Ann. Inst. Statist. Math.
Vol. 52, No. 1, 84–107 (2000)

A STOCHASTIC ADVECTION-DIFFUSION MODEL FOR THE ROCKY FLATS SOIL PLUTONIUM DATA

JAROSLAV MOHAPL*

Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

(Received October 27, 1997; revised September 28, 1998)

Abstract. An advection-diffusion equation with time and space dependent random coefficients is derived as a model for the plutonium concentration changes in the surface soil around the Rocky Flats Plant northwest of Denver, Colorado. The equation is used to fit a set of temporal-spatial data sampled annually over a 23 year period from 71 sites around the plant. The coefficients of the advection-diffusion equation are derived from the estimated covariance function of the observed random field using a combination of maximum likelihood and quasi-likelihood techniques. Goodness-of-fit of the model to the data is also assessed. Finally we interpret the model in terms of the advection-diffusion mechanism.

Key words and phrases: Temporal-spatial data, temporal-spatial process, random field, parameter estimation, stochastic partial differential equation, stochastic modeling.

1. Introduction

The data analyzed in this paper are soil plutonium concentration measurements taken between 1970 and 1992 by the Radiation Control Division of the Colorado Department of Public Health and Environment. The purpose of the survey was to assess the degree of plutonium contamination in the surface soil in the vicinity of what used to be called the Rocky Flats Plant. The contamination was caused by a used leaking oil coolant containing plutonium. Containers with the oil were stored near the south-east security fence of the plant over the period 1958–1968. The leak caused by corrosion of the containers was detected in 1964. In 1969 the containers were removed and the site was cleaned and filled by gravel. Later it was covered by asphalt so that further surface distribution of contaminated particles from the storage site into the area is prevented.

Several studies analyzing the Rocky Flats soil plutonium data were published in the past, all of them trying to draw a contour map depicting the plutonium contamination in the surface soil around the plant. A detailed discussion of their conclusions and other relevant references are given in the report by Jones and Zhang (1994). What makes our study special is the attempt to build a physically meaningful stochastic model that can be used for drawing the contour maps. We assume the radioactive plum was created mainly by contaminated particles of dust and soil airlifted into the area by the wind. The transportation mechanism is thus random and we describe it using an advection-diffusion equation with stochastic space-time dependent coefficients.

* Now at AQRM, Atmospheric Environment Service, Environment Canada, 4905 Dufferin Street, Downsview, Ontario, Canada M3H 5T4.

Literature on applications of partial differential equations (PDE's) to data generated by a random mechanism can be divided in two main groups. The first considers the PDE as a model for an autoregression problem. See e.g. Jones and Vecchia (1993), Huebner and Rozovskii (1995), Jones and Zhang (1997) and others. The second uses the PDE to model the trend in the data. See Itô and Kunisch (1990), Lamm (1992) etc. In the first case, fitting of the model to the data requires estimation of parameters in a covariance matrix of a probability distribution, while in the second case we have to deal with a non-linear regression problem. There are several ways to incorporate randomness into a PDE: through the input, by means of a white noise on the right side of the equation, using random coefficients, random boundary conditions and a combination of the three mentioned. The autoregression approach is preferred by statisticians. Its main goal is to model the covariance structure of the data by means of a stochastic PDE with one or two constant advection-diffusion parameters along with a usually simple linear trend. The equation is used to obtain a relevant covariance matrix for a Gaussian model used subsequently for prediction of the data. The non-linear regression problem usually requires the advection-diffusion parameters in the model to be space dependent in order to achieve a reasonable fit and it is preferred by applied mathematicians. The statistical approach attempts to accommodate irregularly sampled data and emphasizes efficient estimation of the parameters using maximum likelihood, whereas the papers in applied mathematics assume lattice sampling and consider mainly least squares estimation.

The advection-diffusion equation as a model for pollutant transportation appears in both mentioned contexts. It is worthwhile to mention the theoretical papers by Kwakernaak (1974) and Chapter 7 in the monograph by Kallianpur and Xiong (1995) discussing modeling of river water pollution with chemicals deposited according to a Poisson process. Unny (1988) models transportation of chemicals in a groundwater flow by starting with an advection-diffusion equation that has random coefficients and transforms it into a problem with a white noise forcing term. A similar approach is used in the case study on dispersive contaminant transportation in the Borden Aquifer by Vomvoris and Gelhar (1986). The rather deterministic or trend oriented view on the tracer experiment results from Borden Aquifer is apparent in papers by Sudicky *et al.* (1983) and Farrell *et al.* (1994). A model for air lift transportation of chemicals is given in Omatu (1984). As a tool for ocean study we find the advection-diffusion equation in Piterbarg and Ostrovskii (1997).

Despite the number of papers and applications involving a stochastic advection-diffusion equation there does not seem to be a systematic procedure for finding a fit between the equation and the data. Here we provide one possible approach. The advection-diffusion equation with random coefficients and its solution are introduced in Section 2. We assume the logarithm of the solution is a Gaussian linear model with correlated noise commonly used in statistics. The model is specified in Section 3. The covariance matrix of the noise is supplied by a stochastic PDE. In Section 4 we estimate the parameters and validate the model. Finally, in Section 5, we derive the random coefficients of the advection-diffusion equation.

In classical mathematics and theoretical physics coefficients of the PDE are usually derived by a theoretical consideration and then the scientist attempts to decide about the agreement of the solution and the observed reality. Here we proceed in a somewhat reversed way, specifying first the solution from the data and then deriving the equation which can be interpreted in a meaningful manner.

2. The model

Transportation of the plutonium into the surrounding of the plant is caused mainly by the natural diffusion of the contaminated coolant into the soil and by the wind airlifting the radioactive particles into the area. Such a process is often described by the advection-diffusion equation

$$\frac{\partial}{\partial t}c = \left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2}\right)c - \left(v_1\frac{\partial}{\partial x_1} + v_2\frac{\partial}{\partial x_2}\right)c + v_3 c, \tag{2.1}$$

where t is the time and $x = (x_1, x_2)^T$ are the coordinates on the surface. This means, the temporal changes in the concentration $c = c(t, x)$ on the left side of the equation are equal to the changes in the concentration gradient minus a specific discharge of concentration on the right. The discharge is linearly proportional to the wind speed. The functions $v_1 = v_1(t, x)$ and $v_2 = v_2(t, x)$ describe the wind speed along the axes x_1 and x_2. The additive term $v_3 c$, where $v_3 = v_3(t, x)$, characterizes the amount of radiation spread into the system by radioactive particles already deposited at various sites of the monitored area. We assume the coordinate system centered at the leak source.

If v_3 depends on time only then, under mild growth conditions imposed on $v = (v_1, v_2)^T$, equation (2.1) has a unique solution described by the formula

$$\begin{aligned} c(t,x) = {} & \exp\left\{\int_0^t v_3(s)ds\right\} \\ & \times E\left[f(W_s)\exp\left\{\int_0^t v(t+s, W_s)dW_s \right.\right. \\ & \left.\left. - \frac{1}{2}\int_0^t \|v(t+s, W_s)\|^2 ds\right\} \,\middle|\, W_s = x\right], \end{aligned} \tag{2.2}$$

where $f = f(x)$ is the initial condition, $W_t = (W_{1,t}, W_{2,t})^T$ is a random vector of two independent Brownian motions and the first integral in (2.2) is in Itô's sense. Representation (2.2) is a consequence of the Feynman-Kac Theorem and Girsanov Theorem. See Karatzes and Shreve (1988), Section 5.7. The observed concentrations change randomly in time and space. If the wind speed vector is random, then so is c and we can consider (2.2) as a pathwise solution of (2.1). We will assume that with a suitable choice of the initial condition f it is possible to express (2.2) in the rather naive form

$$c(t, x) = \exp\{\mu(t, x) + z(t, x)\}, \tag{2.3}$$

where μ is a deterministic function and z is a zero mean stochastic process. This permits us to find a match between (2.1) and the data.

3. Modeling of stochasticity

The preliminary analysis of the data sampled in equal distance between 1984 and 1992 from two concentric circles with center in the plant (cf. Fig. 1) reveals that it is reasonable to consider $\mu(t, x)$ in (2.3) constant along the circles and model $z(t, x)$ as a stationary Gaussian temporal-spatial process. Let us denote by $u_{k,n}$ the logarithm of concentration observed in year t_k from a circle at location x_n, $k = 1, \ldots, K$, $n = 1, \ldots, N$. Because of the stationarity and the circular equidistant sampling, we can estimate the

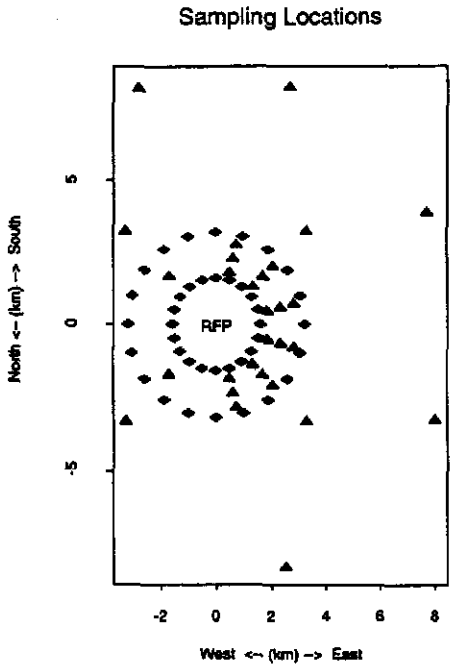

Fig. 1. The sampling locations. The diamonds form two concentric circles with radius 1.6 and 3.2 kilometers. Sampling there was conducted during 1984–1992.

covariance between the random variables $z(t_{k+1}, x_{h+1})$ and $z(t_1, x_1)$ on the same circle by means of the formula

$$\hat{R}(k,h) = \frac{1}{KN} \sum_{j=1}^{K-k} \sum_{n=1}^{N} (u_{j+k,n+h} - \bar{u})(u_{j,n} - \bar{u}), \tag{3.1}$$

where we define $u_{j,N+h} = u_{j,h}$ and $\bar{u}$ is the sample mean. Notice that circularity of the sampling path allows us to utilize all the N observations on the circle. A typical plot for $k = 0$ is shown in Fig. 2. The plot is symmetric because we assume stationarity and the maximum distance between two observations on a circle is limited by the diameter of the circle. Figure 2 gives us an idea about the correlation between the logarithms of the observed concentrations. Plots like these are used in time series analysis quite often as a part of the procedure known also as the Box-Jenkins method. Surprisingly, in kriging and spatial data analysis they occur rather rarely. The graphs in Fig. 2 resemble a Bessel function. Therefore we model the spatial covariance structure of the process z by means of the function $\sigma^2 R_\theta$, where R_θ is defined using the Bessel function of the first kind order zero:

$$R_\theta(0,x) = \frac{1}{\theta_1^2 + \theta_2^2} J_0(\theta_2|x|), \tag{3.2}$$

where $|x| = \sqrt{x_1^2 + x_2^2}$ and $\theta_1, \theta_2 > 0$. A plot of the theoretical covariance function for an estimated value of θ is in Fig. 3.

We recall that $J_0(\theta_2|x|)$ is defined by the relation

$$J_0(\theta_2|x|) = \frac{1}{\pi} \int_0^\pi \cos(\theta_2|x| \sin(\beta)) d\beta. \tag{3.3}$$

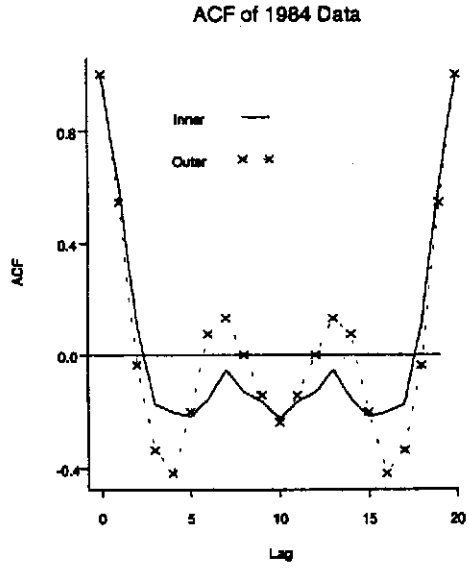

Fig. 2. The empirical autocorrelation function estimated using (3.1) from the circular observations sampled in 1984.

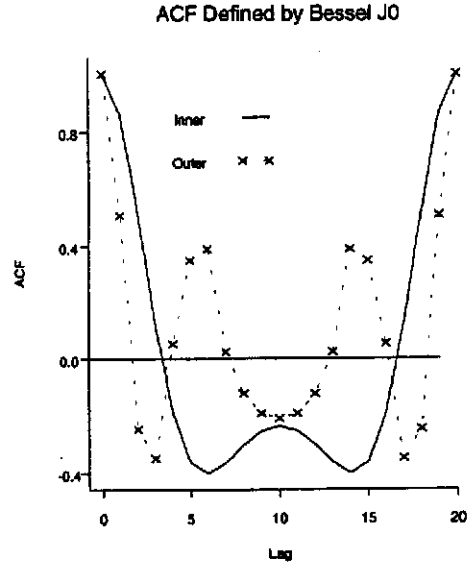

Fig. 3. The theoretical autocorrelation function modeled by the Bessel function $J_0(\theta|x|)$ with $\theta = 1.3$.

A direct investigation of the temporal correlation is complicated by irregular sampling. We do not have observations from more than nine consecutive years at one location. In fact we have no data available from 1979, 1982 and 1983. Hence, the method which led us to the Bessel function cannot be applied due to the small number of temporal observations. To determine the covariance function $R(t,x)$ of $z(t,x)$ in time let us consider the stochastic advection-diffusion equation

$$\left(\frac{\partial}{\partial t} - \frac{1}{2}\left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} - \theta_1^2\right)\right) z dt dx = \sigma Z(dt, dx) \tag{3.4}$$

with a white noise orthogonal measure Z on the right side. This equation may be interpreted in terms of Schwartz distributions. See Itô (1984) and Walsh (1986). Suppose (erroneously) that (3.4) has a sufficiently smooth solution. In addition let $R(t-s, x-y) = Ez(t,x)z(s,y)$. If we multiply both sides of (3.4) by $z(s,y)$, take the expectation and interchange the order of expectation and differentiation symbols we get the equation

$$\left(\frac{\partial}{\partial t} - \frac{1}{2}\left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} - \theta_1^2\right)\right) R = 0. \tag{3.5}$$

This is an analogy of the Yule-Walker equation in the time-series theory. See also Mohapl (1994, 1999). We want the solution of (3.5) to agree with (3.2) at least for $t = 0$. Equation (3.5) with initial condition (3.2) has a unique solution

$$R_\theta(t,x) = \frac{1}{\theta_1^2 + \theta_2^2} \int_{R^2} J_0(\theta_2|y|) \times \frac{1}{2\pi t} \exp\left\{-\frac{1}{2}\frac{|x-y|^2}{t} - \theta_1^2 t\right\} dy \tag{3.6}$$

for every $t > 0$ and $x \in R^2$. Application of $J_0(\theta|x|)$ in spatial data analysis is advocated e.g. in Guttorp (1994). The possibility to represent every homogeneous and isotropic covariance function using J_0 is pointed out in Matérn (1986).

PROPOSITION 3.1. *The function R_θ defined in* (3.6) *determines a stationary temporal-spatial process.*

PROOF. Let S_θ be the circle $\{\omega : |\omega| = \theta_2, \omega \in \boldsymbol{R}^2\}$. We leave the reader to verify the familiar identity

$$J_0(\theta_2|x|) = \frac{1}{2\pi}\int_{S_\theta} e^{i\langle x,\omega\rangle}d\omega, \tag{3.7}$$

where $\langle x,\omega\rangle = x_1\omega_1 + x_2\omega_2$ is the standard inner-product on $\boldsymbol{R}^2$. If we denote by χ_{S_θ} the characteristic function of S_θ, i.e. $\chi_{S_\theta}(\omega) = 1$ if $\omega \in S_\theta$ and $\chi_{S_\theta}(\omega) = 0$ otherwise, then the right side of (3.7) is the backwards Fourier transform of χ_{S_θ}. Using the elementary Fourier transform calculus we obtain:

$$\begin{aligned} R_\theta &= \mathcal{F}^{-1}(\chi_{S_\theta}) * \mathcal{F}^{-1}(e^{-(|\omega|^2+\theta_1^2)t/2}) \\ &= \mathcal{F}^{-1}(\chi_{S_\theta}e^{-(|\omega|^2+\theta_1^2)t/2}). \end{aligned} \tag{3.8}$$

The inverse Fourier transform works on ω while t is fixed. The right side of (3.8) may be further evaluated so that

$$R_\theta(t,x) = \frac{1}{\theta_1^2+\theta_2^2}e^{-(\theta_1^2+\theta_2^2)t/2}J_0(\theta_2|x|) \tag{3.9}$$

for $t \geq 0$. The proof is thus complete.

Proposition 3.1 is somewhat surprising, because the arguments in Whittle (1962) indicate that equation (3.4) does not define a stationary process with a finite variance. The explanation is:

PROPOSITION 3.2. *Let Z be a white noise orthogonal measure which is supported by the Borel σ-algebra of subsets of S_θ and is zero outside S_θ. Let $E(Z(dt,d\omega))^2 = \sigma dtd\omega/2\pi$. Define a measure-valued process $\{\epsilon(\cdot,x) : x \in \boldsymbol{R}^2\}$ by the relation*

$$\epsilon_\theta(\cdot,x) = \int_{S_\theta} e^{i\langle x,\omega\rangle}Z(\cdot,d\omega) \tag{3.10}$$

and consider the stochastic advection-diffusion equation

$$\left(\frac{\partial}{\partial t} - \frac{1}{2}\left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} - \theta_1^2\right)\right) z dtdx = \epsilon_\theta(dt,x). \tag{3.11}$$

If $z_0(x)$ is independent of Z and has a covariance function $\sigma^2 J_0(\theta_2|x|)/(\theta_1^2+\theta_2^2)$ then

$$\begin{aligned} z(t,x) = &\int_{\boldsymbol{R}^2} T(t,x-y)z_0(y)dy \\ &+ \int_0^t \int_{\boldsymbol{R}^2} T(t-s,x-y)\epsilon(ds,y)dy, \end{aligned} \tag{3.12}$$

where

$$T(t,x) = \frac{1}{2\pi t}\exp\left\{-\frac{|x|^2}{2t} - \theta_1^2 t\right\}, \tag{3.13}$$

is a stationary solution of (3.11) *and has correlation function* (3.9).

PROOF. The solution of (3.11) is defined e.g. in Walsh (1986) or Mohapl (1994, 1999). The proof is a direct consequence of Proposition 3.4 in Mohapl (1994). In terms of Schwartz distributions, equation (3.11) can be written as

$$\xi(t) - \xi(0) = \int_0^t A'\xi(s)ds + B'W(t), \tag{3.14}$$

where $\xi(t)$ and $W(t)$ are distribution-valued stochastic processes. Roughly speaking, the equation is obtained from (3.11) if we multiply both sides of (3.11) by a rapidly decreasing function ϕ, integrate over $\boldsymbol{R}^2$ and integrate the left side by parts. This procedure removes the differentiability requirement from z and leads to an equality between integrals. The rapidly decreasing functions form a linear space $\mathcal{D}$ with the known Schwartz topology. We might also say that we replaced the variable $x \in \boldsymbol{R}^2$ by the variable $\phi \in \mathcal{D}$. In this translation

$$\xi(t)\phi = \int_{\boldsymbol{R}^2} z(t,x)\phi(x)ds, \tag{3.15}$$

$$A'c(t)\phi = \int_{\boldsymbol{R}^2} \xi(t,x)\frac{1}{2}\left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} - \theta_1^2\right)\phi(x)dx \tag{3.16}$$

and

$$B'W(t)\phi = \int_0^t \int_{S_\theta} \mathcal{F}^{-1}(\phi)(\omega)Z(dt, d\omega), \tag{3.17}$$

or rather $B'W(t)$ is the continuous modification of this linear function of ϕ. Symbols A' and B' are referred to as dual operators. In this particular situation

$$A = \frac{1}{2}\left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} - \theta_1^2\right) \tag{3.18}$$

and $A = A'$. To determine B we calculate

$$\begin{aligned} E(B'W(t)\phi B'W(s)\psi) &= \rho(t,s)\int_{S_\theta} \mathcal{F}^{-1}(\phi)(\omega)\mathcal{F}^{-1}(\psi)(\omega)d\omega \\ &= \rho(t,s)\int_{\boldsymbol{R}^2} \phi(x)\int_{\boldsymbol{R}^2}\frac{1}{2\pi}\int_{S_\theta} e^{i\langle x-y,\omega\rangle}d\omega\psi(y)dydx \\ &= \rho(t,s)\int_{\boldsymbol{R}^2} \phi(x)B\psi(x)dx, \end{aligned} \tag{3.19}$$

where we define $\rho(t,s) = \sigma^2 \min(t,s)$ and

$$B\psi(x) = \int_{\boldsymbol{R}^2} \mathcal{F}^{-1}(\chi_{S_\theta})(x-y)\psi(y)dy. \tag{3.20}$$

To apply Proposition 3.4 in Mohapl (1994) we must determine the distribution of the initial condition $c(0)$. This is a distribution-valued random element with covariance operator

$$Q = B'(A + A')^{-1}B. \tag{3.21}$$

The proposition will be proved if we show that

$$Q\phi(x) = \int_{R^2} R_\theta(x-y)\phi(y)dy, \tag{3.22}$$

where R_θ is defined by (3.2).

According to (3.18),

$$(A+A')^{-1}\phi(x) = \mathcal{F}^{-1}\left(\frac{1}{|\omega|^2+\theta_1^2}\right) * \phi(x). \tag{3.23}$$

Combined with (3.20) this provides

$$\begin{aligned} B'(A+A')B\phi(x) &= \mathcal{F}^{-1}(\chi_{S_\theta}) * \mathcal{F}^{-1}\left(\frac{1}{|\omega|^2+\theta_1^2}\right) * \mathcal{F}^{-1}(\chi_{S_\theta}) * \phi(x) \\ &= \mathcal{F}^{-1}\left(\chi_{S_\theta}\frac{1}{|\omega|^2+\theta_1^2}\right) * \phi(x) \\ &= \frac{1}{\theta_1^2+\theta_2^2}\int_{R^2}\frac{1}{2\pi}\int_{S_\theta} e^{i\langle x-y,\omega\rangle}d\omega\phi(y)dy \\ &= \frac{1}{\theta_1^2+\theta_2^2}\int_{R^2} J_0(\theta_2|x-y|)\phi(y)dy \end{aligned} \tag{3.24}$$

according to (3.7). This completes the proof.

COROLLARY 3.1. *The process* (3.12) *admits for every* $x \in \boldsymbol{R}^2$ *and* $t \geq s \geq 0$ *a representation*

$$\begin{aligned} z(t,x) &= e^{-(\theta_1^2+\theta_2^2)(t-s)}z(s,x) \\ &\quad + \int_s^t\int_{S_\theta} e^{i\langle x,\omega\rangle-(\theta_1^2+\theta_2^2)(t-u)}Z(du,d\omega). \end{aligned} \tag{3.25}$$

PROOF. This follows by a direct calculation from (3.12) and (3.10).

Considerations in Whittle (1954) and others inspire us to associate the covariance function (3.9) with a formal differential equation

$$\left(\frac{\partial}{\partial t}+\vartheta_1^2\right)\sqrt{\left(\frac{\partial^2}{\partial x_1^2}+\frac{\partial^2}{\partial x_2^2}+\vartheta_2^2\right)}z = \epsilon, \tag{3.26}$$

where $\vartheta_1, \vartheta_2 > 0$ and ϵ is a "white noise". The equation has to do with the relation

$$\left(\frac{\partial^2}{\partial x_1^2}+\frac{\partial^2}{\partial x_2^2}+\vartheta^2\right)J_0(\vartheta|x|) = 0 \tag{3.27}$$

valid for every real number $\vartheta > 0$. It is interesting that Zhang (1995) and his advisor Jones found as best among all the models they fitted to the logarithms of the data a process with covariance function they associated with the equation

$$\left(\frac{\partial}{\partial t}+\vartheta_1^2\right)\left(\frac{\partial^2}{\partial x_1^2}+\frac{\partial^2}{\partial x_2^2}-\vartheta_2^2\right)z = \epsilon. \tag{3.28}$$

The process has obviously the same covariance structure in time as the solution of (3.11), which means what we try to improve using (3.9) is the spatial part of the model for the plutonium concentrations. The equation

$$\left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} - \vartheta^2\right) z = \epsilon \tag{3.29}$$

determines a process with non-negative correlation function $\vartheta|x|K_1(\vartheta|x|)$, where K_1 is the modified Bessel function of second kind order one. The Bessel function K_1 is quite popular in kriging and spatial data analysis. However, it is positive on its whole domain. Our estimates of the autocorrelation attain also negative values. See Fig. 2. That is why we expect to get better fit using the correlation $J_0(\vartheta|x|)$ instead.

The possibility to obtain from the advection-diffusion equation various models for the data by imposing conditions on the covariance structure of the input noise was observed already by Whittle (1962). Our choice is determined by the estimated covariance function plots.

4. Parameter estimation

Occasional difficulties in maximum likelihood estimation of a parameter in the covariance matrix of a spatial Gaussian process are discussed in Warnes and Ripley (1987). We encounter all of them when estimating θ: multimodality, roots of the score function in close distance from each other and ill-conditioness of the covariance matrix causing wild oscillations of the score function around the zero points.

First we estimated the parameters from individual years. We wanted to see whether there is a trend in the estimates suggesting ineptness to describe the mean and correlation of our sample by means of the limited number of parameters we intended to use. The function J_0 was chosen to fit as well as possible the empirical covariance function which was estimated from the circular observations sampled through 1984–1992. To assess relevance of the maximum likelihood estimates we also calculated θ_2 by minimizing the least-squares distance between the estimated and theoretical covariance functions and compared it with θ_2 obtained from the maximum likelihood estimating equation. See Fig. 4. For the data from 1970–1983 we do not have the empirical covariance function and that is why we use for comparison the quasi-likelihood estimating equations derived in Mohapl (1998). See Fig. 5.

Given the satisfactory results from the preliminary estimation we proceeded with calculation of the global maximum likelihood estimates of θ_1, θ_2 and β in the spatial trend. The maximum likelihood estimates were used for calculation and analysis of residual, quantile and autocorrelation plots.

The model for our discrete observations $c_{k,n}$ sampled at time t_k at location x_n is given by equation (2.3):

$$\ln c_{k,n} = z(t_k, x_n) + \mu_\beta(x_n). \tag{4.1}$$

The process z is the solution (3.12) of the stochastic advection-diffusion equation (3.11). The trend μ_β has general form

$$\mu_\beta(x) = \sum_{p=1}^{P} \beta_p \gamma_p(x), \tag{4.2}$$

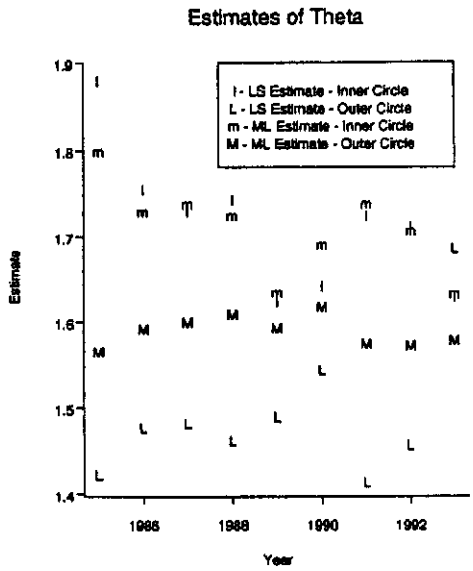

Fig. 4. Estimates of θ from circular observations. Comparison of least squares and maximum likelihood.

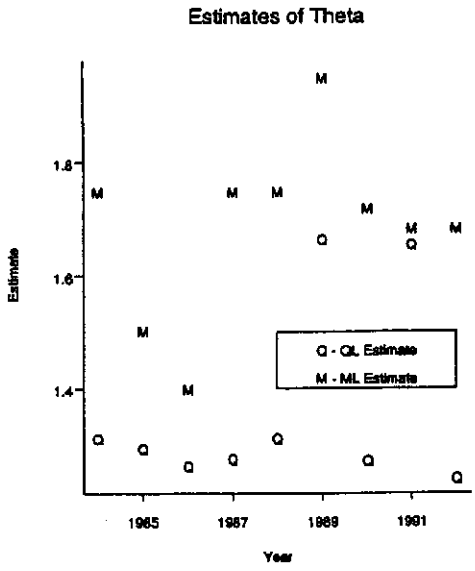

Fig. 5. Estimates of θ from individual planes. Comparison of the MLE and QL estimates.

where $\{\gamma_p, p = 1, \ldots, P\}$ are known base functions and the parameters $\{\beta_p, p = 1, \ldots, P\}$ are to be estimated.

Most of the data were sampled from two or three concentric circles. That does not provide much information for description of the average concentration changes with growing distance from the plant. But in 1989 and 1991 the concentrations were sampled in seven distances from the plant: 1.6, 1.9, 2.4, 2.9, 3.2, 4.64 and 8.64 kilometers. The distance from the plant introduces a natural stratification into the data and the assumption about constant mean along the circles allows us to obtain estimates of the means using plain averages over each stratum. The averages plotted against the distance from RFP are displayed in Fig. 6 along with the regression line. If our assumption about the correlation along circles is correct, the averages are maximum likelihood estimates. See the Appendix, Corollary A.1. Due to the reasonable fit of the regression line we choose

$$\mu_\beta(x) = -\beta_1 |x| + \beta_2, \tag{4.3}$$

where $\beta_1 > 0$.

4.1 *Preliminary estimation*

Prior to fitting the final model we estimated parameters from sub-samples to get an idea about the fit of our model and behavior of estimators. We started with estimation of θ from observations on individual circles. The covariance function of these observations is of the form $\sigma^2 R_\theta$, where $R_\theta(x) = J_0(\theta|x|)$ and $\sigma > 0$. It is obtained by reparametrization from the covariance function defined by (3.2) if we replace $\sigma^2/(\theta_1^2 + \theta_2^2)$ with σ^2 and set $\theta_2 = \theta$. The covariance matrix defined by $J_0(\theta|x|)$ for observations on the circles in Fig. 1 is for small values of θ ill-conditioned in sense of Press *et al.* (1986), Section 2.9. A problem which can be handled using the single value decomposition. See also the Appendix of this paper. We calculated the root of the likelihood function by the

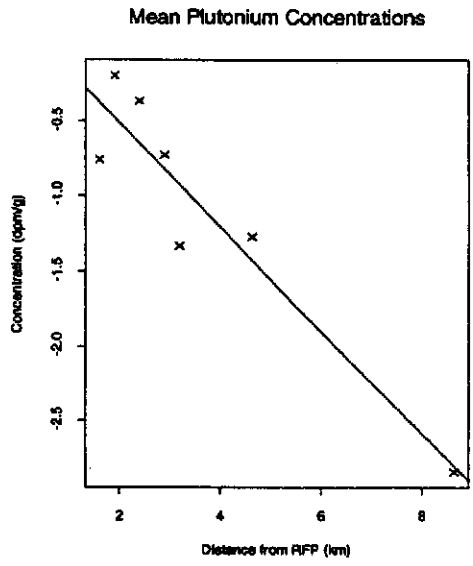

Fig. 6. Mean plutonium concentrations in different distance from RFP estimated from the 1989 sample.

simplest possible method, which is division of the interval containing the root. The ill-conditioness of the covariance matrix caused the score function to oscillate wildly around the zero point where we expected the root. In most cases we were able to separate at least two sets of roots. The criteria for choosing the most relevant maximum likelihood estimate (MLE) was $\hat{\sigma}^2$, the estimate of σ^2. We wanted $\hat{\sigma}^2$ as small as possible. Then we calculated the minimum of the function

$$f(\theta) = \sum_{n=0}^{N-1} (\hat{R}(d_n) - R_\theta(d_n))^2, \tag{4.4}$$

where $d_n = 2r\sin(n\pi\alpha/N)$ is the length of the circular trajectory between two observations on a circle with radius r, and compared it to the MLE. We recall that $r = 1.6$ and 3.2, respectively, $N = 20$ and the observations are in an equal distance from each other. We call the minimum of f in (4.4) the least squares (LS) estimate. The LS estimate minimizes the squared distance between the true and estimated autocorrelation and can thus provide an idea about relevance of the choice of MLE. The estimating equation arising from (4.4) is unbiased and has also several roots, but they are not difficult to calculate and it is easy to choose a proper one according to the autocorrelation plot. The LS estimates we obtained using (4.4) are shown in Fig. 4.

The LS estimator has a close link to the smoothed periodogram estimator. See e.g. Heyde and Gay (1993). Figure 4 indicates absence of an increasing or decreasing trend in the estimates suggesting a time dependence of θ. The shift between LS estimates from inner and outer circles may reflect dependence of the parameter on the distance from the plant. The estimates in Fig. 4 were used for elementary diagnostics like calculation of residuals and inspection of residual plots, normality checkups, comparison of empirical and theoretical autocorrelation function etc. See Box and Jenkins (1970) for details.

It is worthwhile to note that the MLE's and LS estimates obtained from the inner circles agree fairly well, whereas MLE's from the outer circle are shifted. We found

another set of MLE's in close vicinity of 1.45 which also agreed with the LS estimates fairly reasonably, but provided a somewhat larger $\hat{\sigma}^2$. The common $-2\ln$ likelihood function for a column vector $u = (u_1, \ldots, u_N)^T$ of observations from a Gaussian process with covariance function (3.2) and mean (4.2) is

$$l_N(u, \beta, \theta, \sigma) = N\ln(2\pi) + N\ln(\sigma^2) + \ln(\det(\Gamma_\theta)) + \frac{1}{2\sigma^2}(u - \mu_\beta)^T\Gamma_\theta^{-1}(u - \mu_\beta), \tag{4.5}$$

where Γ_θ is the covariance matrix with components

$$\gamma_{n,m}(\theta) = J_0(\theta|x_n - x_m|) \tag{4.6}$$

and

$$\mu_\beta = X\beta. \tag{4.7}$$

The matrix X of explanatory variables has N-dimensional column vectors made up of $\{\gamma_p(x_n) : n = 1, \ldots, N\}$, $p = 1, \ldots, P$. The estimating equations for the parameters in the mean obtained by minimization of the $-2\ln$-likelihood are

$$\hat{\beta} = (X^T\Gamma_\theta^{-1}X)^{-1}X^T\Gamma_\theta^{-1}u. \tag{4.8}$$

Parameters σ and θ are estimated using

$$\hat{\sigma}^2(\theta) = \frac{1}{N}(u - \mu_{\hat{\beta}})^T\Gamma_\theta^{-1}(u - \mu_{\hat{\beta}}), \tag{4.9}$$

$$\sum_{n=1}^{N}\sum_{m=1}^{N}\gamma_{n,m}(\theta)v_{n,m}(\theta) - \frac{1}{\hat{\sigma}^2(\theta)}(u - \mu_{\hat{\beta}})^T V_\theta(u - \mu_{\hat{\beta}}) = 0, \tag{4.10}$$

where $v_{n,m}(\theta)$ are the components of the matrix

$$V_\theta = \Gamma_\theta^{-1}\partial_\theta\Gamma_\theta\Gamma_\theta^{-1} \tag{4.11}$$

and the elements of $\partial_\theta\Gamma_\theta$ are composed of Γ_θ elements differentiated with respect to θ.

Formulas (4.8), (4.9) and (4.10) were used to calculate the values plotted in Fig. 6. The MLE's of β, σ^2 and θ are in Table 3 in the Appendix. To get a feedback about relevance of the MLE's we calculated another set of estimates using the quasi-likelihood (QL) functions suggested in Mohapl (1998):

$$\frac{1}{N}(u - \mu_{\hat{\beta}})^T(u - \mu_{\hat{\beta}}) - \sigma^2 R_\theta(0) = 0, \tag{4.12}$$

$$\sum_{n=1}^{N}\sum_{m=1}^{N}k(x_n - x_m)((u_n - \mu_\beta(x_n))(u_m - \mu_\beta(x_m)) - \sigma^2 R_\theta(x_n - x_m)) = 0. \tag{4.13}$$

The mean μ_β is calculated according to (4.3). We choose $k(x) = J_0(|x|)$. Relation (4.12) produces the familiar moment estimator, arguments leading to (4.13) and a simulation study are given in Mohapl (1998). Table 4 containing the QL estimates is given in the Appendix. Graphical comparison of the MLE's and QL estimates of θ is shown in Fig. 5. The QL function failed to have a root when the sample size was small. The sample sizes used for calculation of the estimates are in Tables 3 and 4, respectively. The QL estimator seems to be more sensitive to the sample size. More details can be found in the Appendix.

4.2 *Maximum likelihood estimation*

The above estimating equations are designed for a set of spatial observations sampled at a fixed year around the plant. Next we construct the overall likelihood function for the temporal-spatial data set. The procedure is rather standard. Therefore we present only what was used for our data and leave the possible generalizations to the reader. Simple autoregression analysis of the observations in time indicates very little temporal dependence. See also Jones and Zhang (1994), p. 4. What else, we have a relatively short period of observation. This motivates us to utilize the fact that the estimated value of θ_2 is between 1.5 and 2 and set $\theta_1 = \theta_2$. This way we obtain a simple model with a reasonable fit and interpretation.

As we mentioned, in this section we assume that our model has mean (4.2) and covariance $Ez(0,0)z(t,x) = \sigma^2 R_\theta(t,x)$, where

$$R_\theta(t,x) = \frac{1}{2\theta^2} e^{-\theta^2|t|} J_0(\theta|x|), \tag{4.14}$$

and $\sigma, \theta > 0$ are parameters to be estimated. According to (3.25), for each pair $t \geq s \geq 0$,

$$\begin{aligned} E(z(t,x) - e^{-\theta^2(t-s)} z(s,x))(z(t,y) - e^{-\theta^2(t-s)} z(s,y)) \\ = \sigma^2(1 - e^{-2\theta^2(t-s)}) J_0(\theta|x-y|). \end{aligned} \tag{4.15}$$

Let us denote

$$z_t = (z(t,x_1), \ldots, z(t,x_N))^T. \tag{4.16}$$

In consequence of (4.15),

$$z_t - e^{-\theta^2(t-s)} z_s \sim \mathcal{N}(0, \sigma^2(1 - e^{-2\theta^2(t-s)})\Gamma_\theta), \tag{4.17}$$

where $\mathcal{N}$ denotes the multivariate normal distribution. Simultaneously,

$$z_s \sim \mathcal{N}(0, \sigma^2 \Gamma_\theta). \tag{4.18}$$

Using the Markov property of the process z and the standard chain rule for conditional probabilities we get the joint probability density function of $z = \{z(t_k, x_n) : k = 1, \ldots, K, n = 1, \ldots, N\}$:

$$\begin{aligned} f(z,\theta,\sigma) = {} & \frac{1}{(2\pi\sigma^2)^{N/2} \det(\Gamma_\theta)} \exp\left\{ -\frac{1}{2\sigma^2} z_{t_1}^T \Gamma_\theta^{-1} z_{t_1} \right\} \\ & \times \left[\frac{1}{(2\pi\sigma^2)^{N/2} \det(\Gamma_\theta)} \right]^{K-1} \left[\prod_{k=2}^{K} \frac{1}{\sqrt{1 - e^{-2\theta^2(t_k - t_{k-1})}}} \right] \\ & \times \exp\Bigg\{ - \sum_{k=2}^{K} \frac{1}{2\sigma^2(1 - e^{-2\theta^2(t_k - t_{k-1})})} \\ & \quad \times (z_{t_k} - e^{-\theta^2(t_k - t_{k-1})} z_{t_{k-1}})^T \\ & \quad \times \Gamma_\theta^{-1} (z_{t_k} - e^{-\theta^2(t_k - t_{k-1})} z_{t_{k-1}}) \Bigg\}. \end{aligned} \tag{4.19}$$

Set $u_k = (u_{k,1}, \ldots, u_{k,N})^T$ and denote $u_{k,n} = \ln c_{k,n}$. The MLE of β is

$$\hat{\beta} = \frac{1}{K}(X^T\Gamma_\theta^{-1}X)^{-1}X^T\Gamma_\theta^{-1}u_1 + \frac{1}{K}\sum_{k=2}^{K}\frac{1}{1-e^{-\theta^2(t_k-t_{k-1})}} \times (X^T\Gamma_\theta^{-1}X)^{-1}X^T\Gamma_\theta^{-1}(u_k - e^{-\theta^2(t_k-t_{k-1})}u_{k-1}). \tag{4.20}$$

If we set

$$\bar{u}_k = u_k - \mu_{\hat{\beta}}, \tag{4.21}$$

then

$$\hat{\sigma}^2(\theta) = \frac{1}{KN}\left\{\bar{u}_1^T\Gamma_\theta^{-1}\bar{u}_1 + \sum_{k=2}^{K}\frac{1}{1-e^{-2\theta^2(t_k-t_{k-1})}} \times (\bar{u}_k - e^{-\theta^2(t_k-t_{k-1})}\bar{u}_{k-1})^T \times \Gamma_\theta^{-1}(\bar{u}_k - e^{-\theta^2(t_k-t_{k-1})}\bar{u}_{k-1})\right\}. \tag{4.22}$$

The parameter θ is estimated from the equation

$$K\sum_{n=1}^{N}\sum_{m=1}^{N}\gamma_{n,m}(\theta)v_{n,m}(\theta) + 4\theta\sum_{k=2}^{K}\frac{e^{-2\theta^2(t_k-t_{k-1})}}{1-e^{-2\theta^2(t_k-t_{k-1})}} - \frac{4\theta}{\hat{\sigma}^2(\theta)}\sum_{k=2}^{K}\frac{e^{-2\theta^2(t_k-t_{k-1})}}{(1-e^{-2\theta^2(t_k-t_{k-1})})^2} \times (\bar{u}_k - e^{-\theta^2(t_k-t_{k-1})}\bar{u}_{k-1})^T\Gamma_\theta^{-1}(\bar{u}_k - e^{-\theta^2(t_k-t_{k-1})}\bar{u}_{k-1}) + \frac{1}{\hat{\sigma}^2(\theta)}\sum_{k=2}^{K}(1-e^{-2\theta^2(t_k-t_{k-1})}) \times \{2\theta e^{-\theta^2(t_k-t_{k-1})}\bar{u}_{k-1}^T\Gamma_\theta^{-1}(\bar{u}_k - e^{-\theta^2(t_k-t_{k-1})}\bar{u}_{k-1}) - (\bar{u}_k - e^{-\theta^2(t_k-t_{k-1})}\bar{u}_{k-1})^T V_\theta(\bar{u}_k - e^{-\theta^2(t_k-t_{k-1})}\bar{u}_{k-1})\} - \frac{1}{\hat{\sigma}^2(\theta)}\bar{u}_1 V_\theta \bar{u}_1 = 0. \tag{4.23}$$

The above formulas can be extended to the case when the spatial sampling scheme is fixed but a different number of observations is sampled each year (missing observations case).

To compare our model to the one used by Jones and Zhang (1994) we first analyzed the data sampled from the two concentric circles with radius 1.6 and 3.2 kilometer. We have 20 data points on each circle sampled in 9 subsequent years 1984–1992. This means we can apply formulas (4.20), (4.22) and (4.23) with $N = 40$ and $K = 9$ and estimate the parameters θ, β and σ. We found two sets of MLE's summarized in the next table.

If we realize that our model has only four parameters while the model suggested by Jones and Zhang (1994) has eleven, then our model fits very well. The interesting thing

is that the lower θ estimate leads to a lower estimate of σ^2 than the higher estimate. From Figs. 3, 4 and 5 we can see that both sets of estimates are quite relevant. Another finding is that the quasi-likelihood estimates of θ from individual years computed by (4.12) do point rather to the smaller estimate of θ.

To estimate the parameters for the whole data set we used the composite likelihood instead of the global likelihood function. This move may be justified by the low temporal correlation and presumption that lost of efficiency caused by use of a composite score function is compensated by numerical precision we gain when avoiding numerous inversion of the ill-conditioned covariance matrix for spatial observations. In general, composite likelihood estimators are not as efficient as MLE's. See Lindsay (1988). But in situations like ours they are much easier to calculate. Our data set can be divided into four main subclasses. Three of them contain yearly data with the same number of observations, the rest consists of observations from some 5 years that do not fit in the previous categories. The spatial sample size N was 13, 40 and 73 for the first three, respectively, one of the remaining five samples had $N = 53$, in the rest $N \leq 10$. We obtain the composite likelihood by neglecting the correlation between the subsamples and treating them as independent. It is easy to see that the composite likelihood estimators of β and σ are a plain average of the MLE's for the individual subsamples, θ must be estimated from the composite score function which sums up the individual scores for θ. We obtained two sets of estimates:

The results agree well with the MLE's in Table 1. When comparing the $-2\ln$-likelihood values in Table 2 with those in Jones and Zhang (1994) we should keep in mind that our model has only four parameters whereas those considered by Jones and Zhang have between ten and fourteen. The apparent increase of $\hat{\sigma}^2$ has to do with the trend in our observations. On the two circles our estimate of trend agrees with the stratified mean. For the whole data set the fit of a linear trend is not that good. Although $\hat{\sigma}^2$ is smaller for $\hat{\theta} = 1.255$, analysis of the residual and autoregression plots we can construct for the observations on circles with $r = 1.6$ and $r = 3.2$ shows that it is more appropriate to choose the set of estimates with larger $\hat{\theta}$. See Figs. 7–10.

Table 1. The maximum likelihood estimates from the concentric circles with radius 1.6 and 3.2 kilometers estimated using formulas (4.20), (4.22) and (4.23). Total number of observations is 360.

$\hat{\theta}$	1.314	1.773
$\hat{\beta}_1$	0.498	0.502
$\hat{\beta}_2$	−0.219	−0.214
$\hat{\sigma}^2$	0.223	0.262
$-2\ln$ likelihood	586.082	655.242

Table 2. The composite likelihood estimates from the whole sample. The total number of observations is 559.

$\hat{\theta}$	1.255	1.662
$\hat{\beta}_1$	0.330	0.241
$\hat{\beta}_2$	−0.117	−0.481
$\hat{\sigma}^2$	1.626	2.421
$-2\ln$ likelihood	1715.331	2014.035

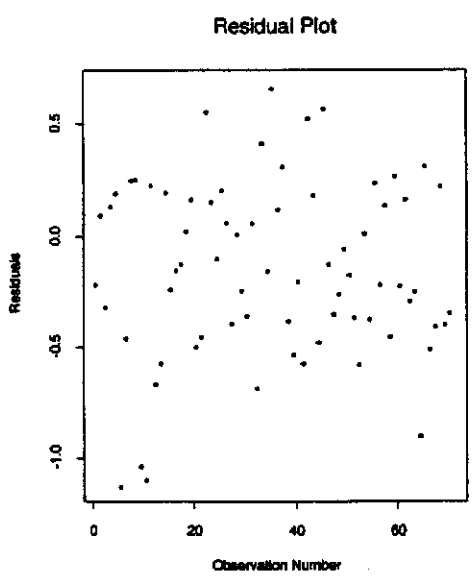

Fig. 7. Residual plot for 71 logarithms of observations sampled in 1991. The residuals are obtained by means of the final model (4.24).

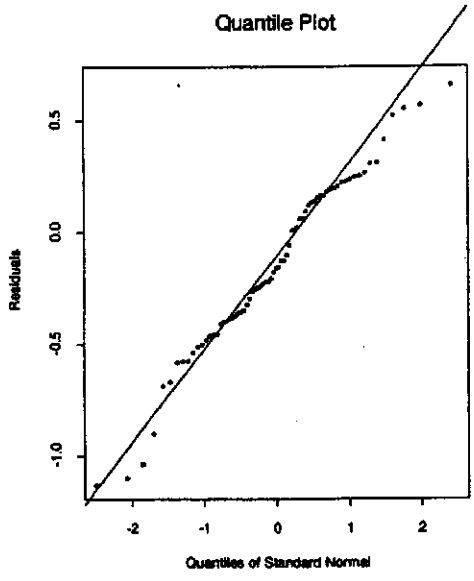

Fig. 8. Quantile plot for 71 logarithms of observations sampled in 1991. The residuals are obtained by means of the final model (4.24).

In consequence of the above arguments, the final model we propose for the RFP Soil Plutonium Data is

$$c(t,x) = \exp\{z(t,x) - 0.241|x| - 0.481\}, \tag{4.24}$$

where $z(t,x)$ is the stationary solution of (3.11) with $\theta_1 = \theta_2$ and covariance function $\sigma^2 R_\theta(t,x)$ described by (4.14). We take $\sigma^2 = 2.421$ and $\theta = 1.662$.

Once we estimated the parameters and accept the assumption that z is a Gaussian process then we can use the logarithms of observations and formula (4.1) to predict further logarithmized data and obtain the properly scaled predictions by exponentiating. This is a rather standard procedure based on Theorem 2.5.1 in Anderson (1958). The result is visualized in Figs. 11 and 12.

5. Advection-diffusion model

Next we return to the general physical model described by the advection-diffusion equation. In Section 3 we derived from the data a Gaussian process $z(t,x)$, in Section 4 we estimated its parameters and next we use $z(t,x)$ to obtain the coefficients v_1, v_2 and v_3 of the equation (2.1) introduced in Section 2. We proceed by differentiation of the solution (2.3). This will provide an insight into the physical meaning of (2.3) and its relation to the actual advection-diffusion transportation process.

PROPOSITION 5.1. *Suppose that γ is a differentiable function. Then the process*

$$c(t,x) = \exp\{z(t,x) - \beta_1\gamma(|x|) + \beta_2\} \tag{5.1}$$

satisfies an advection-diffusion equation

$$\frac{\partial}{\partial t}c = \left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2}\right)c - \left(v_1\frac{\partial}{\partial x_1} + v_2\frac{\partial}{\partial x_2}\right)c + v_3 c. \tag{5.2}$$

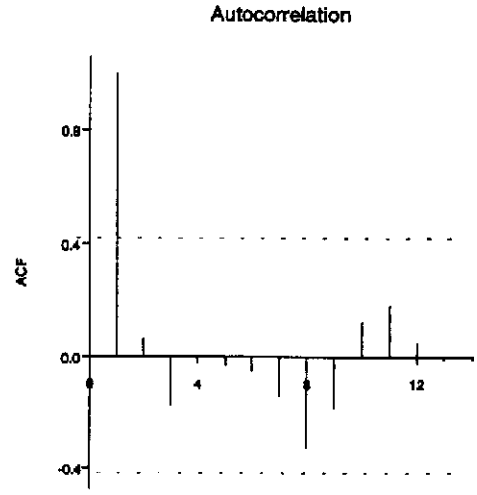

Fig. 9. Autocorrelation plot for a subset of 20 residuals presented in Figs. 7 and 8. The residuals are positioned on the circle with diameter 1.6.

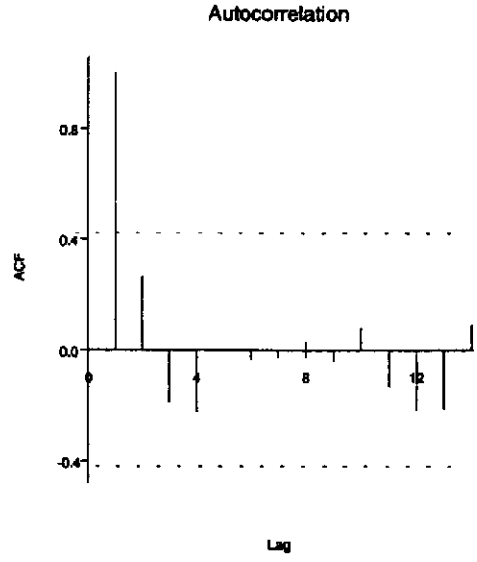

Fig. 10. Autocorrelation plot for a subset of 20 residuals presented in Figs. 7 and 8. The residuals are positioned on the circle with diameter 3.2.

The equation holds almost surely in the sense of Schwartz distributions. The coefficients v_1, v_2 *and* v_3 *are random and depend on time and space*:

$$v_1(t,x) = \frac{\partial}{\partial x_1} z(t,x) - \beta_1 \frac{x_1}{|x|} \dot{\gamma}(|x|), \tag{5.3}$$

$$v_2(t,x) = \frac{\partial}{\partial x_2} z(t,x) - \beta_1 \frac{x_2}{|x|} \dot{\gamma}(|x|) \tag{5.4}$$

and

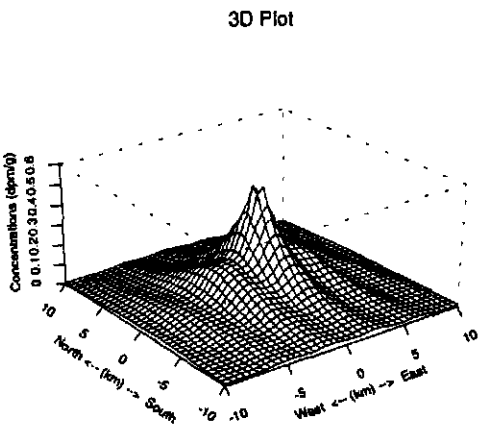

Fig. 11. A 3D plot for concentrations observed in 1991. We used the observations to predict values on a 21 by 21 point rectangular net with center in zero and interpolated the surface.

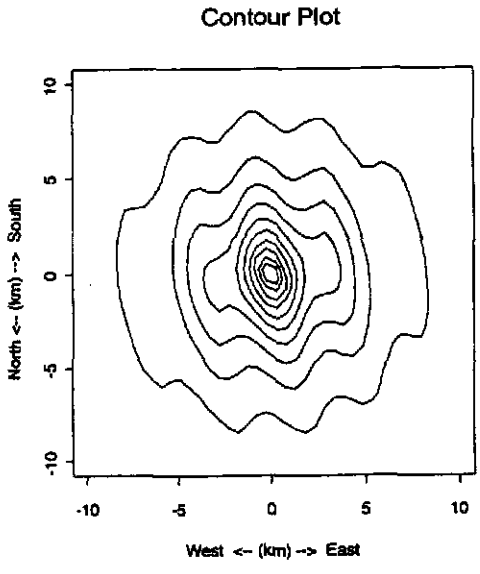

Fig. 12. Contour plot for concentrations observed in 1991. We used the observations to predict values on a 21 by 21 point rectangular net with center in zero and interpolated the contours. The contour lines are 0.05 mdg/g from each other.

$$v_3(dt,x) = \beta_1\ddot{\gamma}(|x|) + \frac{\beta_1}{|x|}\dot{\gamma}(|x|) + J_0((\theta_1^2+\theta_2^2)|x|) + \frac{1}{2}(\theta_2^2-\theta_1^2)z(t,x) + \epsilon_\theta(dt,x). \tag{5.5}$$

PROOF. The process z is, due to Corollary 3.1, an Itô process in time for each fixed $x \in \boldsymbol{R}^2$ and therefore, according to the Itô's formula,

$$\frac{\partial}{\partial t}c = c\frac{\partial}{\partial t}z + \frac{1}{2}J_0((\theta_1^2+\theta_2^2)|x|)c. \tag{5.6}$$

The process ϵ_θ is infinitely many times differentiable in the x variables. This property is inherited by the processes z and c, respectively. Plain differentiation provides:

$$\frac{\partial}{\partial x_1}c = v_1 c, \tag{5.7}$$

$$\frac{\partial^2}{\partial x_1^2}c = \left(\frac{\partial^2}{\partial x_1^2}z - \frac{\partial}{\partial x_1}\left[\frac{\beta_1 x_1}{\sqrt{x_1^2+x_2^2}}\dot{\gamma}(|x|)\right]\right)c + v_1\frac{\partial}{\partial x_1}c. \tag{5.8}$$

The function $\dot{\gamma}$ denotes the derivative of γ. Analogically for x_2:

$$\frac{\partial^2}{\partial x_2^2}c = \left(\frac{\partial^2}{\partial x_2^2}z - \frac{\partial}{\partial x_2}\left[\frac{\beta_1 x_2}{\sqrt{x_1^2+x_2^2}}\dot{\gamma}(|x|)\right]\right)c + v_2\frac{\partial}{\partial x_2}c. \tag{5.9}$$

Since

$$\left(\frac{\partial^2}{\partial x_1^2}+\frac{\partial^2}{\partial x_2^2}+\theta_2^2\right)z=0, \tag{5.10}$$

and

$$\frac{\partial}{\partial t}z+\frac{1}{2}(\theta_1^2+\theta_2^2)z=\epsilon_\theta(dt,\cdot), \tag{5.11}$$

we have

$$\begin{aligned}\frac{\partial}{\partial t}c-\left(\frac{\partial^2}{\partial x_1^2}+\frac{\partial^2}{\partial x_2^2}\right)c&\\ =\left(\frac{\partial}{\partial t}z+\frac{1}{2}(\theta_1^2+\theta_2^2)z\right)c-\left(v_1\frac{\partial}{\partial x_1}+v_2\frac{\partial}{\partial x_2}\right)c&\\ +\left(\frac{\beta_1}{\sqrt{x_1^2+x_2^2}}\dot\gamma(|x|)+\beta_1\ddot\gamma(|x|)\right.&\\ \left.+J_0((\theta_1^2+\theta_2^2)|x|)+\frac{1}{2}(\theta_2^2-\theta_1^2)z\right)c&\\ =-\left(v_1\frac{\partial}{\partial x_1}+v_2\frac{\partial}{\partial x_2}\right)c+v_3c.&\end{aligned} \tag{5.12}$$

The last equality holds almost surely and completes the proof.

According to Section 1, v_1 and v_2 in equation (5.2) characterize the wind speed. Relations (5.3) and (5.4) say that $v=(v_1,v_2)$ at given time and space equals to the gradient of the logarithmized concentration values. Hence, in our final setting, v is a vector orthogonal to the contours of the logarithmized concentration surface. It characterizes the force distributing the contaminated particles from RFP into the area. The mean of v at a given place is a vector independent of time with components

$$Ev_1=\beta_1\frac{x_1}{|x|}\dot\gamma(|x|)\quad\text{and}\quad Ev_2=\beta_1\frac{x_2}{|x|}\dot\gamma(|x|) \tag{5.13}$$

we can interpret as average wind speed along the coordinates. The quantity

$$\sqrt{((Ev_1)^2+(Ev_2)^2)}=\beta_1|\dot\gamma(|x|)| \tag{5.14}$$

is the average magnitude of the wind velocity at location x. When $\gamma(r)=r$, the right side of (5.14) is constant and equals to β_1. Thus β_1 is related to the wind speed magnitude in the area. Notice that a plain linear trend $\beta_1x_1+\beta_2x_2+\beta_3$ provides v_1 and v_2 with a constant trend β_1 and β_2, respectively, and v_3, the Malthusian rate, with zero mean. However, that would violate the assumption about the constant mean of the logarithmized observations along circles with center in the origin.

By the last theorem, v_3 is a stochastic process with a trend that does not depend on time. The product v_3c can be interpreted as the radiation contributed into the system by particles already firmly deposited at various locations. The random part of v_3 reflects the possibility of random redepositing of the surface soil. The mean of v_3 is zero if and only if $\gamma(|x|)$ is a harmonic function, i.e. satisfies the Laplace equation with zero on the right hand side. If $\theta_1=\theta_2$ then v_3 is at each fixed location a plain Brownian motion in

time. If $\gamma(r) = r$ then $Ev_3 = \beta_1/|x|$. This means the pollution rate goes steeply up as we get close to the coordinate origin where the plant is located. This is explained by the fact that the highest plutonium amount is in the soil closest to RFP.

Appendix

Here we provide the supplementary relations used in calculation of the maximum likelihood estimates in Subsections 4.1 and 4.2. The relations are specific for the circular sampling scheme and they were used to overcome the difficulties caused by the ill-conditioned covariance matrix Γ_θ defined by the Bessel function $J_0(\theta|x|)$.

LEMMA A.1. *Let N be even and $\{x_1, \dots, x_N\}$ be a set of points in $\boldsymbol{R}^2$ distributed in equal distance from each other on a circle with diameter $r > 0$. Then the matrix Γ_θ with components* (4.6) *has eigenvalues*

$$\lambda_k(\theta) = \sum_{n=1}^{N} J_0\left(2\theta r \sin\left(\frac{\pi}{N}n\right)\right) \cos\left(\frac{2k\pi}{N}n\right), \tag{A.1}$$

$k = 1, \dots, N/2$. To each eigenvalue belong two eigenvectors

$$\left(\frac{1}{\sqrt{N}} \cos\left(\frac{2k\pi}{N}n\right)\right)_{n=1}^{N} \quad \text{and} \quad \left(\frac{1}{\sqrt{N}} \sin\left(\frac{2k\pi}{N}n\right)\right)_{n=1}^{N}, \tag{A.2}$$

$k = 1, \dots, N/2$.

PROOF. See Grenander and Rosenblatt (1984), Sections 3.4–3.5.

COROLLARY A.1. *Consider N observations with $-2\ln$ likelihood function* (4.5) *sampled in equal distance from a circle of radius r and let N be even.*

If $\mu_\beta = \mathbf{1}\beta$, where $\mathbf{1}$ is a column of ones then the MLE of β is the sample mean.

The matrix V_θ in (4.11) *has eigenvalues $\dot{\lambda}_k$ given by the formula*

$$\dot{\lambda}_k(\theta) = -2r \sum_{n=1}^{N} c_n(\theta) \cos\left(\frac{2k\pi}{N}n\right), \tag{A.3}$$

where

$$c_n(\theta) = \sin\left(\frac{\pi}{N}n\right) \frac{J_1\left(2\theta r \sin\left(\frac{\pi}{N}n\right)\right)}{J_0\left(2\theta r \sin\left(\frac{\pi}{N}n\right)\right)} \qquad n = 1, \dots, N, \tag{A.4}$$

$k = 1, \dots, N/2$. The corresponding eigenvectors are described by (A.2).

PROOF. The previous lemma asserts that $\mathbf{1}$ is an eigenvector of Γ_θ. The result is thus a straight consequence of formula (4.8). The second statement follows from the familiar single value decomposition.

Simple plots of the eigenvalues as functions of θ may clarify the degree of ill-conditioness of the covariance matrix as well as the unsettled behavior of the score

function (4.10). The above lemma can be generalized to the case of two concentric circles.

LEMMA A.2. *Suppose we have two concentric circles with diameters r_1 and r_2, respectively, $0 < r_1 < r_2$. We sample N observations from each circle, N is even. The locations of observation are equally spaced and under the same angle on both circles. Let us represent Γ_θ with components* (4.6) *in the form*

$$\Gamma_\theta = \begin{pmatrix} \Gamma_{1,\theta} & K_\theta \\ K_\theta^T & \Gamma_{2,\theta} \end{pmatrix}, \tag{A.5}$$

where $\Gamma_{i,\theta}$ is the covariance matrix of observations sampled on the circle with diameter r_i, $i = 1, 2$. We claim that

i) *$\Gamma_{1,\theta}$, $\Gamma_{2,\theta}$ and K_θ share the same system of eigenvectors and the eigenvalues of K_θ are non-negative,*

ii) *if λ_i, $i = 1, 2$ and k are eigenvalues for a fixed θ sharing a common (column) eigenvector e then there are constants c_1 and c_2 such that the column vector $E = (c_1 e^T, c_2 e^T)^T$ of dimension $2N$ is the eigenvector of Γ_θ,*

iii) *the pair (c_1, c_2) and the eigenvalue λ corresponding to E are calculated from the equation*

$$\begin{pmatrix} \lambda_1 & k \\ k & \lambda_2 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \lambda \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}. \tag{A.6}$$

We omit the proof and emphasize that λ and c above depend on θ. We have also no guarantee that Γ_θ is invertible. However, if we are given a particular set of sampling locations then we can use the previous Lemma to calculate k from the equality $K_\theta e = ke$, where e is the eigenvector corresponding to λ_1 and λ_2, and check if $k \approx \sqrt{\lambda_1 \lambda_2}$. If the approximate equality is true then Γ_θ may be singular.

COROLLARY A.2. *If Γ_θ is invertible and the vectors u_1, u_2 of observations were sampled from two concentric circles with radii $0 < r_1 < r_2$, respectively, at locations specified in the previous lemma then the MLE of β in* (4.2), (4.3) *can be calculated according to the formula*

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} -r_1 & 1 \\ -r_2 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \bar{u}_1 \\ \bar{u}_2 \end{pmatrix}, \tag{A.7}$$

where $\bar{u}_i$ is the sample mean of u_i, $i = 1, 2$.

PROOF. For sake of brevity we include the minus signs in (6.7) into the r's. In the standard vector notation used in regression problems $\mu_\beta = X\beta$, where

$$X = \begin{pmatrix} r_1 \mathbf{1} & \mathbf{1} \\ r_2 \mathbf{1} & \mathbf{1} \end{pmatrix} \tag{A.8}$$

and $\mathbf{1}$ is a column vector of N ones. The vector $\mathbf{1}$ is a common eigenvector for $\Gamma_{1,\theta}$ and $\Gamma_{2,\theta}$. According to Lemma A.2, i), there are constants c_1, c_2 such that

$$\begin{pmatrix} \Gamma_{1,\theta} & K_\theta \\ K_\theta^T & \Gamma_{2,\theta} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{1} \\ \mathbf{1} \end{pmatrix} = \begin{pmatrix} c_1 \mathbf{1} \\ c_2 \mathbf{1} \end{pmatrix}. \tag{A.9}$$

Table 3. The maximum likelihood estimates from individual planes calculated using formulas (4.8), (4.9) and (4.10).

Year	N	θ	β_1	β_2	σ^2
70	13	1.565	0.305	0.289	3.131
71	13	1.501	0.327	1.061	3.583
72	13	1.501	0.315	0.902	3.571
74	13	1.565	0.271	0.716	2.025
76	13	1.565	0.208	−0.398	1.969
77	13	1.501	0.374	0.769	2.008
80	13	1.501	0.355	0.168	2.287
81	12	1.667	0.047	−0.928	1.775
84	40	1.751	0.783	0.254	1.430
85	40	1.505	0.453	−0.398	0.225
86	53	1.401	0.189	−0.889	3.678
87	40	1.751	0.355	−0.192	0.707
88	40	1.751	0.498	−0.179	0.613
89	71	1.950	0.273	−0.291	2.282
90	40	1.720	0.428	−0.592	0.200
91	71	1.685	0.287	−0.699	2.987
92	40	1.685	0.558	−0.011	1.013

Table 4. The quasi-likelihood estimates from individual planes calculated using formulas (4.8), (4.12) and (4.13). For small sample sizes the equations had no solution.

Year	N	θ	β_1	β_2	σ^2
84	40	1.314	0.783	−0.254	3.273
85	40	1.296	0.453	0.398	1.937
86	53	1.266	0.513	0.019	2.127
87	40	1.278	0.355	0.192	1.609
88	40	1.314	0.498	0.179	2.394
89	71	1.667	0.482	0.384	1.913
90	40	1.275	0.428	0.592	1.981
91	71	1.657	0.506	−0.874	3.744
92	40	1.244	0.558	0.011	2.443

Consequently,

$$X^T\Gamma_\theta^{-1}X = \begin{pmatrix} c_1r_1\mathbf{1} & c_1\mathbf{1} \\ c_2r_1\mathbf{1} & c_2\mathbf{1} \end{pmatrix}^T \begin{pmatrix} r_1\mathbf{1} & \mathbf{1} \\ r_2\mathbf{1} & \mathbf{1} \end{pmatrix} = N \begin{pmatrix} c_1r_1 & c_1 \\ c_2r_1 & c_2 \end{pmatrix}^T \begin{pmatrix} r_1 & 1 \\ r_2 & 1 \end{pmatrix}. \tag{A.10}$$

Analogically we calculate

$$X^T\Gamma_\theta^{-1}u = N \begin{pmatrix} c_1r_1 & c_1 \\ c_2r_1 & c_2 \end{pmatrix}^T \begin{pmatrix} \bar{u}_1 \\ \bar{u}_2 \end{pmatrix}, \tag{A.11}$$

where $u = (u_1^T, u_2^T)^T$ is the vector of observations. The proof is now a direct consequence of (4.8).

Corollary A.2 asserts that the estimator (A.7) agrees with what is obtained by simply interpolating the stratified means received from individual circles using a straight line.

Finally we are in the position briefly describe application of the single value decomposition (SVD) procedure. The idea is (see Press *et al.* (1986)) to replace the eigenvector expansion of the image

$$\Gamma^{-1}x = \sum_{n=1}^{N} \lambda_k (x^T e_k) e_k \tag{A.12}$$

by its truncation to first K summands for which the ratio λ_1/λ_k, $k = 1, \ldots, K$, does not exceed a certain value c_K. We assume $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ with at least some λ positive. The Γ_θ is used for estimation of both mean and variance. Where possible, we utilized Corollaries A.1 and A.2 for estimation of β. Our choice of K was determined by the estimate of σ^2. We did choose c_K in a way which led to estimates of σ^2 that reasonably compared to the sample variance. Using SVD we naturally introduce a bias into our estimates. We noticed that compared to σ^2, estimates of θ and β were not quite sensitive to the choice of K. We did not attempt to adjust for the bias.

References

Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*, Willey, New York.

Box, G. E. P. and Jenkins, G. M. (1970). *Time Series Analysis, Forecasting and Control*, Holden Day, San Francisco.

Farrell, D. A., Woodbury, A. D., Sudicky, E. A. and Rivett, M. O. (1994). Stochastic and deterministic analysis of dispersion in unsteady flow at the Borden Tracer-Test site, *Journal of Contaminant Hydrology*, **15**, 159–185.

Grenander, U. and Rosenblatt, M. (1984). *Statistical Analysis of Stationary Time Series*, Chelsea, New York.

Guttorp, P. (1994). Comment to a paper by Handcock M. S. and Wallis J. R., *J. Amer. Statist. Assoc.*, **89**, 382–384.

Heyde, C. C. and Gay, R. (1993). Smoothed periodogram asymptotics and estimation for processes and fields with possible long-range dependence, *Stochastic Process. Appl.*, **45**, 169–182.

Huebner, M. and Rozovskii, B. L. (1995). On asymptotic properties of maximum likelihood estimators for parabolic stochastic *PDE*'s, *Probab. Theory Related Fields*, **103**, 143–163.

Itô, K. (1984). *Foundations of Stochastic Differential Equations in Infinite Dimensional Spaces, Society for Industrial and Applied Mathematics*, Philadelphia, Pennsylvania.

Itô, K. and Kunisch, K. (1990). The augmented Lagrangian method for parameter estimation in elliptic systems, *SIAM J. Control Optim.*, **28**, 113–136.

Jones, R. H. and Vecchia, A. V. (1993). Fitting Continuous ARMA models to unequally spaced spatial data, *J. Amer. Statist. Assoc.*, **88**, 947–954.

Jones, R. H. and Zhang, Y. (1994). Spatial and temporal analysis of the Rocky Flats Soil Plutonium Data, Report, Colorado Department of Public Health and Environment, Colorado.

Jones, R. H. and Zhang, Y. (1997). Models for continuous stationary space-time processes, *Modeling Longitudinal and Spatially Correlated Data: Methods, Applications and Future Directions* (eds. T. G. Gregoire, P. Bickel, P. Diggle, S. Fienberg, K. Krickeberg, I. Olkin, N. Wermuth and S. Zeger), Lecture Notes in Statist., **122**, 289–298, Springer, New York.

Kallianpur, G. and Xiong, J. (1995). *Stochastic Differential Equations in Infinite Dimensional Spaces*, IMS Lecture Notes—Monograph Series, Vol. 26, Hayward, California.

Karatzas, I. and Shreve, S. E. (1988). *Brownian Motion and Stochastic Calculus*, Springer, New York.

Kwakernaak, H. (1974). Filtering for systems excited by Poisson white noise, *Lecture Notes in Econom. and Math. Systems*, **107**, 468–492.

Lamm, P. (1992). Parameter estimation for distributed equations in parameter-dependent state spaces: Application to shape identification, *SIAM J. Control Optim.*, **30**, 894–925.

Lindsay, B. G. (1988). Composite likelihood methods, *Contemp. Math.*, **80**, 221–239.

Matérn, B. (1986). *Spatial Variation*, 2nd ed., Lecture Notes in Statist., No. 36, Springer, Berlin.

Mohapl, J. (1994). Maximum likelihood estimation in linear infinite dimensional models, *Comm. Statist. Stochastic Models*, **10**, 781–794.

Mohapl, J. (1998). Discrete sample estimation for Gaussian random fields generated by stochastic partial differential equations, *Comm. Statist. Stochastic Models*, **14**, 883–903.

Mohapl, J. (1999). On estimation in random fields generated by linear stochastic partial differential equations, *Mathematica Slovaca*, **49**, 95–115.

Omatu, S. (1984). Estimation theory in Hilbert spaces and its applications, *Advances in Probability and Related Topics*, Vol. 7 (ed. M. Pinsky), 367–410, Marcel Dekker, Inc., New York.

Piterbarg, L. I. and Ostrovskii, G. A. (1997). *Advection and Diffusion in Random Media*, Implications for Sea Surface Temperature Anomalies, Kluwer, Dordrecht.

Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T. (1986). *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, 1st ed., Cambridge University Press.

Sudicky, E. A., Cherry, J. A. and Frind, E. O. (1983). Migration of contaminants in groundwater at a landfill, A case study, *Journal of Hydrology*, **63**, 81–108.

Unny, T. E. (1988). Stochastic partial differential equations in groundwater hydrology, *Stochastic Structural Dynamics, Proceedings of the Symposium held at the University of Illinois at Urbana Champaign October 30–November 1, 1988* (eds. N. S. Namachchivaya, H. H. Hilton and Y. K. Wen).

Vomvoris, E. G. and Gelhar, L. W. (1986). Stochastic prediction of dispersive contaminant transport, Tech. Report, No. EPA/600/2-86/114 of the Civil Engineering Department, Massachusetts Institute of Technology, Cambridge (reproduced by the U. S. Department of Commerce, National Technical Information Service).

Walsh, J. B. (1986). An introduction to stochastic partial differential equations, *Lecture Notes in Math.*, **1180**, 266–437.

Warnes, J. J. and Ripley, B. D. (1987). Problems with likelihood estimation of covariance functions of spatial Gaussian processes, *Biometrika*, **74**, 640–642.

Whittle, P. (1954). On stationary processes in the plane, *Biometrika*, **41**, 434–439.

Whittle, P. (1962). Topographic correlation, power law covariance functions and diffusion, *Biometrika*, **49**, 305–314.

Zhang, Y. (1995). Autoregressive models for continuous stationary space-time processes, Ph.D. Thesis, Department of Preventive Medicine and Biometrics, Graduate School of the University of Colorado.