Conditional Simulation of Two-Dimensional Random Processes using Daubechies Wavelets

T. Tran, U. A. Mueller and L. M. Bloom

School of Engineering and Mathematics, Edith Cowan University, 100 Joondalup Dr. Joondalup, 6027 Australia (t.tran@ecu.edu.au)

Abstract: In this paper we present a new algorithm for the conditional simulation of random fields. This algorithm uses Daubechies wavelets and is effective for both isotropic and anisotropic cases. The simulation is carried out in a moving window, with the movement along a random path. The random process to be simulated is assumed to be Gaussian and so non-Gaussian data need to be suitably transformed before the application of the algorithm and back transformed at the conclusion of the algorithm. Wavelet and scaling coefficients are drawn from a normal distribution with variance equal to the wavelet spectrum and scaling spectrum in such a way that the relevant properties of the conditioning data in the window under consideration are reproduced. The simulated values are obtained by application of the inverse discrete wavelet transform. Although Haar wavelets provide a simpler implementation, the window is too small to capture spatial anisotropy. Therefore, it is only effective for the isotropic case. In this algorithm we overcome the drawback of the Haar-based conditional simulation by using the normal scores of the kriging estimates as the training image to capture the spatial correlation.

Key words: Wavelet spectrum; Wavelet transform; Conditional simulation; Kriging; Spatial correlation.

1. INTRODUCTION

Geostatistical simulation algorithms are modelling methods used to obtain realisations that reproduce relevant properties of the original sample. Such properties include variance, percentiles, histogram and spatial correlation. Simulation is a useful tool for the modelling of spatial attributes that are susceptible to extreme values such as precious metal grades and pollution concentrations and in applications that focus on short range attribute variability.

Simulation methods are said to be conditional if they honour the attribute values at the sampled locations; examples include sequential Gaussian simulation [Deutsch and Journel, 1998] and conditional spectral simulation [Yao, 1998].

In this paper we introduce a geostatistical simulation algorithm that is based on wavelets. Wavelet analysis has its origin in signal processing and image reconstruction and compression. It has also been used in some applications where the phenomena have sharp edges, where a wavelet basis is more suitable than sine and cosine functions. Wavelet analysis has been used in the non-conditional simulation of random processes [Zeldin and Spanos, 1995]. Exploiting the location-independence of the wavelet spectrum of second order stationary Gaussian processes, a conditional simulation of isotropic random processes using Haar wavelets was proposed by Tran et al. [2001]. This algorithm, denoted here as HAARSIM, worked effectively for isotropic cases but failed to adequately reproduce spatial anisotropy.

The conditional simulation algorithm DB2SIM introduced here uses Daubechies wavelets with filters having four non-zero coefficients. The simulation is carried out sequentially in a two-by-two moving window where the associated wavelet and scaling coefficients are simulated and the attribute values in the window are computed via the inverse discrete wavelet transform. In order to capture the spatial correlation we use the normal score transform of kriging estimates as pseudo estimates during our simulation. For global accuracy assessment we investigate the reproduction of the sample histogram and the semivariogram model for two test data sets. The results indicate that the algorithm works effectively for both isotropic and anisotropic cases.

2. BACKGROUND

2.1 Daubechies Wavelets

Finitely supported wavelets [Kahane and Lemarie-Rieusset, 1995] are families of well localised functions, each member of which takes non-zero values only on a small interval, and whose integral over its support is equal to zero. In one dimension these wavelets are the dilations and translations of a mother wavelet ψ and are denoted by

$$\psi_{jk}(t) = 2^{-j/2}\psi(2^{-j}t-k);\, j,k \in Z \qquad (1)$$

Each wavelet family has an associated family of functions called scaling functions which are also the translations and dilations of a function ϕ and are defined by

$$\phi_{jk}(t) = 2^{-j/2}\phi(2^{-j}t-k) \qquad (2)$$

Daubechies wavelets are orthogonal finitely supported functions that are characterised by their low pass and high pass filters. These filters have finite length and the filter coefficients are determined using the properties and restrictions of multiresolution analysis and orthonormality [Daubechies, 1992]. The longer the length of their filters, the smoother the Daubechies wavelet and scaling functions become. In this paper we focus on the four-coefficient Daubechies wavelets whose low pass filter h and high pass filter g are defined by

Table 1. Values of Db2 low and high pass filters.

n	$h(n)$	$g(n)$
0	.482962913145	-.129409522551
1	.836516303738	-.224143868041
2	.224143868041	.836516303738
3	-.129409522551	-.482962913145

Daubechies wavelets have no explicit formulae. For each value of t, the associated value of $\psi(t)$ is determined using the *cacade algorithm* [Daubechies, 1992]. The wavelet and scaling auto-correlation functions $\Psi(h)$ and $\Phi(h)$ for the Daubechies wavelet and scaling functions also have no closed forms and are obtained by numerical integration from their definition:

$$\Psi(h) = \int_{-\infty}^{\infty}\psi(t)\psi(t+h)dt \qquad (3)$$

and

$$\Phi(h) = \int_{-\infty}^{\infty}\phi(t)\phi(t+h)dt \qquad (4)$$

In 2D the scaling function and its associated three wavelet functions are defined via the tensor product of the one dimensional scaling and wavelet functions

$$\phi(x,y) = \phi(x)\phi(y) \qquad (5)$$

$$\psi^1(x,y) = \phi(x)\psi(y) \qquad (6)$$

$$\psi^2(x,y) = \psi(x)\phi(y) \qquad (7)$$

$$\psi^3(x,y) = \psi(x)\psi(y) \qquad (8)$$

2.2 The Discrete Wavelet Transform (DWT)

The discrete wavelet transform (DWT) is a fast algorithm that transforms a set of discrete values into different components called scaling and wavelet coefficients in order to study these components at appropriate scales or to reduce computer storage. In the two-dimensional case, a set of values located on a regular grid of size N by M is identified with the set of scaling coefficients at level j=0, by $\{c_0(n,m): 1 \le n \le N; 1 \le m \le M\}$. The discrete wavelet transform [Mallat, 1998] allows the wavelet and scaling coefficients at level j+1 to be expressed as linear combinations of the scaling coefficients at level j. Each application of the transform reduces the number of wavelet and scaling coefficients to be stored by a factor of four. The inverse discrete transform (IDWT) is the operation used for the reconstruction of the scaling coefficients at level j from the wavelet and scaling coefficients at the coarser level j+1.

2.3 Stochastic Wavelet Analysis

In the cases of interest here, only values at some locations on the regular grid are known. We will assume that the sample is representative of a random variable with approximately standard normal distribution. Moreover, we assume that the underlying random variable is second order stationary. Then both the wavelet and the scaling coefficients are normally distributed with mean 0 and variance approximately equal to the wavelet spectrum / scaling spectrum given by the mathematical expectation of the squares of the wavelet / scaling coefficients respectively. In our case, the wavelet spectrum $\eta^k; k = 1,2,3$ and scaling spectrum τ at level j=1 are independent of location and are computed in terms of the wavelet / scaling auto-correlation functions associated with the level and the covariance model of the sample according to

$$\eta^k = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} C(h_x, h_y)\Psi^k(h_x, h_y)dh_x dh_y \quad (9)$$

and

$$\tau = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} C(h_x, h_y)\Phi_1(h_x)\Phi_1(h_y)dh_x dh_y \quad (10)$$

respectively, where

$$\Psi^1(h_x, h_y) = \Phi_1(h_x)\Psi_1(h_y) \quad (11)$$

$$\Psi^2(h_x, h_y) = \Psi_1(h_x)\Phi_1(h_y) \quad (12)$$

$$\Psi^3(h_x, h_y) = \Psi_1(h_x)\Psi_1(h_y) \quad (13)$$

Here

$$\Phi_1(h) = \Phi(h/2) \quad (14)$$

and

$$\Psi_1(h) = \Psi(h/2) \quad (15)$$

[Nason and Silverman, 1997; Sachs et al., 2000 and Tran et al., 2001] .

3. METHOD

Using DB2SIM the conditional simulation is carried out in a window of size 2 by 2 moving along a random path that visits all locations in the study area. For each window only one location is simulated. The algorithm is a combination of stochastic wavelet analysis and a variant of multi-linear regression called kriging. Stochastic wavelet analysis is used to simulate wavelet and scaling coefficients; kriging estimates are applied as pseudo estimates for capturing the spatial correlation. The conditional simulation algorithm DB2SIM comprises the following 6 steps.

Step 1: Given a random sample which is approximately normally distributed with mean 0 and variance 1 (if this condition is not met then the sample needs to be transformed into normal score space), we obtain the covariance function $C(\mathbf{h})$ from the semivariogram model via the expression $g(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h})$ and compute the wavelet spectrum and the scaling spectrum of the random variable.

Step 2: Generate a random path visiting each grid node once and only once.

Step 3: Move to the first grid node on the path.

Step 4: Construct a window of size 2 by 2; the location to be simulated is the first unestimated node found in the window according to row ascending order. As the size of the window is small the number of conditioning data within the window is not large enough to capture spatial correlation. Therefore, the normal scores of the associated kriging estimates values are assigned to the three remaining locations. These values are expressed as linear combinations of 16 scaling and wavelet coefficients within the window via the inverse discrete wavelet transform.

Step 5: Simulate scaling and wavelet coefficients and compute the simulated value using the inverse discrete wavelet transform. Since there are three equations in 16 unknowns, 13 coefficients are simulated and the other 3 are calculated in terms of the estimates and simulated coefficients. Since the kriging estimates implicitly carry the spatial correlation of the simulated process, the simulated wavelet/scaling coefficients which honour the kriging estimates also implicitly capture this spatial correlation. Having simulated the wavelet and scaling coefficients, the simulated value is computed via the inverse discrete transform. The normal scores of the kriging estimates are discarded

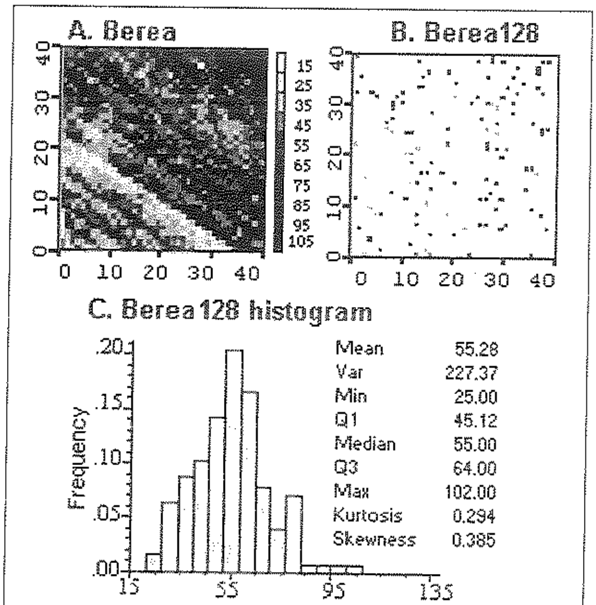

Figure 1. A/. Mosaic plot of *Berea*; B and C/. Post plot and summary statistics of *Berea128*.

before the next location is simulated.
Step 6: Repeat steps 4 and 5 until all locations are simulated.

4. TEST DATA

We use two test data sets to assess the performance of the algorithm with respect to global accuracy.

4.1 The Berea128 Sample

The *Berea128* sample consists of 128 values randomly selected from a real, anisotropic exhaustive data set *Berea*. The *Berea* data set contains 1600 measurements of air permeability on a slab of Berea sandstone [Giordano et al, 1985] with mean 55.53, variance 249.23, minimum 19.50 and maximum 111.50. A mosaic plot of the exhaustive data set, a post plot of the sample and the summary statistics of the sample are shown in Figure 1. The standardised experimental semivariogram of the normal scores is modelled as a combination of a nugget effect and one spherical structure. The values of nugget, sill and range of the model in the directions of maximum (145^0, azimuth 125^0) and minimum continuity (55^0, azimuth 35^0) are 0.2, 0.4 and 10.0 and 0.2, 0.8, 10.0, respectively. The values of nugget, sill and range of our spherical model for the attribute values in the directions of maximum and minimum continuity are 55.0, 90.0, 14.0 and 55.0, 170.0, 6.0, respectively.

4.2 The True97 Sample

The *True97* data set comes from GSLIB [Deutsch and Journel,1998] and consists of 97 values on a pseudo-regular grid. It is a sample from the synthetic exhaustive data set *True* which has two variables, denoted as *Primary* and *Secondary* each of which consists 2500 values on a 50 by 50 regular grid. In this paper we are interested only in the *Primary* variable which has characteristics similar to those from a gold mineralisation, with mean 2.58, variance 26.54, minimum 0.01 and maximum 102.70. A mosaic plot for *True*, a post plot for *True97* together with its summary statistics are shown in Figure 2. A standardised spherical semivariogram model for the normal score transform of the sample has nugget 0.1, sill 0.9 and range 10.0. Our spherical model for the attribute values has nugget 3.68, sill 6.50 and range 10.00.

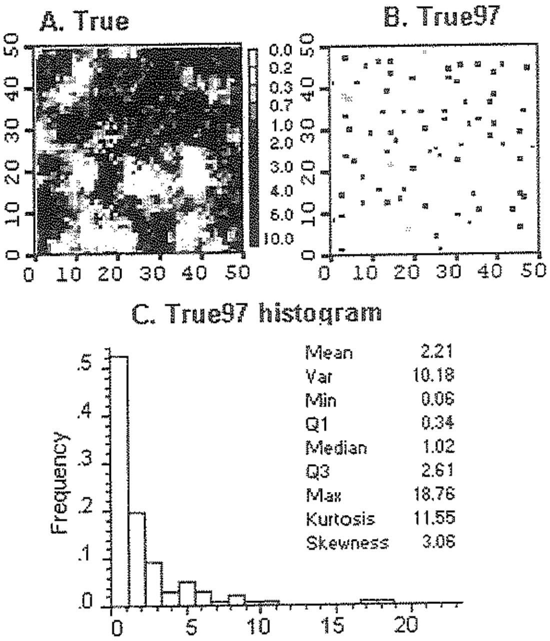

Figure 2. A/. Mosaic plot of *True*; B and C/. Post plot and summary statistics of *True97*.

5. DISCUSSION

In this section we discuss the precision of the algorithm DB2SIM by looking at the global accuracy of 50 single simulated realisations obtained for each data set. We also compare the results with those obtained using HAARSIM, the algorithm that is based on Haar wavelets without any training image.

Global accuracy is measured by the reproduction of the associated sample histogram and semivariogram model. The reproduction of the sample histogram is assessed via the mean absolute deviation (MAD) between the quantiles of the sample and the corresponding quantiles of a simulated realisation. To compare the goodness of DB2SIM with HAARSIM in the sense of histogram reproduction, the MAD between 50 simulations obtained by the two algorithms are computed using 20 histogram

Table 2. Summary statistics for the MAD values.

	DB2SIM		HAARSIM	
	Berea128	*True97*	*Berea128*	*True97*
Min	1.2299	0.0762	0.5460	0.1822
Q1	1.4614	0.0947	1.0760	0.2385
Med	1.5848	0.1029	1.2935	0.3240
Q3	1.6480	0.1103	1.5991	0.4177
Max	1.7818	0.1412	2.6983	0.7300

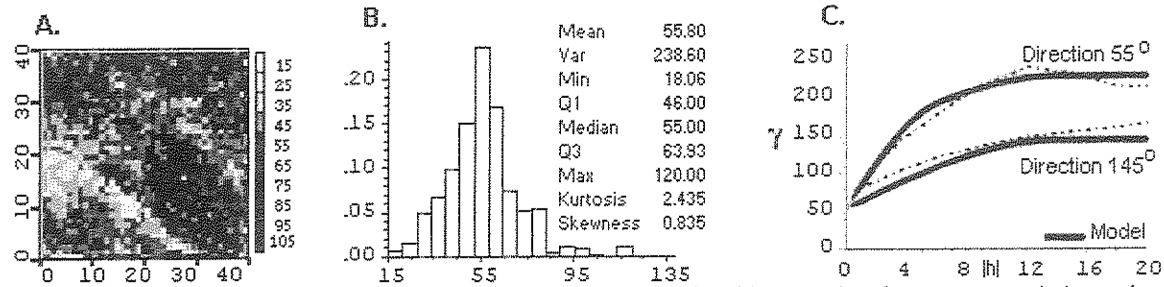

Figure 3. Mosaic plot of *Bereasim36* obtained using DB2SIM and its associated summary statistics and experimental semivariogram.

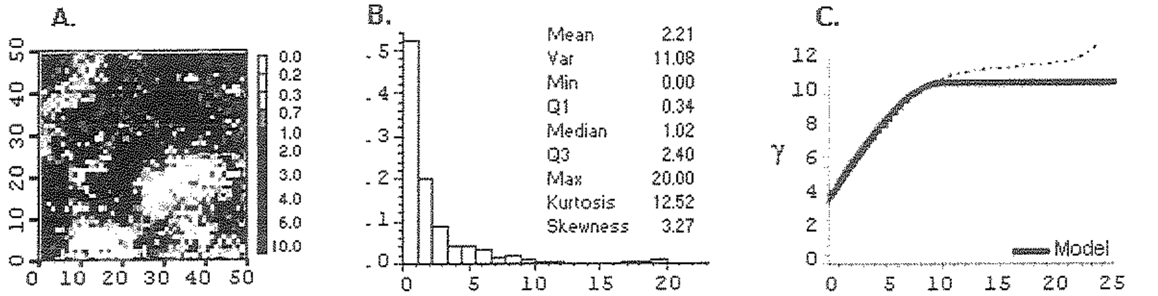

Figure 4. Mosaic plot of *Truesim36* obtained using DB2SIM and its associated summary statistics and experimental semivariogram.

classes. Summary statistics for the MAD values are listed in Table 2. These values indicate that the deviation between the simulated histograms and the associated sample histograms by the two algorithms is insignificant. That means both algorithms reproduce sample histograms, but for the highly skewed *True97*, DB2SIM performs better.

Summary statistics of a typical simulated realisation of *Berea128*, *Bereasim36*, with MAD=0.76 together with its mosaic plot and experimental semivariogram in the directions of maximum and minimum spatial continuity are shown in Figure 3; whereas summary statistics of a typical simulated realisation of *True97*, *Truesim36* with MAD=0.12 together with its mosaic plot and experimental semivariogram are shown in Figure 4. The mosaic plots in both figures show that the simulated realisations capture all features of the associated sample plots; and the summary statistics indicate that the mean and variance of the corresponding samples are reproduced.

While both HAARSIM and DB2SIM reproduce the associated sample histograms, only DB2SIM captures the spatial correlation of the anisotropic data set. The plots of the experimental semivariogram for the single simulations *Bereasim36* (in Fig. 3C) and *Truesim36* (in Fig. 4C) and the plots of 50 simulated experimental semivariograms in comparison with the associated models in Figure 5, indicate that the spatial correlation is reproduced for both anisotropic and isotropic cases using DB2SIM. In contrast, the experimental semivariograms for 50 simulated realisations of *Berea128* using HAARSIM in Figure 6 shows that spatial correlation in the direction of maximum continuity is not captured.

6. CONCLUSION

While stochastic wavelet analysis has been used in many statistical applications, the wavelet based conditional simulation algorithm has been the first attempt to exploit the location-independence of the wavelet spectrum and scaling spectrum in the conditional simulation of spatial processes. To condition the data, both HAARSIM and DB2SIM carry out the simulation in over-lapping windows of size 2 by 2. For isotropic data sets, HAARSIM reproduces the global statistics of the associated sample. Nevertheless, for anisotropic cases it fails to adequately capture the spatial correlation. On the other hand, DB2SIM works effectively for both isotropic and anisotropic cases, but has to make use of a training image to do so. Experiments with an algorithm using a combination of Haar wavelets and a training image indicate it is largely the training image that enables the reproduction of the spatial anisotropy. However, due to the nature of the Haar filters, the simulated images using HAARSIM are smoother than those obtained using DB2SIM. Our next aim is to implement a conditional simulation algorithm using Daubechies wavelets with longer filter length and investigate the impact of the filters on the results.

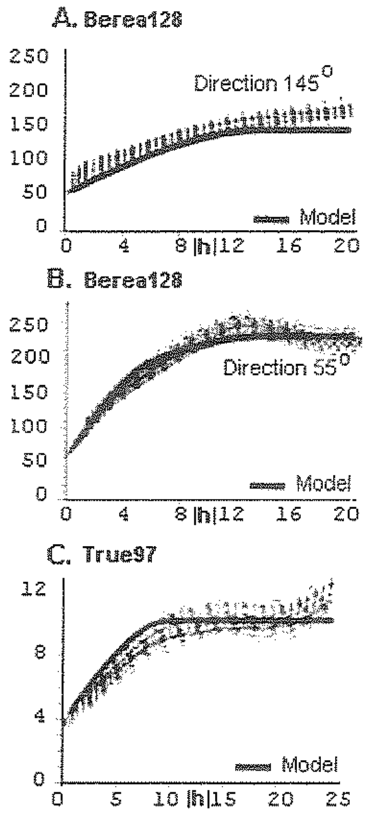

Figure 5. A &B/. Experimental semivariograms for 50 simulated realisations using DB2SIM of *Berea128* at the maximum and minimum spatial continuity; C/. Isotropic experimental semivariograms for 50 simulated realisations using DB2SIM of *True97*.

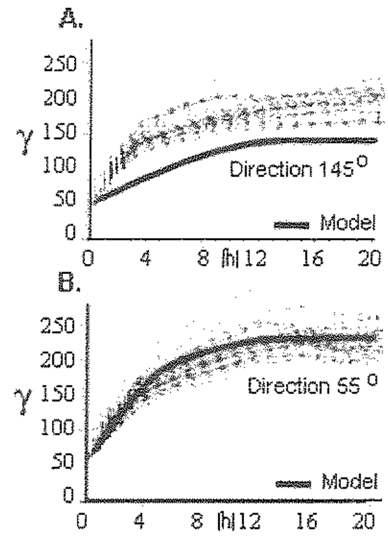

Figure 6. Experimental semivariograms for 50 simulated realisations of *Berea128* obtained using HAARSIM at the maximum and minimum spatial continuity.

7. REFERENCES

Daubechies, I., *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics, Philadelphia, 357pp, 1992.

Deutsch, C. V., and A. G Journel, *GSLIB, Geostatistical Software Library and User's Guide*, 2nd edn. Oxford University Press Inc., New York, 369 pp, 1998.

Giordano, R., S. Slater, and K. Mohanty, The effects of permeability variations on flow in porous media. *SPE Paper 14365, 60th SPE Annual Conference*, Las Vegas, 1985.

Kahane, J. P., and P-G. Lemarie-Rieusset, *Fourier Series and Wavelets*, Gordon and Breach, Australia, 394 pp, 1995.

Mallat, S. A., *Wavelet Tour of Signal Processing*, Academic Press, San Diego, 577 pp, 1998.

Nason, G. P., and B. W. Silverman, The stationary wavelet transform and some statistical applications. In: Antoniadis, A., Oppenheim, G. (Eds), *Wavelets and Statistics, Lecture Notes in Statistics*, Springer, New York, 281-300, 1995.

Sachs, R. V., G. P. Nason, and G. Kroisandt, Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum, *Journal of the Royal Statistical Society, Series B*, 60(2), 271-292, 2000.

Tran, T., U. Mueller, and L. Bloom, Conditional simulation of isotropic random fields using Haar wavelets, preprint 2001.

Yao, T., Conditional spectral simulation with phase identification, *Mathematical Geology*, 30(3), 285-308, 1998.

Zeldin, B. and P. Spanos, Random field simulation using wavelet bases. In: *Applications of Statistics and Probability*; M. Lemaire, J-L. Favre, and A. Mebarki (Eds), Balkema, Rotterdam, 1275-1283, 1995.