163–181.
- Emery, X., 2005b. Conditional simulation of random fields with bivariate gamma isofactorial distributions. Mathematical Geology 37 (4), 419–445.
- Freulon, X., de Fouquet, C., 1991. Remarques sur la pratique des bandes tournantes à trois dimensions. In: Cahiers de géostatistique, Fascicule 1, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau, pp. 101–117.
- Gneiting, T., 1998. Closed form solutions of the two-dimensional turning bands equation. Mathematical Geology 30 (4), 379–390.
- Gneiting, T., 1999. The correlation bias for two-dimensional simulations by turning bands. Mathematical Geology 31 (2), 195–211.
- Journel, A.G., Huijbregts, C.J., 1978. Mining Geostatistics. Academic Press, London 600pp.
- Lantuéjoul, C., 1991. Ergodicity and integral range. Journal of Microscopy 161 (3), 387–404.
- Lantuéjoul, C., 1994. Non conditional simulation of stationary isotropic multigaussian random functions. In: Armstrong, M., Dowd, P.A. (Eds.), Geostatistical Simulations. Kluwer Academic, Dordrecht, pp. 147–177.
- Lantuéjoul, C., 2002. Geostatistical Simulation, Models and Algorithms. Springer, Berlin 256pp.
- Mantoglou, A., 1987. Digital simulation of multivariate two- and three-dimensional stochastic processes with a spectral turning bands method. Mathematical Geology 19 (2), 129–149.

- Mantoglou, A., Wilson, J.L., 1982. The turning bands method for simulation of random fields using line generation by a spectral method. Water Resources Research 18 (5), 1379–1394.
- Matheron, G., 1965. Les variables régionalisées et leur estimation. Masson, Paris 305pp.
- Matheron, G., 1973. The intrinsic random functions and their applications. Advances in Applied Probability 5, 439–468.
- Matheron, G., 1989. Estimating and Choosing. Springer, Berlin 141pp.
- Pardo-Igúzquiza, E., Chica-Olmo, M., 1993. The Fourier integral method—an efficient spectral method for simulation of random fields. Mathematical Geology 25 (2), 177–217.
- Pardo-Igúzquiza, E., Dowd, P.A., 2003. IRFK2D: a computer program for simulating intrinsic random functions of order k. Computers & Geosciences 29 (6), 753–759.
- Schlather, M., 2001. Simulation and analysis of random fields. R News 1 (2), 18–20.
- Shinozuka, M., Jan, C.M., 1972. Digital simulation of random processes and its applications. Journal of Sound and Vibrations 25 (1), 111–128.
- Tompson, A.F.B., Ababou, R., Gelhar, L.W., 1989. Implementation of the three-dimensional turning bands random field generator. Water Resources Research 25 (8), 2227–2243.
- Wood, A.T.A., Chan, G., 1994. Simulation of stationary Gaussian processes in [0,1]<sup>d</sup>. Journal of Computational and Graphical Statistics 3 (4), 409–432.