Prediction and Conditional Simulation of a 2D Lognormal Diffusion Random Field *

Ramón Gutiérrez[1], Concepción Roldán[2], Ramón Gutiérrez-Sánchez[1], and José M. Angulo[1]

[1] Department of Statistics and Operations Research
University of Granada. Campus de Fuentenueva, s/n,
E-18071 Granada, Spain
(e-mail: rgjaimez@ugr.es, ramongs@ugr.es, jmangulo@ugr.es)

[2] Department of Statistics and Operations Research
University of Jaén. Las Lagunillas, s/n,
E-23071 Jaén, Spain
(e-mail: iroldan@ujaen.es)

Abstract. This paper describes techniques for estimation, prediction and simulation of two-parameter lognormal diffusion random fields which are diffusions on each coordinate and satisfy a particular Markov property.
Keywords: Diffusion Random Field, Kriging, Lognormal Diffusion Process.

1 Introduction

Lognormal random fields represent the technically more complex stage of lognormal modelling. Problems as parameter estimation, lognormal simple kriging, estimation based on lognormal maximum entropy, among others, are generally undertaken by simply considering the lognormal random field as the exponential transformation of a Gaussian random field, without reference to any specific diffusion structure. This latter approach, however, constitutes an important alternative in relation to modelling, parameter estimation and inference, analysis of first passage through barriers, associated Îto equations and derivation of discrete simulation schemes, etc.

Among the contribution to theoretical foundations for diffusion random fields, see [Nualart, 1983]. In this context, [Gutiérrez *et al.*, 2004] considered lognormal random field models which are diffusions on each coordinate. Involving exogenous factors affecting the drift term, the drift and diffusion coefficients, which characterize a two-parameter lognormal diffusion under certain conditions, were estimated by maximum likelihood. For data on a regular grid, an alternative method was proposed to estimate the diffusion coefficient.

In this work, the estimates of the drift and the diffusion coefficients given in [Gutiérrez *et al.*, 2004] are used for obtaining predictions and conditional

* *This work has been partially supported by projects BFM2002-03633 and BFM2002-01836 of the DGI, Spain.*

simulations. The contents are organized as follows. First, the 2D lognormal random field model is introduced. Second, estimation of the drift and diffusion coefficients based on a discrete finite set of data is given. Finally, aspects related to kriging and conditional simulation are addressed and illustrated.

2 Lognormal Diffusion Random Fields

Lognormal diffusion processes are commonly used in the analysis of economic variables. When the parameter space is a subset of $\mathbf{R}_+^2$, [Nualart, 1983] introduced a class of two-parameter random fields which are diffusions on each coordinate and satisfy a particular Markov property related to partial ordering in $\mathbf{R}_+^2$. Using this theory, we can introduce a 2D lognormal diffusion random field as follows.

Let $\left\{X(\mathbf{z}) : \mathbf{z} = (s,t) \in I = [0,S] \times [0,T] \subset \mathbf{R}_+^2\right\}$ be a positive-valued Markov random field, defined on a probability space $(\Omega, \mathcal{A}, P)$, where $X(0,0)$ is assumed to be constant or a lognormal random variable with $E[\ln X(0,0)] = \phi_0$ and $var(\ln X(0,0)) = \sigma_0^2$. The distribution of the random field is determined by the following transition probabilities:

$$\begin{aligned} &P(B,(s+h,t+k) \mid (x_1,x,x_2),\mathbf{z}) = \\ &\quad P[X(s+h,t+k) \in B \mid X(s,t+k) = x_1, X(\mathbf{z}) = x, X(s+h,k) = x_2], \end{aligned}$$

where $\mathbf{z} = (s,t) \in I$, $h,k > 0$, $(x_1,x,x_2) \in \mathbf{R}_+^3$ and B is a Borel subset. We suppose that the transition densities exist and are given by

$$\begin{aligned} &g(y,(s+h,t+k) \mid (x_1,x,x_2),\mathbf{z}) \\ &= \frac{1}{y\sqrt{2\pi\sigma_{\mathbf{z};h,k}^2}} \exp\left\{-\frac{1}{2}\left(\frac{\ln\left(\frac{yx}{x_1x_2}\right) - m_{\mathbf{z};h,k}}{\sigma_{\mathbf{z};h,k}}\right)^2\right\}, \end{aligned}$$

for $y \in \mathbf{R}_+$, with

$$m_{\mathbf{z};h,k} = \int_s^{s+h}\int_t^{t+k} \tilde{a}(\sigma,\tau)\, d\sigma d\tau, \quad \sigma_{\mathbf{z};h,k}^2 = \int_s^{s+h}\int_t^{t+k} \tilde{B}(\sigma,\tau)\, d\sigma d\tau,$$

and $\tilde{a}$, $\tilde{B}$ being continuous functions on I. Under these conditions we can assert that $\{X(\mathbf{z}) : \mathbf{z} \in I\}$ is a lognormal diffusion random field. The one-parameter drift and diffusion coefficients associated are given by

$$\begin{aligned} a_1(\mathbf{z})\, x &:= \left(\tilde{a}_1(\mathbf{z}) + \frac{1}{2}\tilde{B}_1(\mathbf{z})\right) x, \quad B_1(\mathbf{z})\, x^2 := \tilde{B}_1(\mathbf{z})\, x^2, \\ a_2(\mathbf{z})\, x &:= \left(\tilde{a}_2(\mathbf{z}) + \frac{1}{2}\tilde{B}_2(\mathbf{z})\right) x, \quad B_2(\mathbf{z})\, x^2 := \tilde{B}_2(\mathbf{z})\, x^2, \end{aligned}$$

where

$$\tilde{a}_1(s,t) = \int_0^t \tilde{a}(s,\tau)\,d\tau, \quad \tilde{B}_1(s,t) = \int_0^t \tilde{B}(s,\tau)\,d\tau,$$
$$\tilde{a}_2(s,t) = \int_0^s \tilde{a}(\sigma,t)\,d\sigma, \quad \tilde{B}_2(s,t) = \int_0^s \tilde{B}(\sigma,t)\,d\sigma,$$

for all $\mathbf{z} = (s,t) \in I$, $x \in \mathbf{R}_+$.

The random field $\{Y(\mathbf{z}) : \mathbf{z} \in I\}$ defined as $Y(\mathbf{z}) = \ln X(\mathbf{z})$ is then a Gaussian diffusion random field, with $\tilde{a}$ and $\tilde{B}$ being, respectively, the drift and diffusion coefficients, and $\tilde{a}_1$, $\tilde{a}_2$, $\tilde{B}_1$ and $\tilde{B}_2$ being the corresponding one-parameter drift and diffusion coefficients. Furthermore, if $\mathbf{z}, \mathbf{z}' \in I$, $\mathbf{z} = (s,t)$, $\mathbf{z}' = (s',t')$, then

$$m_Y(\mathbf{z}) := E[Y(\mathbf{z})] = \phi_0 + \int_0^s \int_0^t \tilde{a}(\sigma,\tau)\,d\sigma d\tau,$$
$$\sigma_Y^2(\mathbf{z}) := var(Y(\mathbf{z})) = \sigma_0^2 + \int_0^s \int_0^t \tilde{B}(\sigma,\tau)\,d\sigma d\tau,$$
$$c_Y(\mathbf{z},\mathbf{z}') := cov(Y(\mathbf{z}), Y(\mathbf{z}')) = \sigma_Y^2(\mathbf{z} \wedge \mathbf{z}'),$$

where we write $\mathbf{z} \wedge \mathbf{z}'$ for $(s \wedge s', t \wedge t')$, with '$\wedge$' denoting the minimum.

Under suitable regularity conditions, it is possible to obtain a SPDE formulation for a two-parameter diffusion RF. In fact, we need hypotheses **I-V** to be satisfied, in order to apply *Theorem 2.8* of [Nualart, 1983]. These hypotheses and the uniqueness of solution have been proved by the authors to hold for the lognormal diffusion RF considered. Thus, there exists a two-parameter Wiener RF $\{W(\mathbf{z}) : \mathbf{z} \in I\}$ (adjoining, if it is necessary, a new probability space) such that $\{X(\mathbf{z}) : \mathbf{z} \in I\}$ is the only diffusion RF satisfying the following partial SPDE:

$$\frac{\partial^2 X(s,t)}{\partial s \partial t} - X^{-1}(s,t)\frac{\partial X(s,t)}{\partial s}\frac{\partial X(s,t)}{\partial t} - \frac{\partial a_2(s,t)}{\partial s}X(s,t) = \left(\frac{\partial B_2(s,t)}{\partial s} + B_1(s,t)\,B_2(s,t)\right)^{1/2} X(s,t)\frac{\partial^2 W(s,t)}{\partial s \partial t}.$$

This aspect is not essential for the approach considered in this work, although it provided an alternative interesting interpretation of the RF formulation considered.

Henceforth we will assume that the conditions usually considered for estimation of the drift and diffusion coefficients in the one-parameter case hold; that is, $P[\ln X(0,0) = \phi_0] = 1$ (i.e. $\sigma_0^2 = 0$) and $\sigma_Y^2(\mathbf{z}) = \tilde{B}st$, $\mathbf{z} = (s,t) \in I$.

3 Estimation of the Drift and Diffusion Coefficients

Let $\{X(\mathbf{z}) : \mathbf{z} \in I\}$ be a lognormal diffusion random field. Data $\mathbf{X} = (X(\mathbf{z}_1), ..., X(\mathbf{z}_n))^t$ are assumed to be observed at known spatial locations $\mathbf{z}_1 = (s_1,t_1)$, $\mathbf{z}_2 = (s_2,t_2), ..., \mathbf{z}_n = (s_n,t_n) \in I$. Let $\mathbf{x} = (x_1, x_2, ..., x_n)^t$ be a sample. Let us consider the log-transformed $n-$dimensional random vector, $\mathbf{Y}=(Y(\mathbf{z}_1),Y(\mathbf{z}_2),...,Y(\mathbf{z}_n))^t=(\ln X(\mathbf{z}_1),\ln X(\mathbf{z}_2),...,\ln X(\mathbf{z}_n))^t=\ln\mathbf{X}$, and the log-transformed sample, $\mathbf{y}=(y_1,y_2,...,y_n)^t=\ln\mathbf{x}$. We denote

$$\mathbf{m}_Y=(m_Y(\mathbf{z}_1),...,m_Y(\mathbf{z}_n))^t, \quad \varSigma_Y=\left(\sigma_Y^2(\mathbf{z}_i\wedge\mathbf{z}_j)\right)_{i,j=1,...,n}.$$

3.1 MLE for the Drift and Diffusion Coefficients Using Exogenous Factors

Suppose that the drift coefficient $\tilde{a}$ of Y is a linear combination of several known functions, set $\{h_1(\mathbf{z}),...,h_p(\mathbf{z}):\mathbf{z}\in I\}$, with real coefficients $\phi_1,...,\phi_p$:

$$\tilde{a}(\mathbf{z})=\sum_{\alpha=1}^{p}\phi_\alpha h_\alpha(\mathbf{z}), \quad \mathbf{z}\in I.$$

Defining, for $\mathbf{z}=(s,t)\in I$,

$$f_0(\mathbf{z})=1, \quad f_\alpha(\mathbf{z})=\int_0^s\int_0^t h_\alpha(\sigma,\tau)\,d\sigma d\tau, \quad \alpha=1,...,p,$$

the mean of Y is given by

$$m_Y(s,t)=\phi_0+\sum_{\alpha=1}^{p}\phi_\alpha\int_0^s\int_0^t h_\alpha(\sigma,\tau)\,d\sigma d\tau=\sum_{\alpha=0}^{p}\phi_\alpha f_\alpha(\mathbf{z}).$$

Thus, denoting $\mathbf{F}=(\mathbf{f}_0,\mathbf{f}_1,...,\mathbf{f}_p)$, with $\mathbf{f}_\alpha=(f_\alpha(\mathbf{z}_1),f_\alpha(\mathbf{z}_2),....,f_\alpha(\mathbf{z}_n))^t$, for $\alpha=0,1,...,p$, and $\boldsymbol{\phi}=(\phi_0,\phi_1,...,\phi_p)^t$, we have

$$\mathbf{m}_Y=(\phi_0\mathbf{f}_0+\phi_1\mathbf{f}_1+...+\phi_p\mathbf{f}_p)=\mathbf{F}\boldsymbol{\phi}.$$

Let us write

$$\varSigma_Y=\tilde{B}\mathbf{M}:=\tilde{B}\begin{pmatrix} s_1t_1 & (s_1\wedge s_2)(t_1\wedge t_2) & \cdots & (s_1\wedge s_n)(t_1\wedge t_n)\\ (s_1\wedge s_2)(t_1\wedge t_2) & s_2t_2 & \cdots & (s_2\wedge s_n)(t_2\wedge t_n)\\ \vdots & \vdots & \ddots & \vdots\\ (s_1\wedge s_n)(t_1\wedge t_n) & (s_2\wedge s_n)(t_2\wedge t_n) & \cdots & s_nt_n \end{pmatrix}.$$

With this notation, the MLE for the drift and diffusion coefficients are, respectively,

$$\boldsymbol{\phi}^*=\left(\phi_0^*,\phi_1^*,...,\phi_p^*\right)^t=\left(\mathbf{F}^t\mathbf{M}^{-1}\mathbf{F}\right)^{-1}\mathbf{F}^t\mathbf{M}^{-1}\ln\mathbf{x} \tag{1}$$

and

$$\tilde{B}^*=\frac{1}{n}(\ln\mathbf{x}-\mathbf{m}_Y^*)^t\,\mathbf{M}^{-1}(\ln\mathbf{x}-\mathbf{m}_Y^*), \tag{2}$$

where $\mathbf{m}_Y^*=\mathbf{F}\boldsymbol{\phi}^*$.

3.2 Estimation of the Drift and Diffusion Coefficients from Data on a Regular Grid

Suppose now that the data are obtained on a regular grid in $\mathbf{R}^2_+$. Let $\mathbf{z} = (s,t)$ be a point in a set S of locations included in the regular grid and let us denote the 2D four-point increment of Y by

$$Y\left(\Delta_{hk}\left(\mathbf{z}\right)\right) = Y\left(s+h,t+k\right) - Y\left(s,t+k\right) - Y\left(s+h,t\right) + Y\left(s,t\right),$$

for $h,k > 0$. Taking into account that the variance of this increment,

$$var\left(Y\left(\Delta_{hk}\left(\mathbf{z}\right)\right)\right) = \sigma^2_{\mathbf{z};h,k} = \int_s^{s+h}\int_t^{t+k} \tilde{B}\left(\sigma,\tau\right) d\sigma d\tau = \tilde{B}hk,$$

does not depend on the location $\mathbf{z}$, but only on the area hk, the diffusion coefficient $\tilde{B}$ can be estimated using a similar approach to Matheron's estimator for the variogram (see, for example, [Cressie, 1993]), considering here 2D four-point increments, as follows.

Under the implicit condition that $\mathbf{z}_i = (s_i,t_i) < \mathbf{z}_j = (s_j,t_j)$, we denote

$$[\mathbf{z}_i,\mathbf{z}_j] = \left\{(s_i,t_i),(s_i,t_j),(s_j,s_i),(s_j,t_j)\right\}.$$

The estimator, for $\mathbf{z} = (s,t)$, is

$$\begin{aligned}&\widehat{var}\left(Y\left(\Delta_{hk}\left(\mathbf{z}\right)\right)\right)\\ &= \frac{1}{\left|N\left(hk\right)\right|}\sum_{N(hk)}\left(Y\left(s+h,t+k\right) - Y\left(s,t+k\right) - Y\left(s+h,t\right) + Y\left(s,t\right)\right.\\ &\qquad \left. - m_Y\left(s+h,t+k\right) + m_Y\left(s,t+k\right) + m_Y\left(s+h,t\right) - m_Y\left(s,t\right)\right)^2,\end{aligned}$$

where

$$N\left(hk\right) \equiv \left\{(\mathbf{z}_i,\mathbf{z}_j) : [\mathbf{z}_i,\mathbf{z}_j] \in S, \left(s_j - s_i\right)\left(t_j - t_i\right) = hk, \quad i,j = 1,...,n\right\}$$

and $|N(hk)|$ is the number of different elements of $N(hk)$. If the mean is unknown, it can be estimated using (1) by $m^*_Y(\mathbf{z}) = \sum_{\alpha=0}^{p}\phi^*_\alpha f_\alpha(\mathbf{z})$.

4 Numerical Examples

In this section we describe some numerical examples illustrating estimation for a lognormal diffusion random field under the approaches considered and an example of prediction and conditional simulation. First, using simulated data on a regular grid, the two estimation methods for the diffusion coefficient respectively described in Sections 3.1 and 3.2 are compared, considering the case of known non constant mean (for the associated Gaussian random field). Second, we obtain a conditional simulation for a lognormal diffusion random field.

The parameter space considered is $I = [0, 1.65] \times [0, 1.05]$. Realizations are generated on a regular 19×19 grid, S, with SW corner at the origin $(0, 0)$ and NE corner at point (1.65,1.05). Parameter estimates, kriging predictions and simulations are obtained on this grid based on the data $\mathbf{X}$, consisting of the values corresponding to the 7×7 regular grid, subset determined by the same corner points. We will obtain unconditionally simulated realizations by the method of unconstrained simulation described in [Christakos, 1992].

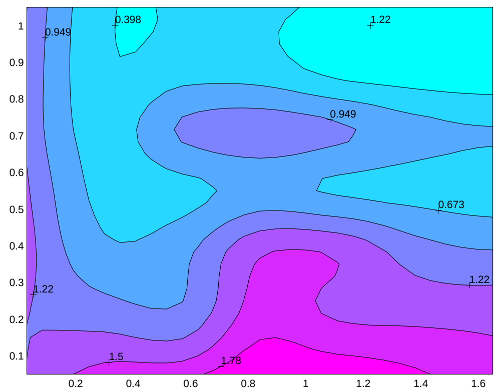

Fig. 1. Contour-level plot of 49 values generated (simulation 1) for the lognormal diffusion random field (known non constant mean case)

Sim. no.	$\tilde{B}^{*}$	$\tilde{B}^{**}$	Sim. no.	$\tilde{B}^{*}$	$\tilde{B}^{**}$
1	1.1115	0.8199	9	0.9911	0.8236
2	1.0605	0.5584	10	0.9909	0.4597
3	1.2060	1.0004	11	1.0107	0.4792
4	1.2153	0.5457	12	0.9016	0.3990
5	1.1324	0.8595	13	0.8914	1.5703
6	0.8309	0.6103	14	1.1850	0.9163
7	0.6138	0.3944	15	0.9870	1.2419
8	1.3243	0.4456	16	1.0684	1.1750

Table 1. Estimates of $\tilde{B}$ by the two methods considered, for 16 simulations of the lognormal diffusion random field (known non constant mean case)

We consider a lognormal diffusion random field with non constant mean, with $\phi_0 = 0.25$, $\tilde{a}(\mathbf{z}) = -2$, for all $\mathbf{z} \in I$, and $\tilde{B} = 1$. Table 1 gives the estimates of $\tilde{B}^{*}$ and $\tilde{B}^{**}$ obtained for 16 independent unconstrained simulations for this random field, assuming that the mean of the associated Gaussian random field is known.

From the results obtained in both cases studied, we can observe that the maximum likelihood estimation method overall provides more accurate estimates for the diffusion coefficient than the alternative method based on evaluation of 2D four-point increments. A similar behavior has been observed in several other cases studied by the authors. Lack of stability in the estimate $\tilde{B}^*$ can be possibly overcome by robust estimation of the slope of $\widehat{var}\left(Y\left(\Delta_{hk}\left(\mathbf{z}\right)\right)\right)$ vs. hk instead of using the least-squares approach.

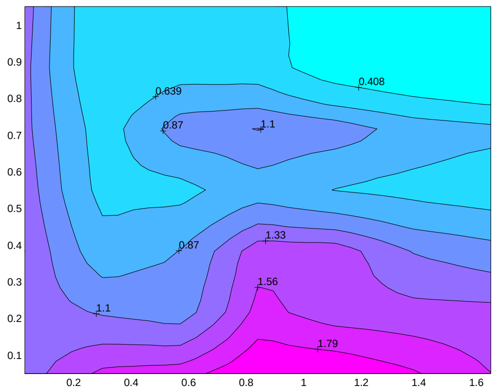

Fig. 2. Contour-level plot of the 361 predictions obtained by simple lognormal kriging using the 49 values plotted in Figure 1

As for simulation, we have considered a practical method for generating conditional simulations that combines unconditional simulation and kriging, described in [Yuh-Ming and Hugh Ellis, 1997]. This technique yields an unbiased conditional simulation (with respect to sample data) and reproduces conditional variances. We can summarize the procedure as follows:

Step 1 Predict $\{\widehat{y}\left(\mathbf{z}_i\right) : \mathbf{z}_i \in S\}$ based on the data $\mathbf{Y}$ and on the predictor of simple lognormal kriging.

Step 2 Calculate unconditionally simulated realizations $\{y^u\left(\mathbf{z}_i\right) : \mathbf{z}_i \in S\}$ based on the method of unconstrained simulation and using the estimates given in (1) and (2).

Step 3 Calculate the set of predictions $\left\{\widehat{y^u}\left(\mathbf{z}_i\right) : \mathbf{z}_i \in S\right\}$ based on the data $\{y^u\left(\mathbf{z}_i\right) : \mathbf{z}_i \in G\}$ and on the predictor of simple lognormal kriging.

Step 4 Calculate conditional simulation realizations of Y by

$$y^c\left(\mathbf{z}_i\right) = y^u\left(\mathbf{z}_i\right) + \left[\widehat{y}\left(\mathbf{z}_i\right) - \widehat{y^u}\left(\mathbf{z}_i\right)\right], \quad \forall \mathbf{z}_i \in S.$$

Step 5 Calculate conditional simulation realizations of X by

$$x^c\left(\mathbf{z}_i\right) = \frac{\exp\left\{y^u\left(\mathbf{z}_i\right)\right\}\exp\left\{\widehat{y}\left(\mathbf{z}_i\right)\right\}}{\exp\left\{\widehat{y^u}\left(\mathbf{z}_i\right)\right\}} \equiv \frac{x^u\left(\mathbf{z}_i\right)\widehat{x}\left(\mathbf{z}_i\right)}{\widehat{x^u}\left(\mathbf{z}_i\right)}, \quad \forall \mathbf{z}_i \in S.$$

For the example of prediction and conditional simulation we consider the previous diffusion. That is, a lognormal diffusion random field with non constant mean, $\phi_0 = 0.25$, $\tilde{a}(\mathbf{z}) = -2$, for all $\mathbf{z} \in I$, and $\tilde{B} = 1$. Using the 49 values obtained from simulation 1 (see Figure 1) we have obtained $\tilde{B}^* = 1.1115$ and using this estimate we have calculated 19×19 predictions by simple lognormal kriging. The results are plotted in Figure 2.

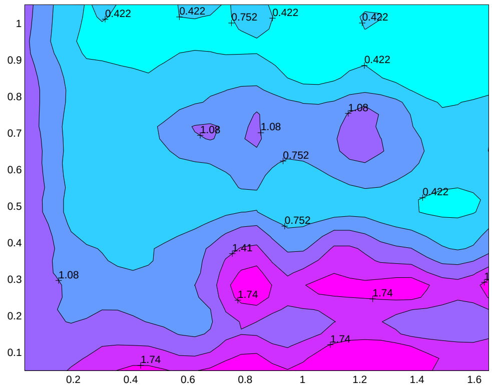

Fig. 3. Contour-level plot of the 361 simulations obtained by conditional simulation using the 49 values plotted in Figure 1

Figure 3 displays a contour-level plot for the 19×19 conditional simulation realization based on the data of simulation 1, and Figure 4 displays the original contour-level plot.

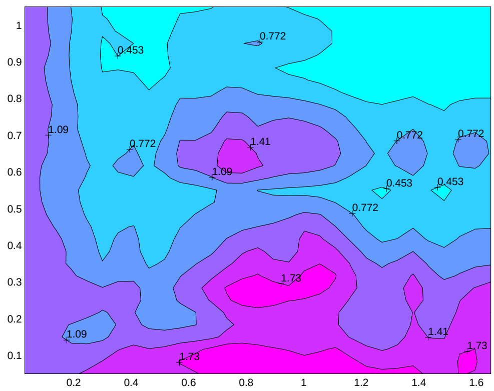

Fig. 4. Contour-level plot of the 361 values (including the 49 values used for estimating $\tilde{B}$) generated (simulation 1) for the 2D lognormal diffusion considered

5 Conclusions

In this paper we study prediction and conditional simulation for a 2D lognormal diffusion random field, including exogenous factors in its formulation. This is an important case of random fields which are not intrinsically stationary, then well-known related techniques cannot be applied. Such models are useful to represent diffusion-type positive valued characteristics, like pollutant indicators in environmental studies. The approach considered allows us to use well-known techniques for estimation and prediction, such as simple kriging, and for conditional simulation.

Possible extensions under investigation by the authors include consideration of non-constant diffusion-type values at the boundary axes as well as higher-dimension spatial and spatio-temporal formulations. Also, development of validation techniques in this context would be most important for real applications.

References

[Christakos, 1992]G. Christakos. *Random Field Models in Earth Sciences.* Academic Press, San Diego, 1992.

[Cressie, 1993]N. Cressie. *Statistics for Spatial Data.* Wiley & Sons, New York, 1993.

[Gutiérrez *et al.*, 2004]R. Gutiérrez, C. Roldán, R. Gutiérrez-Sánchez, and J.M. Angulo. Estimation and prediction of a 2D lognormal diffusion random field. *Stochastic Environmental Research and Risk Assessment*, in press, 2004.

[Nualart, 1983]D. Nualart. Two-parameter diffusion processes and martingales. *Stochastic Processes and their Applications*, 15:31–57, 1983.

[Yuh-Ming and Hugh Ellis, 1997]L. Yuh-Ming and J. Hugh Ellis. Estimation and simulation of lognormal random fields. *Computers and Geosciences*, 23(1):19–31, 1997.