Proofs subject to correction. Not to be reproduced without permission. Contributions to the discussion must not exceed 400 words. Contributions longer than 400 words will be cut by the editor.

Appl. Statist. (2010)
59, *Part* 2, *pp.* 1–20

Geostatistical inference under preferential sampling

Peter J. Diggle,

Lancaster University, UK, and Johns Hopkins University School of Public Health, Baltimore, USA

Raquel Menezes

University of Minho, Braga, Portugal

and Ting-li Su

Lancaster University, UK

[*Read before* The Royal Statistical Society *at a meeting organized by the* Environmental Statistics Section *on Wednesday, September 23rd, 2009, the President* Professor D. J. Hand, *in the Chair*]

Summary. Geostatistics involves the fitting of spatially continuous models to spatially discrete data. Preferential sampling arises when the process that determines the data locations and the process being modelled are stochastically dependent. Conventional geostatistical methods assume, if only implicitly, that sampling is non-preferential. However, these methods are often used in situations where sampling is likely to be preferential. For example, in mineral exploration, samples may be concentrated in areas that are thought likely to yield high grade ore. We give a general expression for the likelihood function of preferentially sampled geostatistical data and describe how this can be evaluated approximately by using Monte Carlo methods. We present a model for preferential sampling and demonstrate through simulated examples that ignoring preferential sampling can lead to misleading inferences. We describe an application of the model to a set of biomonitoring data from Galicia, northern Spain, in which making allowance for preferential sampling materially changes the results of the analysis.

Keywords: Environmental monitoring; Geostatistics; Log-Gaussian Cox process; Marked point process; Monte Carlo inference; Preferential sampling

1. Introduction

The term *geostatistics* describes the branch of spatial statistics in which data are obtained by sampling a spatially continuous phenomenon $S(x): x \in \mathbb{R}^2$ at a discrete set of locations $x_i, i=1,\ldots,n$, in a spatial region of interest $A \subset \mathbb{R}^2$. In many cases, $S(x)$ cannot be measured without error. Measurement errors in geostatistical data are typically assumed to be additive, possibly on a transformed scale. Hence, if Y_i denotes the measured value at the location x_i, a simple model for the data takes the form

$$Y_i = \mu + S(x_i) + Z_i, \qquad i=1,\ldots,n, \tag{1}$$

where the Z_i are mutually independent, zero-mean random variables with variance τ^2, often in this context called the *nugget variance*. One interpretation of the Z_i in model (1) is as measure-

---

Address for correspondence: Peter J. Diggle, School of Health and Medicine, Lancaster University, Bailrigg, Lancaster, LA1 4YB, UK
E-mail: p.diggle@lancaster.ac.uk

© 2009 Royal Statistical Society 0035–9254/10/59000

RSSC c2081 Dispatch: 16.7.2009 No. of pages:20

ment errors in the Y_i. Another, which explains the more colourful terminology, is as a device to model spatial variation on a scale that is smaller than the shortest distance between any two sample locations x_i. We adopt the convention that $E[S(x)]=0$ for all x; hence in model (1) $E[Y_i]=\mu$ for all i. Model (1) extends easily to the regression setting, in which $E[Y_i]=\mu_i=d_i'\beta$, with d_i a vector of explanatory variables associated with Y_i. The objectives of a geostatistical analysis typically focus on prediction of properties of the realization of $S(x)$ throughout the region of interest A. Targets for prediction might include, according to context, the value of $S(x)$ at an unsampled location, the spatial average of $S(x)$ over A or subsets thereof, the minimum or maximum value of $S(x)$ or subregions in which $S(x)$ exceeds a particular threshold. Chilès and Delfiner (1999) have given a comprehensive account of classical geostatistical models and methods.

Diggle *et al.* (1998) introduced the term *model-based geostatistics* to mean the application of general principles of statistical modelling and inference to geostatistical problems. In particular, they added Gaussian distributional assumptions to the classical model (1) and re-expressed it as a two-level hierarchical linear model, in which $S(x)$ is the value at location x of a latent Gaussian stochastic process and, conditional on $S(x_i), i=1,\ldots,n$, the measured values $Y_i, i=1,\ldots,n$, are mutually independent, normally distributed with means $\mu+S(x_i)$ and common variance τ^2. Diggle *et al.* (1998) then extended this model, retaining the Gaussian assumption for $S(x)$ but allowing a generalized linear model (McCullagh and Nelder, 1989) for the mutually independent conditional distributions of the Y_i given $S(x_i)$.

As a convenient shorthand notation to describe the hierarchical structure of a geostatistical model, we use $[\cdot]$ to mean 'the distribution of' and write $S=\{S(x): x\in\mathbb{R}^2\}$, $X=(x_1,\ldots,x_n)$, $S(X)=\{S(x_1),\ldots,S(x_n)\}$ and $Y=(Y_1,\ldots,Y_n)$. Then, the model of Diggle *et al.* (1998) implicitly treats X as being deterministic and has the structure $[S,Y]=[S][Y|S(X)]=[S][Y_1|S(x_1)][Y_2|S(x_2)]\ldots[Y_n|S(x_n)]$. Furthermore, in model (1) the $[Y_i|S(x_i)]$ are univariate Gaussian distributions with means $\mu+S(x_i)$ and common variance τ^2.

As presented above, and in almost all of the geostatistical literature, models for the data treat the sampling locations x_i either as fixed by design or otherwise stochastically independent of the process $S(x)$. Admitting the possibility that the sampling design may be stochastic, a complete model needs to specify the joint distribution of S, X and Y. Under the assumption that X is independent of S we can write the required joint distribution as $[S,X,Y]=[S][X][Y|S(X)]$, from which it is clear that for inferences about S or Y we can legitimately condition on X and use standard geostatistical methods. We refer to this as *non-preferential sampling* of geostatistical data. Conversely, *preferential sampling* refers to any situation in which $[S,X]\neq[S][X]$.

We contrast the term *non-preferential* with the term *uniform*, the latter meaning that, beforehand, all locations in A are equally likely to be sampled. Examples of designs which are both uniform and non-preferential include completely random designs and regular lattice designs (strictly, in the latter case, if the lattice origin is chosen at random). An example of a non-uniform, non-preferential design would be one in which sample locations are an independent random sample from a prescribed non-uniform distribution on A. Preferential designs can arise either because sampling locations are deliberately concentrated in subregions of A where the underlying values of $S(x)$ are thought likely to be larger (or smaller) than average, or more generally when X and Y together form a marked point process in which there is dependence between the points X and the marks Y.

We emphasize at this point that our definition of preferential sampling involves a stochastic dependence, as opposed to a functional dependence, between the process S and the sampling design X. For example, a model in which the mean of S and the intensity of X share a dependence on a common set of explanatory variables does not constitute preferential sampling. In most geostatistical applications it is difficult to maintain a sharp distinction between the treatment of variation $S(x)$ as deterministic or stochastic because of the absence of independent replication of the process under investigation. Our pragmatic stance is to represent by a stochastic model the portion of the total variation in S that cannot be captured by extant explanatory variables.

Curriero *et al.* (2002) evaluated a class of non-ergodic estimators for the covariance structure of geostatistical data, which had been proposed by Isaaks and Srivastava (1988) and Srivastava and Parker (1989) as a way of dealing with preferential sampling. They concluded that the non-ergodic estimators 'possess no clear advantage' over the traditional estimators that we describe in Section 3.1. Schlather *et al.* (2004) developed two tests for preferential sampling, which treat a set of geostatistical data as a realization of a marked point process. Their null hypothesis is that the data are a realization of a *random-field model*. This model assumes that the sample locations X are a realization of a point process $\mathcal{P}$ on A, that the mark of a point at location x is the value at x of the realization of a random field S on A, and that $\mathcal{P}$ and S are independent processes. This is therefore equivalent to our notion of non-preferential sampling. Their test statistics are based on the following idea. Assume that S is stationary, and let $M_k(h) = E[S(x)^k | x, x+h \in \mathcal{P}]$. Under the null hypothesis that sampling is non-preferential, the conditioning on $x+h \in \mathcal{P}$ is irrelevant; hence $M_k(h)$ is a constant. Schlather *et al.* (2004) proposed as test statistics the empirical counterparts of $M_1(h)$ and $M_2(h)$, and implemented the resulting tests by comparing the observed value of each chosen test statistic with values calculated from simulations of a conventional geostatistical model fitted to the data on the assumption that sampling is non-preferential. Guan and Afshartous (2007) avoided the need for simulation and parametric model fitting by dividing the observation into non-overlapping subregions that can be assumed to provide approximately independent replicates of the test statistics. In practice, this requires a large data set; their application has a sample size $n = 4358$.

In this paper, we propose a class of stochastic models and associated methods of likelihood-based inference for preferentially sampled geostatistical data. In Section 2 we define our model for preferential sampling. In Section 3 we use the model to illustrate the potential for inferences to be misleading when conventional geostatistical methods are applied to preferentially sampled data. Section 4 discusses likelihood-based inference using Monte Carlo methods and suggests a simple diagnostic for the fitted model. Section 5 applies our model and methods to a set of biomonitoring data from Galicia, northern Spain, in which the data derive from two surveys of the same region, one of which is preferentially sampled, the other not. Section 6 is a concluding discussion.

The data that are analysed in the paper can be obtained from

`http://www.rss.org.uk/main.asp?page=1836`

2. A shared latent process model for preferential sampling

Recall that S denotes an unobserved, spatially continuous process on a spatial region A, X denotes a point process on A and Y denotes a set of measured values, one at each point of X. The focus of scientific interest is on properties of S, as revealed by the data (X, Y), rather than on the joint properties of S and X, but we wish to protect against incorrect inferences that might arise because of stochastic dependence between S and X.

To clarify the distinction between preferential and non-preferential sampling, and the inferential consequences of the former, we first examine a related situation that was considered by Rathbun (1996), in which S and X are stochastically dependent but measurements Y are taken only at a different, prespecified set of locations, i.e. independently of X. Then, the joint distribution of S, X and Y takes the form

$$[S, X, Y] = [S][X|S][Y|S]. \tag{2}$$

It follows immediately on integrating equation (2) with respect to X that the joint distribution of S and Y has the standard form, $[S, Y] = [S][Y|S]$. Hence, for inference about S it is valid, if potentially inefficient, to ignore X, i.e. to use conventional geostatistical methods. Models that are analogous to equation (2) have also been proposed in a longitudinal setting, where the analogues of Y and X are a time sequence of repeated measurements at prespecified times and a related time-to-event outcome respectively. See, for example, Wulfsohn and Tsiatis (1997) or Henderson *et al.* (2000).

In contrast, if Y is observed at the points of X, the appropriate factorization is

$$[S, X, Y] = [S][X|S][Y|X, S]. \tag{3}$$

Even when, as is typical in geostatistical modelling, equation (3) takes the form

$$[S, X, Y] = [S][X|S][Y|S(X)], \tag{4}$$

so that the algebraic form of $[Y|X, S]$ reduces to $[Y|S(X)]$, an important distinction between preferential and non-preferential sampling is that in equation (4) the functional dependence between S and X in the term $[Y|S(X)]$ cannot be ignored, because the implicit specification of $[S, Y]$ resulting from equation (4) is non-standard. Conventional geostatistical inferences which ignore the stochastic nature of X are therefore potentially misleading. The longitudinal analogue of equation (4) arises when subjects in a longitudinal study provide measurements at time points which are not prespecified as part of the study design; see, for example, Lipsitz *et al.* (2002), Lin *et al.* (2004) or Ryu *et al.* (2007).

We now define a specific class of models through the following additional assumptions.

Assumption 1. S is a stationary Gaussian process with mean 0, variance σ^2 and correlation function $\rho(u;\phi) = \mathrm{corr}\{S(x), S(x')\}$ for any x and x' a distance u apart.

Assumption 2. Conditional on S, X is an inhomogeneous Poisson process with intensity

$$\lambda(x) = \exp\{\alpha + \beta\, S(x)\}. \tag{5}$$

Assumption 3. Conditional on S and X, Y is a set of mutually independent Gaussian variates with $Y_i \sim N\{\mu + S(x_i), \tau^2\}$.

It follows from assumptions 1 and 2 that, unconditionally, X is a log-Gaussian Cox process (Møller *et al.* 1998). If $\beta = 0$ in equation (5), then it follows from assumptions 1 and 3 that the unconditional distribution of Y is multivariate Gaussian with mean $\mu\mathbf{1}$ and variance matrix $\tau^2 I + \sigma^2 R$, where I is the identity matrix and R has elements $r_{ij} = \rho(||x_i - x_j||;\phi)$.

Ho and Stoyan (2008) discussed essentially the same construction as assumptions 1–3 viewed as a model for a marked point process of locations X and marks Y, and derived its first- and second-moment properties.

We do not suggest that this model will be adequate for all applications. However, it is sufficiently rich to provide a vehicle for investigating the consequences of preferential sampling, and for the application that is described in Section 5 of the paper.

3. Effect of preferential sampling on geostatistical inference

To illustrate how preferential sampling affects the performance of standard geostatistical methods, we have conducted a simulation experiment as follows. For each run of the experiment, we first simulated an approximate realization of a stationary Gaussian process on the unit square by simulating its values on a finely spaced lattice and treating the spatially continuous process $S(\cdot)$ as constant within each lattice cell. We then sampled the values of $S(\cdot)$, with additive Gaussian measurement error, either non-preferentially or preferentially according to each of the following sampling designs. For the *completely random* sampling design, sample locations x_i were an independent random sample from the uniform distribution on A. For the *preferential* design, we first generated a realization of S, then a realization of $X = \{x_i : i = 1, \ldots, n\}$ conditional on S by using the model that is defined by equation (5) with parameter $\beta = 2$, and finally a realization of $Y = \{y_i : i = 1, \ldots, n\}$ conditional on S and on X by using the conditional model that is defined by assumption 3 above. For the *clustered* design, we used the same procedure to generate a realization of X, but then generated a realization of Y on the locations X by using a second, independent, realization of S, so that the resulting Y is a realization of the standard geostatistical model (1). This gives a non-preferential design with the same marginal properties for X and Y as the preferential design.

The model for the spatial process S was stationary Gaussian, with mean $\mu = 4$, variance $\sigma^2 = 1.5$ and Matérn correlation with scale parameter $\phi = 0.15$ and shape parameter $\kappa = 1$. In each case, we set the nugget variance, $\tau^2 = 0$; hence the data y_i consisted of the realized values of $S(\cdot)$ as the sample locations x_i.

The Matérn (1986) class of correlation functions takes the form

$$\rho(u; \phi, \kappa) = \{2^{\kappa-1}\,\Gamma(\kappa)\}^{-1} (u/\phi)^{\kappa}\, K_{\kappa}(u/\phi), \qquad u > 0,$$

where $K_{\kappa}(\cdot)$ denotes the modified Bessel function of the second kind, of order $\kappa > 0$. This class is widely used because of its flexibility. Although κ is difficult to estimate without extensive data, the integral part of κ determines the degree of mean-square differentiability of the corresponding process $S(\cdot)$, giving both a nice interpretation and, in at least some contexts, a rationale for choosing a particular value for κ. The special case $\kappa = 0.5$ gives an exponential correlation function, $\rho(u; \phi) = \exp(-u/\phi)$.

Fig. 1 shows a realization of each of the three sampling designs superimposed on a single realization of the process S. The preferential nature of the sampling in Fig. 1(b) results in the sample locations falling predominantly within the darker shaded areas.

3.1. *Variogram estimation*

The theoretical variogram of a stationary spatial process $Y(x)$ is the function $V(u) = \frac{1}{2}\,\mathrm{var}\{Y(x) - Y(x')\}$ where u denotes the distance between x and x'. Non-parametric estimates of $V(u)$ are widely used in geostatistical work, both for exploratory data analysis and for diagnostic checking.

Consider a set of data $(x_i, y_i), i = 1, \ldots, n$, where x_i denotes a location and y_i a corresponding measured value. The *empirical variogram ordinates* are the quantities $v_{ij} = (y_i - y_j)^2/2$. Under non-preferential sampling, each v_{ij} is an unbiased estimator for $V(u_{ij})$, where u_{ij} is the distance between x_i and x_j. A scatterplot of v_{ij} against u_{ij} is called the *variogram cloud*. A smoothed version of the variogram cloud can be used to suggest appropriate parametric models for the spatial covariance structure of the data; in what follows, we use simple binned estimators. For more information on variogram estimation, see for example Cressie (1985) Cressie (1991), section 2.4, Chilès and Delfiner (1999), section 2.2, or Diggle and Ribeiro (2007), chapter 5.

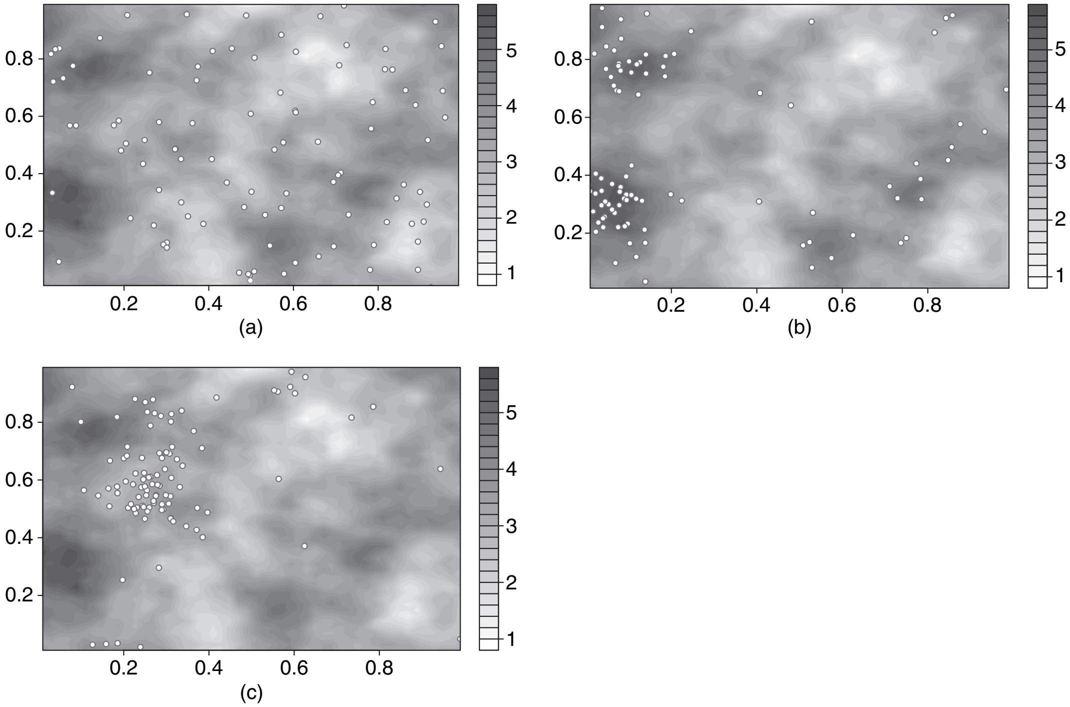
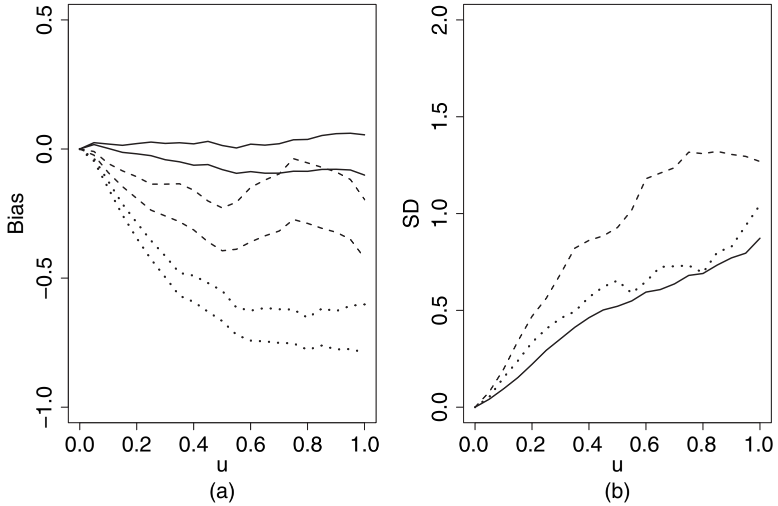


Fig. 1. Sample locations and underlying realizations of the signal process for the model that was used in the simulation study (in each case, the background image represents the realization of the signal process $S(x)$ that was used to generate the associated measurement data; the model parameter values are $\mu=4$, $\sigma^2=1.5$, $\phi=0.15$, $\kappa=1$, $\beta=2$ and $\tau^2=0$): (a) completely random sample; (b) preferential sample; (c) clustered sample

Fig. 2 shows simulation-based estimates of the pointwise bias and standard deviation of smoothed empirical variograms, derived from 500 replicate simulations of each of our three sampling designs. With regard to bias, the results under both uniform and clustered non-preferential sampling designs are consistent with the unbiasedness of the empirical variogram ordinates; although smoothing the empirical variogram ordinates does induce some bias, this effect is small in the current setting. In contrast, under preferential sampling the results show severe bias. With regard to efficiency, Fig. 2(b) illustrates that clustered sampling designs, whether preferential or not, are also less efficient than uniform sampling. The bias that is induced by preferential sampling is qualitatively unsurprising; the effect of the preferential sampling is that sample locations predominantly span a reduced range of values of $S(x)$, which in turn reduces the expectation of pairwise squared differences at any given spatial separation. Note, incidentally, that the sample variogram has substantially smaller variance under preferential than under non-preferential clustered sampling. However, this is of little practical interest in view of its severe bias under preferential sampling. In general, the implicit estimand of the empirical variogram is the variance of $Y(x)-Y(x')$ conditional on both x and x' belonging to X, which under preferential sampling differs from the unconditional variance; see, for example, Wälder and Stoyan (1996) or Schlather (2001).

3.2. *Spatial prediction*

Suppose that our target for prediction is $S(x_0)$, which is the value of the process S at a generic



(a) (b)

Fig. 2. Estimated bias and standard deviation of the sample variogram under random (———), preferential (· · · · · ·) and clustered (- - - - - - -) sampling (see the text for a detailed description of the simulation model): (a) pointwise means plus and minus two pointwise standard errors; (b) pointwise standard deviations

location x_0, given sample data $(x_i, y_i), i=1,2,\ldots,n$. The widely used ordinary kriging predictor estimates the unconditional expectation of $S(x_0)$ by generalized least squares, but using plug-in estimates of the parameters that define the covariance structure of Y. In classical geostatistics, these plug-in estimates are obtained either subjectively (Chilès and Delfiner (1999), section 2.6) or by non-linear least squares (Cressie (1991), section 2.6.2). We used maximum likelihood estimation under the assumed Gaussian model for Y.

Each simulation yields an estimate of the bias $\hat{S}(x_0) - S(x_0)$ and the mean-square error $\{\hat{S}(x_0) - S(x_0)\}^2$, for the ordinary kriging predictor $\hat{S}(x_0)$. The first two rows of Table 1 show approximate 95% confidence intervals, calculated as means plus and minus 2 standard errors over 500 replicate simulations, for the bias and root-mean-square error at the prediction location $x_0=(0.49, 0.49)$.

The bias is large and positive under preferential sampling with $\beta>0$. This prediction bias is a direct consequence of the bias in the estimation of the model parameters, which in turn arises because the preferential sampling model leads to the oversampling of locations corresponding to high values of the underlying process S. The correct predictive distribution for S is $[S|Y, X]$ which, with known parameter values, takes a standard multivariate Gaussian form whether or not sampling is preferential. The two non-preferential sampling designs both lead to approximately unbiased prediction, as predicted by theory. The substantially larger mean-square error for clustered sampling by comparison with completely random sampling reflects the inefficiency of the latter, as already illustrated in the context of variogram estimation.

In a second set of simulations, we set the values of the model parameters to correspond to the maximum likelihood estimates that were obtained in the analysis of the 1997 Galicia biomonitoring data reported in Section 5 below; hence $\mu=1.515, \sigma^2=0.138, \phi=0.313, \kappa=0.5, \beta=-2.198$ and $\tau^2=0.059$. The results are qualitatively as expected, but the differences among the three sampling designs are much smaller for two reasons. Firstly, the degree of preferentiality is much weaker; a measure of this is the product $\beta\sigma$, which takes the values 3 and 0.815 for the first and second simulation models respectively. Secondly, the effect is further diluted by the inclusion of a non-zero nugget variance.

Table 1. Effect of sampling design on the bias and mean-square error of the ordinary kriging *predictor* $\hat{S}(x_0)$, when $x_0 = (0.49, 0.49)$ and each sample consists of 100 locations on the unit square†

Model	*Parameter*	*Confidence intervals for the following sampling designs:*		
		Completely random	*Preferential*	*Clustered*
1	Bias	$(-0.014, 0.055)$	$(0.951, 1.145)$	$(-0.048, 0.102)$
1	Root-mean-square error	$(0.345, 0.422)$	$(1.387, 1.618)$	$(0.758, 0.915)$
2	Bias	$(0.003, 0.042)$	$(-0.134, -0.090)$	$(-0.018, 0.023)$
2	Root-mean-square error	$(0.202, 0.228)$	$(0.247, 0.292)$	$(0.214, 0.247)$

†Each entry is an approximate 95% confidence interval calculated from 500 independent simulations. See the text for a detailed description of the simulation models 1 and 2.

4. Fitting the shared latent process model

4.1. Monte Carlo maximum likelihood estimation

For the shared latent process model (3), the likelihood function for data X and Y can be expressed as

$$L(\theta) = [X, Y] = E_S[[X|S][Y|X, S]], \tag{6}$$

where θ represents all the model parameters and the expectation is with respect to the unconditional distribution of S. Evaluation of the conditional distribution $[X|S]$ strictly requires the realization of S to be available at all $x \in A$. However, and as previously noted in Section 3.1, we approximate the spatially continuous realization of S by the set of values of S on a finely spaced lattice to cover A and replace the exact locations X by their closest lattice points. We then partition S into $S = \{S_0, S_1\}$, where S_0 denotes the values of S at each of n data locations $x_i \in X$, and S_1 denotes the values of S at the remaining $N - n$ lattice points.

To evaluate $L(\theta)$ approximately, a naive strategy would be to replace the intractable expectation on the right-hand side of equation (6) by a sample average over simulations S_j. This strategy fails when the measurement error variance τ^2 is 0, because unconditional simulations of S will then be incompatible with the observed Y. It also fails in practice when the measurement error is small relative to the variance of S, which is the case of most practical interest.

We therefore rewrite the exact likelihood (6) as the integral

$$L(\theta) = \int [X|S][Y|X, S] \frac{[S|Y]}{[S|Y]} [S] \, \mathrm{d}S. \tag{7}$$

Now, write $[S] = [S_0][S_1|S_0]$ and replace the term $[S|Y]$ in the denominator of expression (7) by $[S_0|Y][S_1|S_0, Y] = [S_0|Y][S_1|S_0]$. Note also that $[Y|X, S] = [Y|S_0]$. Then, equation (7) becomes

$$\begin{aligned} L(\theta) &= \int [X|S] \frac{[Y|S_0]}{[S_0|Y]} [S_0][S|Y] \, \mathrm{d}S \\ &= E_{S|Y}\left[[X|S] \frac{[Y|S_0]}{[S_0|Y]} [S_0] \right] \end{aligned} \tag{8}$$

and a Monte Carlo approximation is

$$L_{\mathrm{MC}}(\theta) = m^{-1} \sum_{j=1}^{m} [X|S_j] \frac{[Y|S_{0j}]}{[S_{0j}|Y]} [S_{0j}], \tag{9}$$

where now the S_j are simulations of S conditional on Y. Note that, when Y is measured without error, $[Y|S_{0j}]/[S_{0j}|Y] = 1$. To reduce the Monte Carlo variance, we also use antithetic pairs of realizations, i.e. for each $j = 1, \ldots, m/2$ set $S_{2j} = 2\mu_{\mathrm{c}} - S_{2j-1}$, where μ_{c} denotes the conditional mean of S given Y. The use of conditional simulation in equation (9) bypasses the difficulty with the naive strategy by guaranteeing that the simulated realizations of S are compatible with the data Y.

To simulate a realization from $[S|Y]$, we use the following construction. Recall that the data locations $X = \{x_1, \ldots, x_n\}$ constitute a subset of the $N \geqslant n$ prediction locations, $X^* = \{x_1^*, \ldots, x_N^*\}$ say. Define C to be the $n \times N$ matrix whose ith row consists of $N-1$ 0s and a single 1 to identify the position of x_i within X^*. Note that, unconditionally, $S \sim \mathrm{MVN}(0, \Sigma)$ and $Y \sim \mathrm{MVN}(\mu, \Sigma_0)$ with $\Sigma_0 = C \Sigma C' + \tau^2 I$. Then, if Z denotes an independent random sample of size n from $N(0, \tau^2)$ and y denotes the observed value of Y, it follows that

$$S_c = S + \Sigma C' \Sigma_0^{-1} (y - \mu + Z - CS) \tag{10}$$

has the required multivariate Gaussian distribution of S given $Y = y$ (Rue and Held (2005), chapter 2, and Eidsvik *et al.* (2006)). Hence, for conditional simulation when N is large, we need a fast algorithm for unconditional simulation of S. We have used the circulant embedding algorithm of Wood and Chan (1994) applied to a rectangular region containing the region of interest, A. The subsequent calculations for S_c then involve only the relatively straightforward inversion of the $n \times n$ matrix Σ_0 and simulation of the n independent Gaussian random variables that make up the vector Z in equation (10).

4.2. Goodness of fit

We have already noted in Section 1 the availability of tests for preferential sampling; see, for example, Schlather *et al.* (2004) or Guan and Afshartous (2007). Here, we suggest a way of assessing the goodness of fit of the preferential sampling model that was described in Section 2, by comparing the sample locations with realizations of the fitted Cox model for their unconditional distribution.

A standard diagnostic tool for stationary spatial point processes is the reduced second-moment measure, or K-function (Ripley, 1977), that is defined by $\lambda K(s) = E[N_0(s)]$ where $N_0(s)$ denotes the number of points of the process within distance s of an arbitrary origin of measurement, conditional on there being a point of the process at the origin, and λ is the expected number of points of the process per unit area. Under the preferential sampling model, the marginal model for the sample locations X is a log-Gaussian Cox process with stochastic intensity $\Lambda(x) = \exp\{\alpha + \beta S(x)\}$. For this process, the K-function is of the form

$$K(s) = \pi s^2 + 2\pi \int_0^s \gamma(u) u \, \mathrm{d}u, \tag{11}$$

where $\gamma(u) = \exp\{\beta^2 \sigma^2 \rho(u; \phi)\} - 1$ is the covariance function of $\Lambda(x)$ (Diggle (2003), section 5.5) To assess the goodness of fit informally, we compare the estimated K-function of the data with the envelope of estimates that is obtained from sets of sample locations generated from simulated realizations of the fitted model. For a formal Monte Carlo test, we use a goodness of fit statistic that measures the discrepancy between estimated and theoretical K-functions, as described in Section 5.2.2. Note that this aspect of the model is not considered explicitly in the fitting process that was described in Section 4.1.

5. Heavy metal biomonitoring in Galicia

Our application concerns biomonitoring of lead pollution in Galicia, northern Spain, by using the concentrations in moss samples, in micrograms per gram dry weight, as the measured variable. An initial survey was conducted in the spring of 1995 ‘to select the most suitable moss species and collection sites’ (Fernández *et al.*, 2000). Two further surveys of lead concentrations in samples of the moss species (*Scleropodium purum*) took place in October 1997 and July 2000. Fig. 3 shows the sampling locations that were used in these two surveys. Note that some locations appear to lie outside Galicia, and others in the sea to the north. However, this is an artefact of the fact that the boundary is both approximate and imperfectly registered; it plays no role in the analysis and is included only to add context to the map.

In the 1997 survey, sampling was conducted more intensively in subregions where large gradients in lead concentrations were expected, in line with suggestions in Ruhling (1994). The resulting design was highly non-uniform and potentially preferential. The second survey used an approximately regular lattice design, which is therefore non-preferential; gaps in the lattice arose only where a different species of moss was collected. For further details, see Fernández *et al.* (2000) and Aboal *et al.* (2006). In particular, Fernández *et al.* (2000) studied the changes in heavy metal concentrations between the two years ignoring the corresponding spatial distributions. Our objective in analysing these data is to estimate, and compare, maps of lead concentrations in 1997 and 2000.

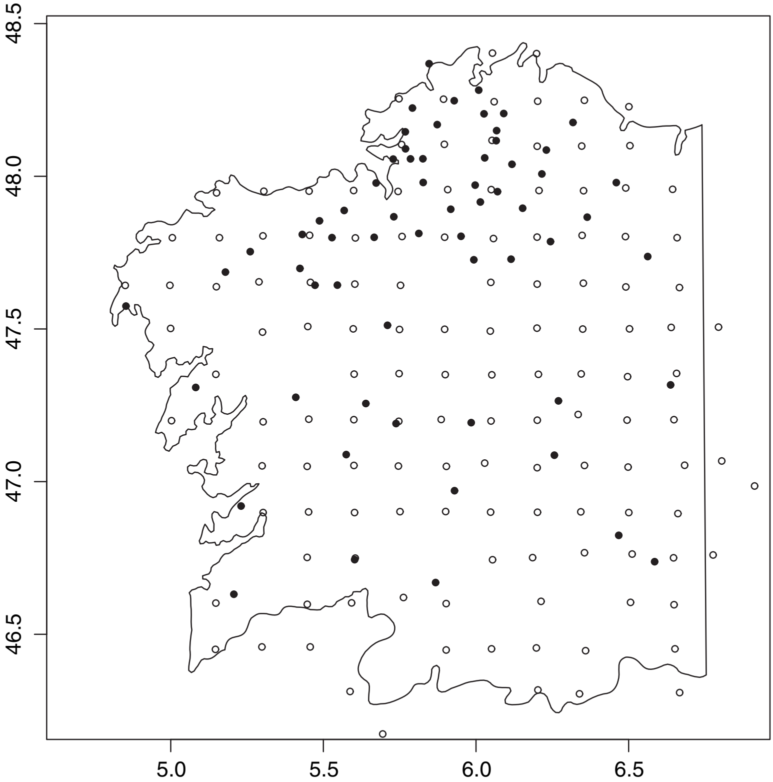


Fig. 3. Sampling locations for 1997 (●) and 2000 (○): the unit of distance is 100 km; two outliers in the 1997 data were at locations (6.50, 46.90) and (6.65, 46.75)

The measured lead concentrations included two gross outliers in 2000, each of which we replaced by the average of the remaining values from that year's survey. Table 2 gives summary statistics for the resulting 1997 and 2000 data. Note that the mean response is higher for the 1997 data than for the 2000 data, which would be consistent either with the former being preferentially sampled near potential sources of pollutant, or with an overall reduction in levels of pollution over the 3 years between the two surveys. Also, the log-transformation eliminates an apparent variance–mean relationship in the data and leads to more symmetric distributions of measured values (Fig. 4).

5.1. Standard geostatistical analysis

For an initial analysis, we assumed the standard Gaussian model (1) with the underlying signal

Table 2. Summary statistics for lead pollution levels measured in 1997 and 2000

	Levels (μg (g dry weight)$^{-1}$) for the following scales and years:			
	Untransformed		*Log-transformed*	
	1997	*2000*	*1997*	*2000*
Number of locations	63	132	63	132
Mean	4.72	2.15	1.44	0.66
Standard deviation	2.21	1.18	0.48	0.43
Minimum	1.67	0.80	0.52	−0.22
Maximum	9.51	8.70	2.25	2.16

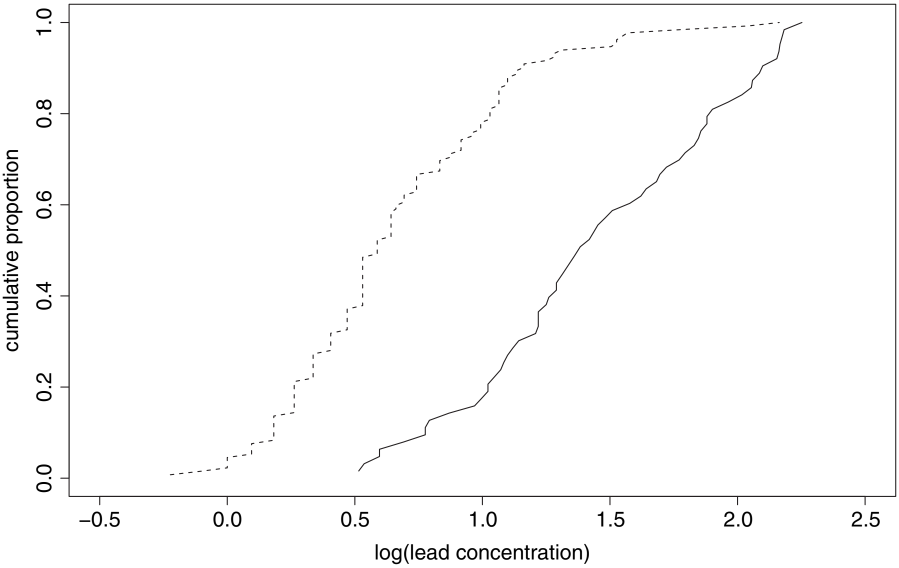


Fig. 4. Empirical distributions of log-transformed lead concentrations in the 1997 (———) and 2000 (-------) samples

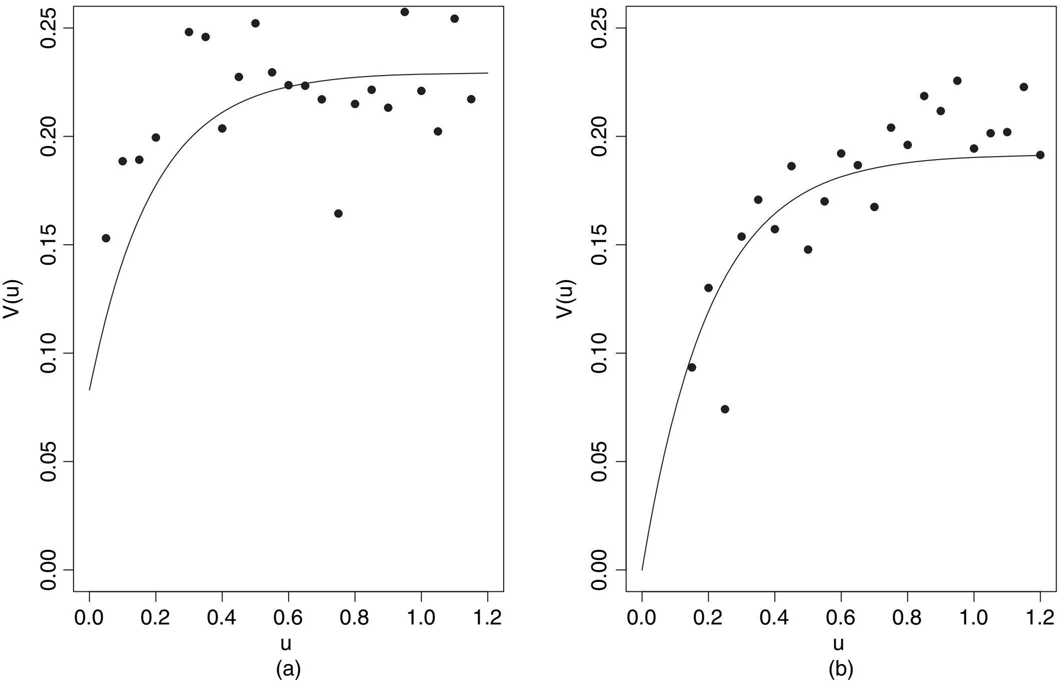
$S(x)$ specified as a zero-mean stationary Gaussian process with variance σ^2 and Matérn correlation function $\rho(u;\phi,\kappa)$, and Gaussian measurement errors, $Z_i \sim N(0,\tau^2)$, and fitted this model separately to the 1997 and 2000 data.

Fig. 5 shows, for each of 1997 and 2000, smoothed empirical variograms and theoretical variograms with parameters fitted by maximum likelihood. On the basis of the general shape of the two empirical variograms, we used a fixed value $\kappa=0.5$ for the shape parameter of the Matérn correlation function. The estimated variograms differ in some respects, notably the absence of a nugget component (i.e. $\hat{\tau}^2 \approx 0$) in the variogram that was estimated from the 2000 data; however, this parameter is poorly identified because of the lattice-like arrangement of the 2000 sampling design, whereas the inclusion of close pairs of locations in the 1997 sampling design enables better estimation of τ^2. Other features of the two fitted variograms are similar, e.g. the height of the asymptote (i.e. $\hat{\tau}^2+\hat{\sigma}^2$) and the approximate range (i.e. $\hat{\phi}$). These observations support the idea that a joint model for the two data sets might allow at least some parameters in common between the two years. The generalized likelihood ratio test statistic (Cox and Hinkley (1974), section 9.3) to test the hypothesis of common σ, ϕ and τ, under the admittedly dubious assumption that neither sample is preferential, was 7.66 on 3 degrees of freedom ($p=0.054$). We revisit this question in Section 5.2.

5.2. Analysis under preferential sampling

5.2.1. Parameter estimation

We now investigate whether the 1997 sampling is indeed preferential. We used the Nelder–Mead simplex algorithm (Nelder and Mead, 1965) to estimate the model parameters, increasing the number of Monte Carlo samples m progressively to avoid finding a false maximum. With $m=$



(a) (b)

Fig. 5. Smoothed empirical (●) and fitted theoretical (———) variograms for (a) 1997 and (b) 2000 log-transformed lead concentration data

100 000, the Monte Carlo standard error in the evaluation of the log-likelihood ratio was reduced to approximately 0.3 (the actual value varies over the parameter space) and the approximate generalized likelihood ratio test statistic to test $\beta = 0$ was 27.7 on 1 degree of freedom ($p < 0.001$).

We then fitted a joint model to the two data sets, treating the 1997 and 2000 data as preferentially and non-preferentially sampled respectively. To test the hypothesis of shared values for σ, ϕ and τ, we fitted the model with and without these constraints, obtaining a generalized likelihood ratio test statistic of 6.2 on 3 degrees of freedom ($p = 0.102$). The advantage of using shared parameter values when justified is that the parameters in the joint model are then estimated more efficiently and the model is consequently better identified (Altham, 1984). This is particularly important in the geostatistical setting, where the inherent correlation structure of the data reduces their information content by comparison with independent data having the same sample size.

Table 3 shows the Monte Carlo maximum likelihood estimates together with estimated standard errors and correlations for the model with shared σ, ϕ and τ. Standard errors and correlations were evaluated by fitting a quadratic surface to Monte Carlo log-likelihoods by ordinary least squares. Parameter combinations were initially set as a 3^6 factorial design centred on the Monte Carlo maximum likelihood estimates, with parameter values chosen subjectively after examining the trajectories through the parameter space taken by the various runs of the Nelder–Mead optimization algorithm. The quadratic surface was then refitted after augmenting this design with a 2^6-factorial on a more closely spaced set of parameter values, to check the stability of the results. Each evaluation of the log-likelihood used $m = 10000$ conditional simulations. The non-negative parameters σ, ϕ and τ are estimated on a log-transformed scale, to improve the quadratic approximation to the log-likelihood surface.

Note that the expectation of $S(\cdot)$ shows a substantial fall between 1997 and 2000, and that the preferential sampling parameter estimate is negative, $\hat{\beta} = -2.198$. The latter finding is both counterintuitive, because the oversampled northern half of the region is more industrialized than the undersampled southern half, and critically dependent on our allowing the two mean parameters to differ. Otherwise, because the observed average pollution level is substantially higher in 1997 than in 2000, we would have been forced to conclude that the 1997 sampling was preferential with a positive value of β. One piece of evidence against this alternative interpretation is that, within the 1997 data, the observed pollution levels are lower in the oversampled northern half of the region ($n = 47$; mean log-concentration 1.38; standard deviation SD = 0.49) than in the undersampled southern half ($n = 16$; mean 1.62; SD = 0.40), which is consistent with a negative value of β.

Table 3. Monte Carlo maximum likelihood estimates of parameters in the joint model for the 1997 and 2000 Galicia biomonitoring data†

Parameter	*Estimate*	*Standard error*	*Correlation matrix*					
μ_{97}	1.515	0.136	1.000	0.023	0.095	−0.243	−0.222	0.167
μ_{00}	0.762	0.110		1.000	0.230	−0.229	−0.281	0.342
$\log(\sigma)$	−0.992	0.049			1.000	−0.217	−0.744	0.469
$\log(\phi)$	−1.163	0.075				1.000	0.604	−0.675
$\log(\tau)$	−1.419	0.042					1.000	−0.652
β	−2.198	0.336						1.000

†Approximate standard errors and correlations are computed from a quadratic fit to the Monte Carlo log-likelihood surface (see the text for details).

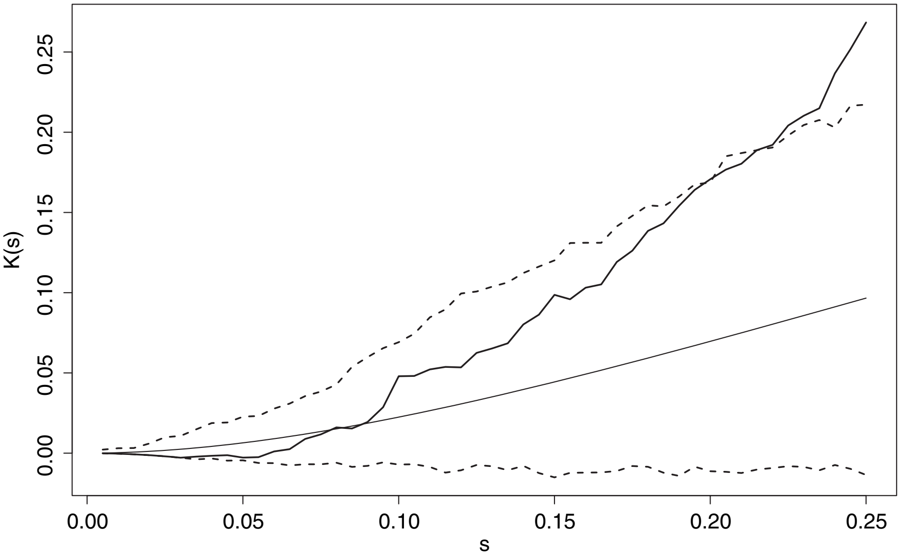


Fig. 6. Estimated K-function of the 1997 sample locations (———) and envelope from 99 simulations of the fitted log-Gaussian Cox process (-------)

5.2.2. *Goodness of fit*

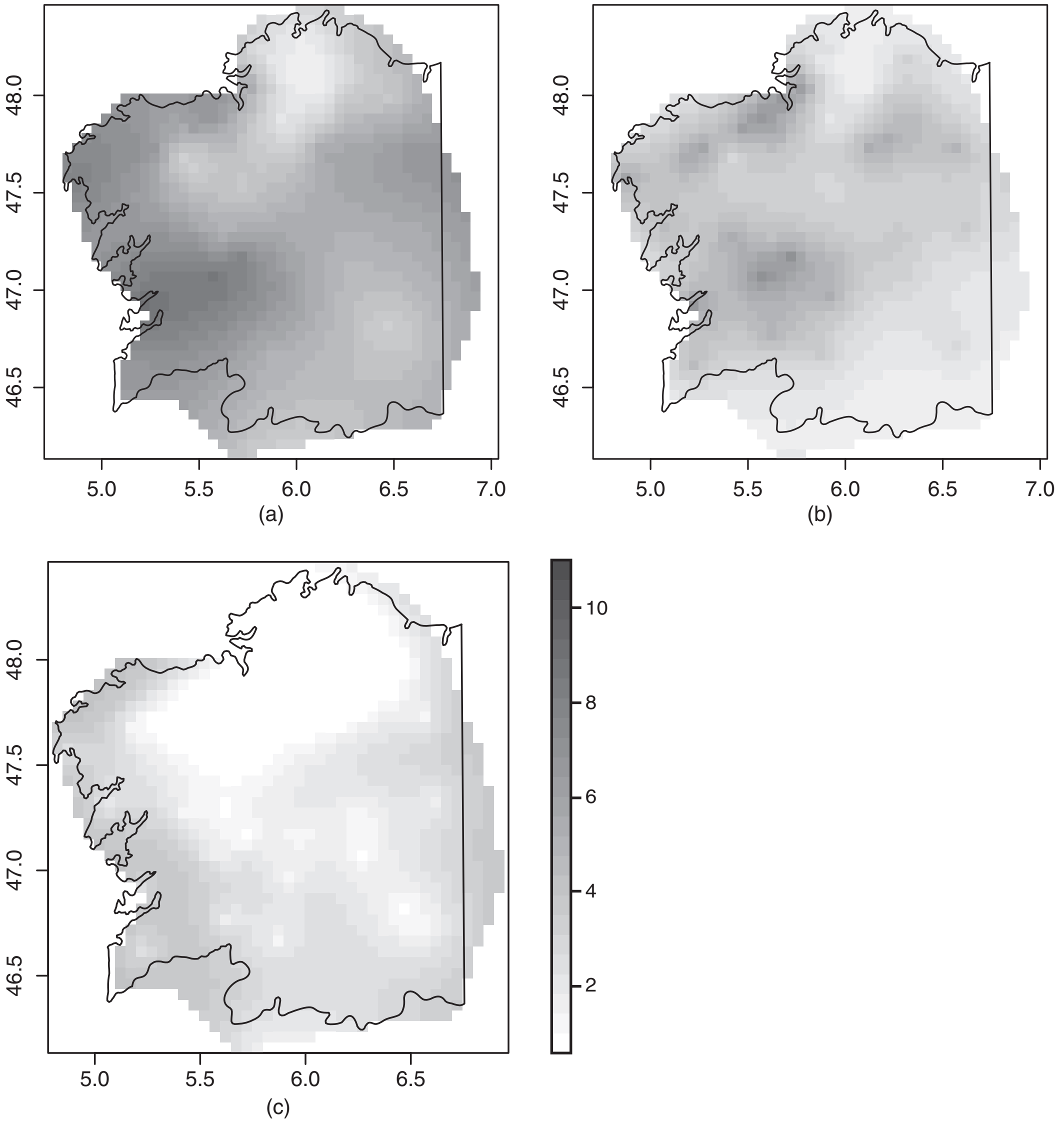
Fig. 6 shows the estimated K-function for the 1997 sampling locations together with the envelope of 99 simulations of the fitted Cox process, and the theoretical K-function. The estimate lies within the simulation envelope for distances up to 0.22 (22 km). For a formal Monte Carlo goodness-of-fit test, we define the test statistic

$$T = \int_0^{0.25} \frac{\{\hat{K}(s) - K(s)\}^2}{v(s)} \mathrm{d}s$$

where $K(s)$ is given by equation (11) and $v(s)$ is the variance of $\hat{K}(s)$, estimated from the simulations of the fitted Cox process. This gives $p = 0.03$. The Cox model slightly underestimates the extent of spatial aggregation in the data locations.

5.2.3. *Prediction*

What effect does the acknowledgement of preferential sampling make on the predicted 1997 pollution surface? Fig. 7 shows the predicted surfaces $\hat{T}(x) = E[T(x)|X, Y]$, where $T(x) = \exp\{S(x)\}$ denotes lead concentration on the untransformed scale, together with the pointwise differences between the two. Each surface is a Monte Carlo estimate based on $m = 10000$ simulations, resulting in Monte Carlo standard errors of 0.026 or less. The predictions that are based on the preferential sampling model have a substantially wider range, over the lattice of prediction locations, than those that assume non-preferential sampling (1.310–7.654 and 1.286–5.976 respectively). The difference surface also covers a relatively large range (from -0.715 to 3.693) and shows strong spatial structure. The size of the difference between the two predicted surfaces is at first sight surprising, as both are partially constrained by the observed concentrations. However, in the presence of a nugget effect the predictions are not constrained to interpolate the data. Also, ignoring the preferential nature of the sampling leads to biased parameter estimates. Finally, the effect of the back-transformation from log-concentrations to concentrations of lead, this being the scale on which predictions are required, is to amplify the differences.



(a)

(b)

(c)

Fig. 7. Predicted surface of lead concentrations in 1997 under (a) preferential and (b) non-preferential assumptions, together with (c) the pointwise difference between the two: all three surfaces are plotted on a common scale, as shown

Using the conditional expectation as a point predictor is conventional, but questionable when, as here, the measurement process has a highly skewed distribution. As an alternative summary, Fig. 8 compares pointwise 5%, 50% and 95% limits of the plug-in predictive distribution of lead concentrations under preferential and non-preferential modelling assumptions, holding the model parameters fixed at their estimated values. The differences between the two are smaller than in Fig. 7, but still non-negligible.

Finally, in Fig. 9, we show kernel density estimates of the plug-in predictive distributions for the areal proportion of Galicia in which 1997 lead concentrations exceed 3, 5 or 7 μg (g dry weight)$^{-1}$. In all three cases, recognition of the preferential sampling results in a pronounced

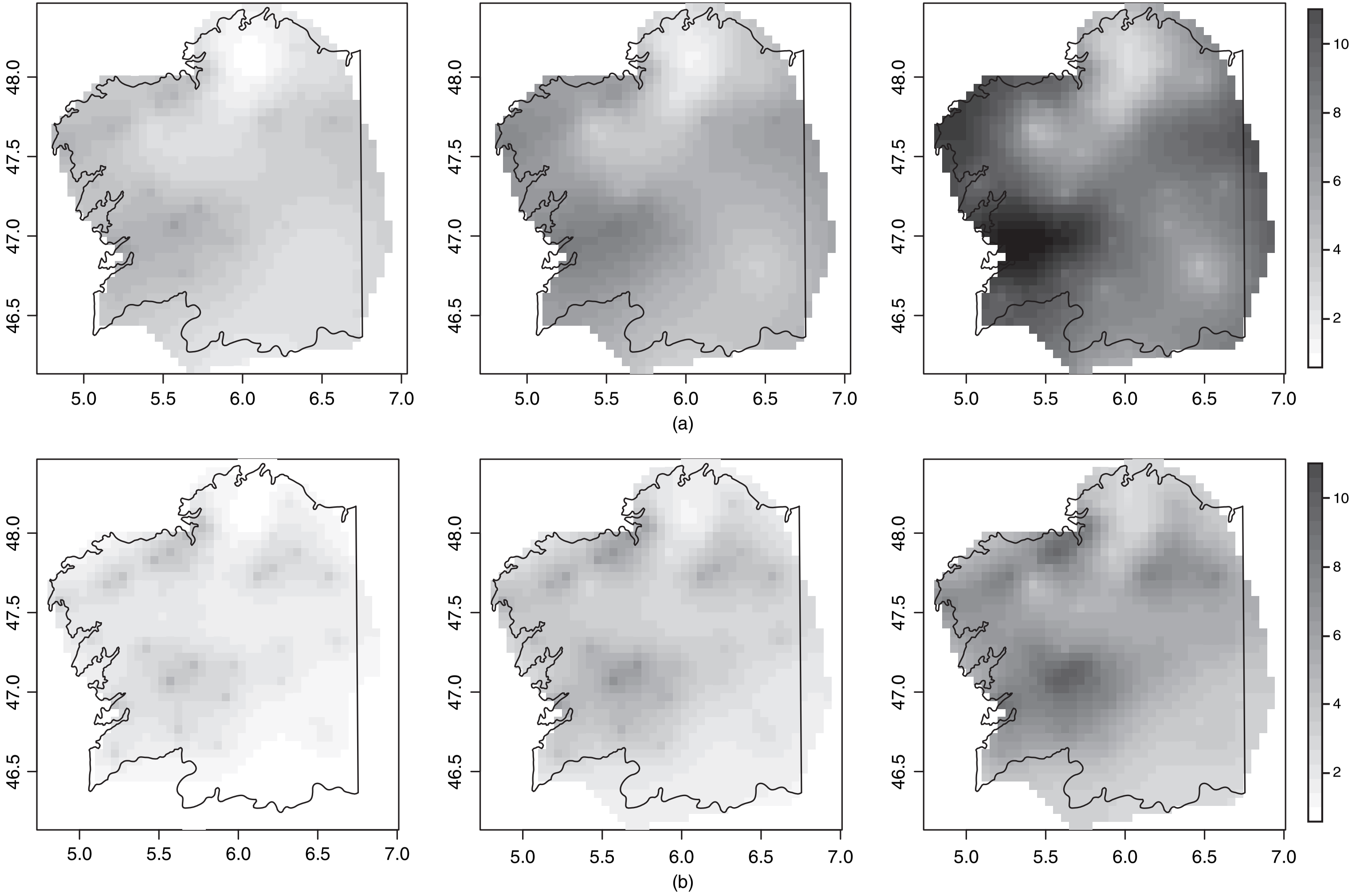
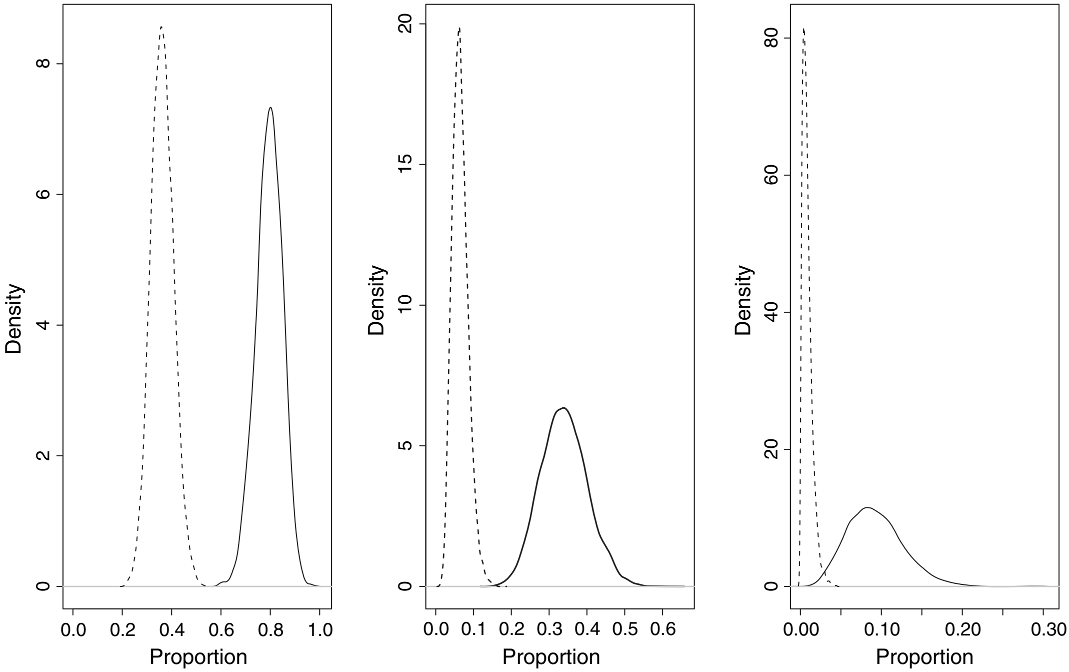

48.0
47.5
47.0
46.5
5.0
5.5
6.0
6.5
7.0
10
8
6
4
2
(a)
(b)

Fig. 8. Pointwise 5%, 50% and 95% limits of the predictive distribution of lead concentrations in 1997 under (a) preferential and (b) non-preferential assumptions: all six surfaces are plotted on a common scale, as shown



(a) (b) (c)

Fig. 9. Predictive distributions for the areal proportion of Galicia in which 1997 lead concentrations exceed (a) 3, (b) 5 and (c) 7 μg (g dry weight)$^{-1}$ under preferential (———) and non-preferential (-------) assumptions

shift in the predictive distribution. Note, however, that these plug-in predictive distributions do not account for parameter uncertainty.

Our overall conclusion is that the preferential sampling has made a material difference to our predictive inferences for the 1997 pollution surface.

6. Discussion

We have shown that conventional geostatistical models and associated statistical methods can lead to misleading inferences if the underlying data have been preferentially sampled. We have proposed a simple model to take account of preferential sampling and developed associated Monte Carlo methods to enable maximum likelihood estimation and likelihood ratio testing within the class of models proposed. The resulting methods are computationally intensive, each run taking several hours of central processor unit time. The computations that were reported in the paper were run on a Dell Latitude D620 laptop personal computer, using the R software environment (R Development Core Team (2008); see also `www.r-project.org`) and associated Comprehensive R Archive Network packages `fields`, `geoR` and `splancs`. The data and R code are available from `www.lancs.ac.uk/staff/diggle/`. There is undoubtedly very considerable scope to improve the efficiency of the authors' code. In particular, we did not seek to automate the progressive increase in the number of Monte Carlo samples as we explored the log-likelhood surface.

The computation of the Monte Carlo likelihood uses direct simulation, as in Diggle and Gratton (1984), rather than Markov chain Monte Carlo sampling. Hence, issues concerning convergence of the simulations do not arise, and the variablity between replicate simulations gives a direct estimate of the size of the Monte Carlo error.

We have described an application to a set of environmental biomonitoring data from Galicia, northern Spain. An important feature of these data is that they are derived from two surveys of the region of interest, the first of which used a spatially irregular set of sampling locations and has been shown to be preferentially sampled. In the second survey the sampling locations formed a nearly regular grid over the study region and we have therefore taken it to be non-preferentially sampled. This, coupled with our finding that several of the model parameters can be assumed to take a common value for the two samples, led to a better identified joint model for the two surveys. To illustrate this point, we also fitted the preferential sampling model to the 1997 data alone. Although, as reported earlier, the value of the maximized log-likelihood was obtained relatively easily, the subsequent quadratic fitting method to estimate the standard errors of the maximum likelihood estimates proved problematic. Using a 3^5+2^5 factorial design analogous to the earlier 3^6+2^6 design for the model fitted to the 1997 and 2000 data jointly, and with 10000 simulations for each log-likelihood evaluation as before, the quadratic fit explained only 72% of the variation in the Monte Carlo log-likelihoods, compared with 93% for the joint model, the implied estimate of $\partial^2 L/\partial\beta^2$ was not significantly different from 0, and the ratio of largest to smallest eigenvalues of the Hessian matrix was 34.5, compared with 22.3 for the joint model.

Alternative strategies for dealing with poorly identified model parameters could include treating the preferential sampling parameter β as a sensitivity parameter, since its value is typically not of direct scientific interest, or using Bayesian methods with informative priors.

A natural response to a strongly non-uniform sampling design is to ask whether its spatial pattern could be explained by the pattern of spatial variation in a relevant covariate. Suppose, for illustration, that S is observed without error, that dependence between X and S arises through their shared dependence on a latent variable U and that the joint distribution of X and S is of the form

$$[X,S]=\int [X|U][S|U][U]\,\mathrm{d}U, \qquad (12)$$

so that X and S are conditionally independent given U. If the values of U were to be observed, we could then legitimately work with the conditional likelihood $[X,S|U]=[X|U][S|U]$ and eliminate X by marginalization, exactly as is done implicitly when conventional geostatistical methods are used. In practice, 'observing' U means finding explanatory variables which are associated both with X and with S, adjusting for their effects and checking that after this adjustment there is little or no residual dependence between X and S. If so, the analysis could then proceed on the assumption that sampling is no longer preferential. In this context, any of the existing tests for preferential sampling can be applied, albeit approximately, to residuals after fitting a regression model for the mean response.

The value of seeking relevant explanatory variables to contribute to a spatial statistical model cannot be overstated. We hold the view that, in most geostatistical applications, spatial correlation reflects, at least in part, smooth spatial variation in relevant, but unobserved, explanatory variables rather than being an inherent property of the phenomenon being studied; an example to the contrary would be the spatial distribution of the prevalence of an infectious disease during an epidemic where, even for a uniformly distributed population in a completely homogeneous environment, the process of transmission from infectious to susceptible individuals would induce spatial correlation in the prevalence surface. This in turn leads us to emphasize that our paper is not a plea for uniform sampling, but rather for ensuring that any model for a set of data should respect whatever sampling design has been used to generate the data; for a thorough discussion that also uses point process models of sampling designs, albeit in a very different setting, see McCullagh (2008).

Returning to the geostatistical setting, and specifically to the application that was described in Section 5, Fernández *et al.* (2005) gave a practitioner's perspective on the ways in which different sampling designs can materially affect any analysis of spatial variation. To meet their primary objective of mapping concentration surfaces, they favoured regular lattice designs but noted that, for other purposes, 'additional sampling in areas of anomalously high concentrations of contaminants makes good sense'. We agree. Our paper formalizes this idea, while acknowledging that it necessarily complicates the subsequent modelling and inference.

Acknowledgements

This work was supported by the UK Engineering and Physical Sciences Research Council through the award of a Senior Fellowship to Peter Diggle.

We thank the Ecotoxicology Group, University of Santiago de Compostela, for permission to use the Galicia data and, in particular, José Angel Fernández, for helpful discussions concerning the data.

We also thank Håvard Rue for advice on efficient conditional simulation of spatially continuous Gaussian processes.

References

Aboal, J. R., Real, C., Fernández, J. A. and Carballeira, A. (2006) Mapping the results of extensive surveys: the case of atmospheric biomonitoring and terrestrial mosses. *Sci. Total Environ.*, **356**, 256–274.

Altham, P. M. E. (1984) Improving the precision of estimation by fitting a model. *J. R. Statist. Soc.* B, **46**, 118–119.

Chilès, J.-P. and Delfiner, P. (1999) *Geostatistics*. New York: Wiley.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. London: Chapman and Hall.

Cressie, N. A. C. (1985) Fitting variogram models by weighted least squares. *J. Int. Ass. Math. Geol.*, **17**, 563–586.

Cressie, N. A. C. (1991) *Statistics for Spatial Data*. New York: Wiley.

Curriero, F. C., Hohn, M. E., Liebhold, A. M. and Lele, S. R. (2002) A statistical evaluation of non-ergodic variogram estimators. *Environ. Ecol. Statist.*, **9**, 89–110.

Diggle, P. J. (2003) *Statistical Analysis of Spatial Point Patterns*, 2nd edn. London: Arnold.

Diggle, P. J. and Gratton, R. J. (1984) Monte Carlo methods of inference for implicit statistical models (with discussion). *J. R. Statist. Soc.* B, **46**, 193–227.

Diggle, P. J., Tawn, J. A. and Moyeed, R. A. (1998) Model-based geostatistics (with discussion). *Appl. Statist.*, **47**, 299–350.

Diggle, P. J. and Ribeiro, P. J. (2007) *Model-based Geostatistics*. New York: Springer.

Eidsvik, J., Martino, S. and Rue, H. (2006) Approximate Bayesian inference in spatial generalized linear mixed models. *Technical Report STATISTICS 2/2006*. Norwegian University of Science and Technology, Trondheim.

Fernández, J. A., Real, C., Couto, J. A., Aboal, J. R. and Carballeira, A. (2005) The effect of sampling design on extensive bryomonitoring surveys of air pollution. *Sci. Total Environ.*, **337**, 11–21.

Fernández, J. A., Rey, A. and Carballeira, A. (2000) An extended study of heavy metal deposition in Galicia (NW Spain) based on moss analysis. *Sci. Total Environ.*, **254**, 31–44.

Guan, Y. and Afshartous, D. R. (2007) Test for independence between marks and points of marked point processes: a subsampling approach. *Environ. Ecol. Statist.*, **14**, 101–111.

Henderson, R., Diggle, P. and Dobson, A. (2000) Joint modelling of measurements and event time data. *Biostatistics*, **1**, 465–480.

Ho, L. P. and Stoyan, D. (2008) Modelling marked point patterns by intensity-marked Cox processes. *Statist. Probab. Lett.*, **78**, 1194–1199.

Isaaks, E. H. and Srivastava, R. M. (1988) Spatial continuity measures for probabilistic and deterministic geostatistics. *Math. Geol.*, **20**, 313–341.

Lin, H., Scharfstein, D. O. and Rosenheck, R. A. (2004) Analysis of longitudinal data with irregular, outcome-dependent follow-up. *J. R. Statist. Soc.* B, **66**, 791–813.

Lipsitz, S. R., Fitzmaurice, G. M., Ibrahim, J. G., Gelber, R. and Lipshultz, S. (2002) Parameter estimation in longitudinal studies with outcome-dependent follow-up. *Biometrics*, **58**, 621–630.

Matérn, B. (1986) *Spatial Variation*, 2nd edn. Berlin: Springer.

McCullagh, P. (2008) Sampling bias and logistic models (with discussion). *J. R. Statist. Soc.* B, **70**, 643–677.

McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models*, 2nd edn. London: Chapman and Hall.

Møller, J., Syversveen, A. and Waagepetersen, R. (1998) Log Gaussian Cox processes. *Scand. J. Statist.*, **25**, 451–482.

Nelder, J. A. and Mead, R. (1965) A simplex method for function minimization. *Comput. J.*, **7**, 308–313.

Rathbun, S. L. (1996) Estimation of Poisson intensity using partially observed concomitant variables. *Biometrics*, **52**, 226–242.

R Development Core Team (2008) *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Ripley, B. D. (1977). Modelling spatial patterns (with discussion). *J. R. Statist. Soc.* B, **39**, 172–212.

Rue, H. and Held, L. (2005) *Gaussian Markov Random Fields: Theory and Applications*. London: Chapman and Hall.

Ruhling, A. (1994) Atmospheric heavy metal deposition in Europe: estimation based on moss analysis. *Nord*, **8**, 53.

Ryu, D., Sinha, D., Mallick, B., Lipsitz, S. R. and Lipshultz, S. E. (2007) Longitudinal studies with outcome-dependent follow-up: models and Bayesian regression. *J. Am. Statist. Ass.*, **102**, 952–961.

Schlather, M. (2001) On the second-order characteristics of marked point processes. *Bernoulli*, **7**, 99–117.

Schlather, M., Ribeiro, Jr, P. J. and Diggle, P. J. (2004) Detecting dependence between marks and locations of marked point processes. *J. R. Statist. Soc.* B, **66**, 79–93.

Srivastava, R. M., and Parker, H. M. (1989) Robust measures of spatial continuity. In *Geostatistics*, vol. 1 (ed. M. Armstrong), pp. 295–308. Boston: Kluwer.

Wälder, O. and Stoyan, D. (1996) On variograms of point process statistics. *Biometr. J.*, **38**, 895–905.

Wood, A. T. A. and Chan, G. (1994) Simulation of stationary Gaussian processes in $[0, 1]^d$. *J. Computnl Graph. Statist.*, **3**, 409–432.

Wulfsohn, M. S. and Tsiatis, A. A. (1997) A joint model for survival and longitudinal data measured with error. *Biometrics*, **53**, 330–339.